Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation- supported training with a conventional training approach

Ph.D Thesis

Amal M Alsalamah

MARCH 2016

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Obstetrics and Gynaecology Ultrasound
School of Medicine, Cardiff University

Acknowledgment

This study was funded by Ministry of Higher Education, Riyadh, Saudi Arabia. Portions of this research were presented in the UK and international. Special thanks for Saudi Arabia government for funding this research.

I would like to express my gratitude to my supervisors, professor Nazar Amso and Neil Pugh, whose constant guidance and encouragement has made it possible for me to get through this project.

This gratitude is extended to all the university staff especially professor Kerry Hood whose showered me with endless help with data statistical analysis and make it sounds enjoyable. My special thanks goes to my dearest colleague Gaynor Looker who has had a hand in helping me over the duration of the project.

I would also like to thank all the people I have met on the project who have made it possible for me to enjoy and get through the Ph.D years. My thanks are extended to my closest friends Azza, Dr. Hanan and Linda and to cousins especially Dr. Shada who have stood by me during the good and bad times. Without their support, this project may not have been possible. Special thanks to my sisters Asma and Latifa and all brothers especially Adil, Sami and Bandar, for their endless assisting to make a lot of things easier for me.

Finally, I would like to dedicate this project to my gorgeous parent, where without their continual encouragement and support, I would not have got through this. I would like to thank them for all the time, patience, support and endless love they have given me throughout the year. They were the best thing had happened in my life. Thank you so much.

Amal Alsalamah

Cardiff University, Cardiff

List of conference abstracts

1. **Alsalamah A,** D'Angelo A., Williams D. and Amso N (2008), Prevalence of incidental abnormal extra-ovarian findings in asymptomatic postmenopausal women undergoing Screening for Ovarian Cancer, Annual Scientific Meeting BSGI, London

2. **Alsalamah A**, Pugh N, Amso N, (2009) The effectiveness of using blended ultrasound training in obstetrics and gynaecology skills and on competencies of trainees, 3RD Saudi international conference, Surrey University, Guildford ***Awarded as*** *best outstanding paper of medical science.*

3. **Alsalamah A**, Pugh N, Amso N, (2009) The effectiveness of using blended ultrasound training in improving trainees ultrasound skills, The 8th International Scientific Meeting of Obstetrics and Gynaecology, AbuDhabi UAE

4. **Alsalamah A**, Pugh N, Amso N, (2011) The influence of a virtual simulator on the acquisition of trainee's ultrasound skills, Annual Scientific Meeting BSGI, London

5. Morgan H, **Alsalamah A**, Amso N, (2011). Comparison of Two Different Methods of Teaching Ultrasound Scanning to Medical Students. Annual Scientific Meeting BSGI, London.

6. **Alsalamah A**, Pugh N, Amso N, (2012) Learning through simulation: Training and assessment of ultrasound skills using ScanTrainer Simulator, Annual Scientific Meeting BSGI, London

7. **Alsalamah A**, Pugh N, Amso N, (2012) Assessment and learning curve evaluation of TVUS skills following simulation and clinical training, 21st Annual Congress ESGE, Paris
8. Amso N, Mullins S, Langan G, **Alsalamah A**, Pugh N (2012) Ultrasound simulation training in the clinical setting: Impact on skills development and barriers to implementation, 21st Annual Congress ESGE, Paris

9. **Alsalamah A**, Pugh N, Amso N, (2013) Construct validity for ScanTrainer ultrasound simulator, Annual Scientific Meeting BSGI, London

10. **Alsalamah A**, Pugh N, Amso N, (2013) Study the construct validity for ScanTrainer ultrasound simulator, 22nd Annual Congress ESGE, Berlin

11. **Alsalamah A**, Albalushi, Hood K, Pugh N, Amso N, (2013) Intra and inter- rater reliability of ultrasound skill assessment using video-recordings method in the ScanTrainer simulator, Postgraduate Research Day, Cardiff

12. **Alsalamah A**, Albalushi D, Hood K, Pugh N, Amso N (2014) Validation of structured scoring sheet of recorded videos for assessing trainees' competency in ultrasound, Annual Scientific Meeting BSGI, London

13. **Alsalamah A**, Albalushi D, Hood K, Pugh N, Amso N (2014) validation of video recording scoring system in assessing trainee competency in transvaginal ultrasound scanning skills, 24th World Congress in Ultrasound in obstetrics and gynaecology ISUOG, Barcelona.

14. **Alsalamah A**, Campo R, Tanos V, Grimbizis G, Van Belle Y, Hood K, Amso N (2014) Face and content validity of the ScanTrainer ultrasound simulator, 23rd Annual Congress ESGE, Brussels

15. **Alsalamah A**, Hood K, Pugh N, Amso N (2016) Experts and novices performance of basic transvaginal ultrasound (TVUS) using the ultrasound simulator, Annual Scientific Meeting BSGI, London

16. **Alsalamah A**, Hood K, Pugh N, Amso N (2016) Acquisition of ultrasound skills with or without simulation support: A Randomised controlled trial Annual Scientific Meeting BSGI, London

Abstract

Delivery of ultrasound training remains a challenge. This thesis presents a series of projects that investigated a new approach in acquiring transvaginal ultrasound skills (TVS) in obstetrics and gynaecology using a novel virtual reality simulator (ScanTrainer®, Medaphor plc, Cardiff, Wales).

Aims and objectives: (1) To evaluate the validity and reliability of the simulator, (2) to assess the learning curves of trainees' competence in performing TVS, and (3) to define potential benefits and limitations of simulation training from the trainee's perspective. These were achieved by undertaking the following studies (1) face, content and construct validity of the simulator, (2) reliability of scoring systems developed for the assessment of ultrasound in obstetrics and gynaecology, (3) validation of simulation scoring system against experts, and (4) evaluating the role of simulation on TVUS skill acquisition (learning curve) in the clinical training environment.

Methods: The projects included observational, comparative and semi-qualitative studies and randomised controlled trial (RCT) comparing conventional with simulation supported training.

Results: (1) Face and content validity study demonstrated high acceptability of the simulator. (2) Construct validity study showed significant differences between novices and experts' performances, $p \leq 0.05$. (3) Validation studies showed excellent agreement (i) between two observers in assessing simulated TVUS performances and (ii) between the observer and the simulator scoring system (*intra-class correlation coefficient* ≥ 0.75). (4) In the RCT, the overall analysis according to the randomisation arm showed no statistically significant difference between the intervention and control groups. (5) Fifty-seven percent of trainees agreed that simulation was a flexible learning platform in practicing TVUS as an adjunct to clinical training.

Conclusion: The ScanTrainer® simulator has high face, content and constructs' validity that support the research hypotheses. It also has a potential role in the assessment of clinical skills. However, the impact of simulation on the learning curves requires further evaluation.

Abbreviations

TVUS	Transvaginal Ultrasound Scan
TAS	Transabdominal Ultrasound Scan
RCOG	Royal College of Obstetrics and Gynaecology
RCR	Royal College of Radiologists
BEME	The Best Evidence Medical Education
NHS REC	National Health Service Research ethical committee
SEWREC	South East Wales Research Ethics Committee
LMS	Learning Management System
CRM	Crisis Resource Management
CVS	Chorionic Villus Sampling
VR	Virtual reality
SDD	Shoulder Dystocia Delivery
PPH	Post-Partum Haemorrhage
VBD	Vaginal Breech Delivery
CRL	Crown-rump length
GS	Gestational sac
YS	Yolk sac
POD	Pouch of Douglas
TURP	TransUrethral Prostate Resection
ST	Speciality Trainee
ICC	Intra-class coefficient
GRS	Global rating scale
RCT	Randomised controlled trial
IS	Individual skill
FE	Full examination
EPAU	Early Pregnancy Assessment Unit
OSATS	Objective structured assessment technical skills
OSAUS	Objective structured assessment of ultrasound skills
HSCIC	Health and Social Care Information Centre
SRS	The Safety Reporting Systems
NTSB	National Transportation Safety Board
VAS	Visual Analogue Score
SD	Standard deviation
KM	Kaplan Meier
GMC	General medical council
NICE	National Institute for Health and Care Excellence
AIUM	American Institute of Ultrasound in Medicine
ISUOG	International Society of Ultrasound in Obstetrics and Gynaecology
EFSUMB	European Federation of Societies for Ultrasound in Medicine and Biology

EFAST	Extended focused assessment sonography in trauma
TRUS	Transrectal ultrasound
dVSSS	de Vinci Surgical Skills Simulator
SSC	Student selected component
URS	Semirigid uretero-renoscopy
HFS	High-fidelity simulator
EM	Emergency medicine
CPR	Cardiopulmonary resuscitation
CT	Computed Tomography
MRI	Magnetic resonance imaging
3D	Three dimension
mini-BAL	mini broncholeovar lavage
CAD	Computer-aided design
MIST VR	Mimic invasive surgical trainer virtual reality
RoSS	Robotix Surgical Simulated Systems
EBUS-TBNA	Endobronchial ultrasound-guided - Transbronchial needle aspiration
DUOSATS	The objective structured assessment of technical skills for duplex assessment of arterial stenosis
3DE IESS	Three Dimension echocardiographic intracardiac endoscopic simulation system

Table of contents

Chapter one

A review and study hypothesis

CHAPTER 1

1.Review and study hypothesis

Ultrasonography is considered the first line imaging method of choice in women's health. Pelvic sonography is commonly performed as a diagnostic procedure to evaluate a variety of obstetrics and gynaecological clinical conditions, which should include transabdominal (TAS) and transvaginal (TVUS) scanning techniques (ISOUG, 2014; Manting M, 2014). Transvaginal ultrasonography became central to the current the Royal College of Obstetrics and Gynaecology (RCOG) guidelines for the management of ectopic pregnancies for its ability to identify and assess female pelvic anatomy accurately (Amso and Griffiths, 2005). Ectopic pregnancy is the third-leading cause of maternal morbidity and mortality. The estimated prevalence of ectopic pregnancy is 1-2% of live births in developed countries, though it may be as high as 4% among those using assisted reproductive technology (Santos-Ribero et al., 2016).

Moreover, in an NHS maternity statistical report, dated February 2012, it is stated that the rate of ectopic pregnancy in England and Wales has increased in the last 5 years; 1.6 per 100 deliveries in 2006-2007 compared to 1.7 per 100 deliveries in 2010-2011 (HSCIC, 2012). According to a report of the National Institute for Health and Care Excellence NICE (2012), two thirds of ectopic pregnancy deaths are associated with substandard care and it was noted that improvement in diagnosis and management is of vital importance. It has been suggested that in healthcare in the UK, up to 10% of patients admitted to hospital are at risk of a patient safety event as reported by Mark (Hellaby M, 2013). One of the areas of concern is the improvement of medical diagnoses and management and to ensure a quality of service delivery by healthcare professionals

TVUS is an initial step in assessing the female pelvis and should be performed by practitioners with a high level of training. Learning ultrasonography has increased

gradually with an increased need of trained physicians. Ultrasound is an integral part of the RCOG's ultrasound training programme to ensure trainees develop the skills they need to use and understand in clinical Obstetric and Gynaecological ultrasound practice. The challenge of acquiring sufficient skills in a reduced training time, in order to function safely at skilled practitioners' level, is a problem not only confined to obstetrics and gynaecology but also applies to all specialties where trainees need to acquire practical skills (Moss et al., 2011; Moglia et al., 2015). Trainees find experience in scanning difficult to access and struggle to reach the RCOG required competencies (Burden et al., 2011; Maul et al., 2004). The specific causes of reduced training opportunities presented as pressure on scanning lists, decreased doctors' hours and increased patient expectation. A need to provide general exposure to specialties has raised concerns that traditional ultrasound training in a clinical setting has become very limited and unsatisfactory. In contrast to TAS training, the opportunities for trainees to hone their skills in TVUS may still be limited not only because of the nature of intimate scanning but on local referral patterns and structure of clinical rotations during residency.

Although ultrasound training is essential for obstetrics and gynaecology practices, the use of simulation in training is still in its infancy and there is very limited evidence in the literature on standard methods in teaching and assessing ultrasound skills using the virtual reality simulators. The rationale of using simulation training is mainly for patient safety as the simulation setting provides (1) a safe environment where the trainees can learn without risk of harm to patients, (2) an environment that is fully attentive to the learner's needs, (3) an opportunity for repetitive training, (4) can be adjusted according to learners need, (4) enables exposure to gradually more complex clinical challenges, (5) supports experiential learning. Despite a huge number of useful published works on simulation-based training and how that dramatically improves trainees' skills, numerous questions remain, pertaining to the learning curve, long-term skill retention and measurement of trainees' competency after ultrasound simulation training (Sidhu et al., 2012; Madsen et al., 2014; Chalouhi et al., 2015a).

1. Overview

A review of the available literature was conducted to present the current evidence regarding the benefits of simulation training in improving trainees' performance in

clinical practice. Simulation is a new educational method of learning TVUS. Relevant evidence was explored in the literature to evaluate and critique the simulation learning elements in the assessment of ultrasound skills and to address the potential factors that minimise progress in learning basic TVUS skills and offer possible solutions.

Since there is limited evidence available in the literature that specifically identified the role of simulation in assessing TVUS, other evidence from different medical specialities was considered. The review in this chapter clarifies the search methods applied, the inclusion/exclusion criteria for searching, and the key words used.

2. Search strategy

Literature searches were conducted to comprise five search aspects:

SEARCH 1: aimed to review evidence of simulation in medical education and the learning theory used in practice.

SEARCH 2: aimed to review evidence of assessment of skills and performance in transvaginal ultrasound when using simulation training.

SEARCH 3: aimed to review evidence of validation studies, which includes face, content and construct validity studies of the simulators.

SEARCH 4: aimed to review evidence of simulation history in other specialities and in transvaginal ultrasonography in particular, stating the types and development of training simulators in obstetrics and gynaecology ultrasound.

SEARCH 5: aimed to review evidence of recently published studies of the virtual reality ultrasound simulator ScanTrainer®.

The literature search comprising the five searches was conducted in English language journals and studies. A comprehensive search was carried out using the following electronic databases: PubMed (1977- February 2016), EMBASE (1947- February 2016), MEDLINE and OVID MEDLINE (1946- February 2016). Bibliographies of all relevant studies and systematic reviews were searched by hand. EndNote, versionX7, the software was used to manage bibliographies, citations and references generated from the searches. Key words for the five searches are described in Table (1.1). Inclusion criteria applied to all studies that used the simulator as an educational, learning or assessment tool to measure learning curves and competence including validation studies. All publication types were considered including: review papers,

conference proceedings, experimental trials, quasi-experimental trials, and observational trials. Exclusion criteria included grey literature, also all studies that aimed to assess the computerised system or software, logarithms or statistics built-in to the simulator. These studies were excluded by the title, abstract, method or even the full text.

Table (1.1): shows the search strategy used in PubMed, EMBASE and MEDLINE databases.

	SEARCH 1	**SAERCH 2**	**SERACH 3**	**SEARCH 4**	**SEARCH 5**
Search strategy	1=simulation* 2=medical education 3=1 OR 2 4=theory of learning 5=clinical education 6=clinical behaviour* 7= learning 8=conceptual framework 9=cognitive 10=outcome assessment 11=practice 12=skills 13=performance 14=practical proficiency 15=theory of adult learning 16=independent learning 17=teaching 18=simulator* 19= 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 20= 3 AND 19	1=simulation* 2=simulator* 3=ScanTrainer* 4=Virtual reality 5=ultrasound 6=1 OR 4 OR 5 7=transvaginal ultrasound* 8=sonography 9=obstetrics 10=gynaecology* 11=early pregnancy* 12=skill acquisition* 13=competency* 14=randomised controlled trial 15=simulation training 16=skill 17=performance 18=learning curve 19=sample size 20=training 21=learning 22=hypothesis 23=patient safety 24=video recording 25=checklist 26=global rate scale 27=experts 28=novice 29-trainee 30=rater 31=7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29 32= 6 AND 31	1=simulation* 2=simulator * 3=ScanTrainer* 4=1 OR 2 OR 3 5=validation process 6=validity 7=face validity 8=content validity 9=construct validity 10=evaluation 11=reliability 12=agreement 13=virtual reality 14=assessment 15=skill 16=performance 17=ultrasound 18=ultrasonography 19= gynaecology 20=expert 21=intermediate 22=novice 23=competence 24=standard setting 25=contrasting-groups 26=borderline 27=clinical performance 28= 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18 OR 19 OR 20 OR 21 OR 22 OR 23 OR 24 OR 25 OR 26 OR 27 29= 4 AND 28	1=simulation* 2=simulator * 3=virtual reality 4=1 OR 2 OR 3 5=aviation 6=games 7=surgery 8=ultrasound 9=anaesthesia 10=transvaginal ultrasound 11=training 12=basic skills 13=performance 14=5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 15= 4 AND 14	1=simulation* 2=simulator * 3=ScanTrainer* 4=1 AND 3 5=ultrasound* 6=transvaginal scan 7=obstetrics 8=gynaecology 9=early pregnancy 10=skill 11=performance 12=assessment 13=learning curve 14=validity 15= 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 16=15 AND 4

3. Learning theory in simulation

The use of simulation to enhance learning is increasingly used in health profession education. Simulation training can be used to address basic science concepts and clinical medical expertise, including both cognitive and practical skills in an interactive way that reflects the workplace (Issenberg et al., 2005). The Best Evidence Medical Education (BEME) review describes the features that lead to effective learning. Those are based on a range of learning theories, i.e. from novice to expert. The simulation seems to be a good learning technique, as it provides the opportunity to plan according to the needs of the learner and incorporate feedback easily. The rationale for using simulation training in medical education is to provide the opportunity for formative assessment, feedback, learning how to learn and summative assessment.

A number of learning theories are available in the literature to aid designing and delivering simulated clinical experiences, to be used not only to develop the appropriate research questions but also to affirm learning credibility. The two main educational theories associated with simulation-based education are deliberate practice 'behaviourism' and experiential learning 'constructivism' (Hellaby M, 2013). In simulation training, the focus is primarily on learning the skills for healthcare education and clinical practice (Gaba D, 2004). True learning is the ability to provide an appropriate response or feedback to a given stimulus as defined by behaviourists. Therefore, simulation learning is a process of behaviours/actions that are performed with desirable consequences where the student is more likely to repeat the skill/performance if the skills are desirable; if not, the student is less likely to repeat it, simply because behaviourism focuses on the behaviours of students. Moreover, the teaching is to ensure that the student can flawlessly repeat the information or the performance being taught under appropriate circumstances or received from the simulator. Bandura (2001) pointed out that the feedback provided from simulation training causes a behavioural change in skills and performance. In addition, Piaget (1973) and Bruner (1966) accentuated that learners could be more effective in developing new ideas and hypotheses from the knowledge received instead of being merely recipients.

With regard to constructivism, cognitive learning theorists defined learning as a highly individualistic process where knowledge should be seen as relevant to the learner, despite the differences in the motivations for learning (Wulf et al., 2010). Conversely to behaviourism, cognitive learning theory accentuates learners as active constructors of knowledge who are constantly trying not only to understand new information but also to expand the understanding of it, due to become more competent. This is precisely how the simulator plays the role of teacher in structuring the information and presenting it formulaically to ensure it is timely, relevant and meaningful, hence considered a useful tool (Distlehorst et al., 2000). It has been agreed that the appropriate training setting for clinical practice is to be in clinic, however, that is not always possible (Gallagher and Tan, 2010). Therefore, rehearsing the performance in a similar, alternative, practical environment is required, in order to develop trainees' ability and confidence (Piaget, 1973).

3.1 Theory of adult learning:

In the literature, theories on learning and teaching are voluminous, thus only the key principles for simulation training are considered. According to the renowned American educator, Malcolm Knowles in 1980, there are five principles or assumptions concerning adult learning (andragogy), i.e. how adults learn and their attitude towards motivation for learning (Knowles M, 1975; Candy P, 1991; Kaufman D, 2003; Kearsley G, 2010). Those assumptions are listed as: (1) adults are self-directing and independent learners and so are motivated to learn, (2) adults have accumulated a great deal of experience which (including mistakes) provides the basis for learning activities, (3) adults are most interested in learning tasks that have immediate relevance and impact to their job or personal life, (4) adults are more interested in immediate problem-centred learning rather than being subjects centred or content-oriented, (5) adults need to be involved in the planning and evaluation of their instruction.

Simulation-based training is offered to a wide range of learners, from novices to experts, representing a variety of health professions, targeting a number of skills and offering a diversity of contexts, i.e. simulation settings. The rationale for using simulation training is mainly for patient safety as the simulation setting provides: (1) a safe environment where the trainees can learn without risk of harming the patients, (2)

an environment that is fully attentive to the learner's needs, (3) the opportunity for repetitive training, (4) training that can be adjusted according to the learner's need, (5) exposure to gradually more complex clinical challenges, and (6) support for experiential learning (Parush et al., 2002; Good M, 2003; Cook et al., 2011).

According to others' experiences in evaluating clinical simulations, there are four broad areas that have been addressed to provide a conceptual framework for judging the usefulness of a given simulator (Huang H, 2002; Scalese et al., 2008). Firstly, gain and maintain practical proficiency. Ericsson and colleagues (1993) compellingly argued that the primary goal of practice is to improve some specific aspect of performance and thus practice must focus on a well-defined area, supported by detailed immediate feedback, and provide opportunities for the gradual improvement of the same tasks. Moreover, practice with simulators should be broken down into smaller learning elements or contents along with long continuous, uninterrupted practical sessions (Donovan and Radosevich, 1999). Secondly, the place of expert assistance in task-based learning provides opportunities to deal with not only core skills but also advanced and complex ones. Thirdly, learning within a professional context makes the simulation reflect the contextual realities of everyday practice to provide an effective adjunct to clinical experience. Practising tasks on isolated models may inevitably present only a partial picture, thus practice with simulation should aim to improve integrated performance (Ericsson et al., 1993). Fourthly, the affective component of learning to any educational experience may exert a powerful positive or negative effect (Ferro T, 1993). Although, the clinical needs of the patient must always take priority over the educational needs of the learner, simulations deliberately place the learner's needs at the centre of attention and provide the opportunity to create conditions of best practice for teaching (Mann K, 1999).

Together, these theories provide important insights into design structured learning through the simulation to fulfil the learner's needs. Although there is increased attention to simulation effectiveness in enhancing learner practical skills, most studies conducted in medical education have been at the lower levels of Kirkpatrick's evaluation hierarchy (Kirkpatrick D, 1994). It is important to note that the four levels of Kirkpatrick in analysing training effectiveness are widely used in assessing the simulators. In terms of evaluating training and learning outcomes and deliverables,

those four levels are classified as:

Level 1- Participant reaction toward the training: to feel that training is a valuable experience.

Level 2- Improved learning: it looks at how much the trainee's knowledge has increased as a result of the training.

Level 3- Clinical behaviour change (transferability): it looks at how trainees apply the information given and to evaluate the change.

Level 4- Outcome results: by analysing the final results of training including the outcomes determined for the training.

The Kirkpatrick model is a great way to evaluate training in a scientific way, as well as providing a useful hierarchal framework against which to evaluate the strength of the evidence on the effectiveness of educational intervention (Kirkpatrick D, 1994).

With regard to simulation training in learning ultrasound skills in obstetrics and gynaecology, some studies have demonstrated the satisfaction of trainees toward their performances after simulation training (*Level 1*) (Monsky et al., 2002; Blum et al., 2013; Williams et al., 2013; Chalouhi et al., 2015b). There seem to be adequate reviews in the literature that demonstrate improvement of skills and performance of trainees after simulation training where their skills were assessed and measured by validated assessment scoring methods, i.e. checklist or global rating scale (*Level 2*) (Mandavia et al., 2000; Janse et al., 2013; Williams et al., 2013; Madsen et al., 2014; Tolsgaard et al., 2014a). In terms of evaluating the utility of simulation in medical education, only a few studies show direct improvement in clinical outcomes from the use of simulation training (*Levels 3 and 4*) (Okuda et al., 2009). Examples of improved patient care practices linked directly to simulation training include studies of better management obstetrics deliveries and endoscopic surgery (Kohn et al., 2000; Mishra et al., 2010). When using a combination of standard educational principles and practices, the Kirkpatrick hierarchy can be a powerful tool for assessing the impact of education and learning outcomes (Kirkpatrick D, 1994).

4. Relevant work in simulation: Aviation and games

As an effort to develop the education in terms of exploring new methods for trainees to learn, most educators and national societies and organisations embraced the simulation idea following the examples of other industries such as aviation and games

(Fisher et al., 2010; de Wit-Zuurendonk and Oei, 2011). With the aviation application, the Antoinette trainer was the first pilot training device (flight simulation) that was developed in 1909 to help pilots fly (Grossman E, 1919). Afterwards, a number of pilot training devices were developed during World War I such as the aircraft simulator "UK Trainer" which was developed by H.G. Anderson in 1915, then the "Cockpit Trainer" by Lender & Heidelberg in France and the "Ruggles Orientor" by W. Ruggles in the U.S. in 1917 (Higdon D, 2008). It was demonstrated that those devices could help pilots to learn the feel of the controls while proceeding along the ground; however, none of these trainers had significant value for training. That may referred to unreliability of the wind created at that time (Hume W, 1983). Indeed, flying schools were using the simulation trainers to train pilots in order to reduce training costs and improve safety for pilots and aircraft.

The efforts to develop a valid and robust flight simulator were continuous until the "Link Trainer" simulator was developed and designed carefully to teach new pilots flying skills and overcome previous flight training limitations. The Link Trainer which was known as the 'Blue box' and 'pilot Trainer' were continuously refined, with the U.S. military becoming one of the most effective users and buyers of simulators. In 1934, after a series of tragic 'Air-Mail' flying accidents, the Army Air Corps bought six Link Trainers to assist in training mail delivery pilots to fly at night and in bad weather (Link E, 2000). Between 1941 and 1945, over 10,000 Blue box Trainers were used to train more than 500,000 pilots. Following World War II, the Link Trainer was developed to become an electronic training device more suited to new high performance aircraft. In 1960, Link developed the simulators beyond just being for flight training to include training for space programmes including the Apollo mission for moon landing (Link E, 2000; Blättler el al., 2011).

As simulation technology advanced in the early 1970s, simulators became more professional and demonstrated effective outcome especially in comparison with aircraft, in terms of critical manoeuvres training and learning other technical flight skills (Abrahamson et al., 1969). The aircraft was used as a primary trainer for pilots that required them to fly countless hours without income from passengers or cargo for the airlines. In terms of safety transportation, one of the most critical tasks in aviation was the ability to land a large number of passengers (200 passengers). The Safety

Reporting Systems (SRS) and National Transportation Safety Board (NTSB) investigated the causes of air carrier accidents and showed that human error was a contributing factor in 60–80% of accidents (Billings and Reynard, 1984; Sexton et al., 2000). Another study notified that many problems were encountered by flight crews, having little to do with technical aspects, when the operation dealt with a multi-person cockpit (Dillard A, 2002). In fact, problems were associated with poor group decision-making, ineffective communication, inadequate training and poor tasks or resource management (Hume W, 1983). The benefit of simulation training over the aircraft itself, is that the simulator allows the individual pilot to practise very specific technical skills, while aircraft training certified flying with two or three pilots. Nowadays, flight simulators form an integral part of pilots' training as simulation training becomes more realistic for both new and experienced pilots (Roscoe S, 1980; Dillard A, 2002; Higdon D, 2008).

Despite that some studies has applied Kirkpatrick's four levels evaluation model to their evaluation of simulation skills, it was hard to realise the full four level evaluation. Kirkpatrick's evaluation model contains four levels: reaction, learning, behavior and results. Reaction level evaluates what participants thought and felt about the training; learning level evaluates the resulting increase in knowledge and/or skills and change in attitudes; behaviour level evaluates change in job behaviour due to training program; results level evaluates the final results that occurred because of attendance and participation in a training program. For example, as aircraft carrier marshalling training has high requirement of skills, and only questionnaire used to evaluate these skills. Tain et al. (2015) pointed out that organisational constraints substantially limit opportunities for collecting results data to evaluate transfer of training. However, Kirkpatrick's model is not easy to be carried out from the first level to the fourth level, as proceeding at each level, the evaluation becomes more difficult and requires more time. This reminds that sponsors of training might have unrealistic expectations with regard to results level outcomes.

With regard to simulation in games, simulation has a long history in the gaming industry. Between the 1940s and 1960s there were various attempts to develop computer games and Thomas Goldsmith and Estle Mann worked on developing the first simulator for games in 1947 which was based on simulating a missile fired at a

specific target. The idea for that computer game was inspired by radar display, which was used during World War II (Roscoe S, 1980; Flight Simulation, Online). Later, in 1970, the advances in technology and massive use of computers made them more affordable than before (Denson and Abrahamson, 1969). As a result, computer-generated imagery was the first video game released in 1970. In the early 1980s, this application continued to develop to become the computer-generated imagery used in the film industry and later it incorporated 3D and 4D formats, which were introduced into ultrasound simulation later (Gaba and DeAnda, 1988).

A number of published studies in the literature demonstrated the effectiveness of simulation in games for learning; interestingly some of these studies used games during clinical training and the results showed that the participants enjoyed this type of training (Raybourn et al., 2007; Honey and Hilton, 2011; Jalink et al., 2014). For instance, in learning technical skills in laparoscopy, De Wit-Zuurendok and Oei (2011) conducted a systematic review of computer-based (serious) gaming as a new field in medical education and evaluated the current status of serious gaming in medicine. In that study, there were two groups: one group trained using games and another group trained in the traditional method. The findings demonstrated the effectiveness of simulation training in improving trainees' performance, also proved that a good method for learning clinical decision-making and patient interaction. Although there is very little definitive proof that gaming simulation learning is good for training healthcare professionals, there are areas outside healthcare such as aviation, when gaming simulation learning is considered to be effective (Lynch et al., 2010; Raybourn et al., 2011; De Wit-Zuurendok and Oei, 2011; van der Spek et al., 2013). We believe that on-going research in this field should provide evidence of the important role of gaming simulation learning in enhancing learner understanding, knowledge and skills and incorporating that into medical education in a motivated and interesting way.

5. Medical Training Simulation Models

Simulation is defined as the technique of imitating the behaviour of some situation or process by means of a suitably analogous simulation or apparatus, especially for the purpose of study or personnel training (Oxford English Dictionary, 2016). Simulation in medical education has a range of different types developed according to the

purpose of the learning outcomes. Simulation in medical training has a long history that started with the use of very simple models to enable learners to practise skills and technique e.g. in obstetrics. The first mannequin used for obstetrics dated back to the eighteenth century when Madame Du Coudray used a fetal model and pelvis to train midwives in France (Cox et al., 1994; Bradly P, 2006). Medical simulation did not gain widespread use in the following centuries, principally for reasons of cost and a reluctance to adopt new teaching methods. With advances in materials and computer sciences, a wide range of modalities has been developed including virtual reality and high-fidelity models often located in deducted simulation centres (Forrest and McKimm, 2010).

Published work on medical simulation identified that the main aim of developing simulators has been to improve the performance of trainees and thereby improve the quality of healthcare delivery and patient safety (Gaba D, 2004). Drawing on an extensive range of sources, simulation models range from simple models such as using an orange for injection training to complex ones such as virtual reality simulator models. With regard to medical simulators, those include a variety of devices that began with Box-Trainers to high-fidelity and virtual reality Trainers (Okuda et al., 2009). In reviewing a history of simulation, the developers of medical simulation models categorised those tools into part-task trainers, training task Trainers, Full-body mannequin trainers or high-fidelity simulator (HFS), screen-based simulation and virtual reality simulators (Cooper and Taqueti, 2004). Major attempts to produce usable simulators for training have focused on the physical appearance of the simulator, in terms of scope, realism or fidelity (Rosen K, 2008).

5.1 Partial/part-task trainer

However, the way a simulator is used depends greatly on the educational context and one physical simulator can be used in a variety of contexts. For instance, intravenous access training arm is usually cited as an example of a simple part-task trainer. In addition, a partial-task trainer is frequently an ideal tool when the learning objective is confidence or competency in performing the procedure. This type of trainer has been successfully used to teach clinical pharmacology (Muller et al., 2005), airway management (Kovacs et al., 2000), cardiovascular examinations (Jones et al., 1997), surgical cricothyroidotomy (Pettineo et al., 2009), central-line catheterisation (Barsuk

et al., 2009) and also for paediatric emergency procedures (Selbst et al., 1989). A considerable amount of literature has been published on the uses of part-task trainers in clinical settings. For instance, Brehmer and Swartz (2005) evaluated the effect of training on the bench model 'Mediskills' dexterity during semirigid uretero-renoscopy (URS) and the significant improvement after simulation training was indicated, $p<0.001$. Similar findings with Bowling et al.'s (2010) study, the evaluation of a low cost cystoscopy model, which includes freshly frozen cadavers, found that the procedure time for trainees was minimised significantly from 128.8 seconds at the baseline to 54.9 seconds after intervention. Wang and colleagues (2008) reviewed the literature addressing simulation partial-task training for emergency medicine (EM) residents perform critical procedures. The authors highlighted the importance of using simulations to practise infrequent procedures that would enhance patients' safety especially when performed by inexperienced residents.

5.2 Training Task Trainer

These simulators are widely used in anaesthesia (Rosenblatt et al., 2002), laparoscopic (Paisly et al., 2001; Gupta and Devarajan, 2004), obstetrics and gynaecology (Macedonia et al., 2003) and cardiology procedures (Hatala et al., 2008; Hunt et al., 2008). For instance, the Resusci Annie medical model for cardiopulmonary resuscitation (CPR) developed in the 1950s (Rosen K, 2008; Singh et al., 2013; Chalouhi et al., 2015a) and the Harvey model developed a few years after Annie (Jones et al., 1997). Rapid development took place for task trainer simulators and particularly targeted the training systems and produced imaging data from actual human cadavers to develop anatomical models with 3D viewing (Ackerman J, 1998). In that way, the images became the basis for many works in surgical simulations, medical virtual reality and were also used for internet-based simulation (Satava R, 2001). Like with many disciplines, a variety of new task trainers outside the surgical domains were developed later, in the 1990s. With regard to ultrasound simulation trainers, the UltraSim ultrasound simulation mannequin was introduced in 1995 and based on real sonographic patients' datasets. However, the model had initially replicated abdominal pathology relevant to obstetrics and gynaecology, this had expanded later to represent diverse intra-abdominal problems. UltraSim trainer was one of the first mannequin products used for instruction manuals and clinical case presentations (Monsky et al., 2002).

5.3 Screen-based simulation models

This model is defined as a program that is exclusively computer-based allows learners to interview, examine, diagnose and work through clinical cases in a manner that mirrors clinical practice. The Air Medic 1 is an example of screen-based simulation and it is defined as a system that integrates screen-based teaching/simulation around communication and teamwork with a biofeedback sensor. In that way, students are able to manage their own emotional responses in stressful situations. Biese et al's (2009) study demonstrated improvements in resident knowledge, confidence, and performance of certain skills in simulated paediatric cardiac arrest scenarios. Also it suggested that screen-based simulation might be an effective way to enhance resuscitation skills of paediatric providers. With regard to ultrasound simulation, the SonoSim abdomen ultrasound simulator is considered one of the screen-based simulations allows trainees to choose from a variety of different cases/pathologies and also to switch between different display modes on patients, such as Computed Tomographic (CT), Magnetic resonance imaging (MRI) or ultrasound (Ehricke H, 1998).

5.4 Full body Mannequin Trainers/ high fidelity simulators (HFS)

These trainer simulations are highly sophisticated systems linked to haptic devices and able to be used by individuals or teams. Some mannequins are used solely on a specific region, such as head and neck (Adler et al., 2007) or full body length and able to present breathing, blinking, heartbeat and pulse. Usually these models are used for training differential diagnosis and treatment in surgical and obstetrics procedures. These simulation systems consist of many steps and tasks for trainees to accomplish a procedure, such as in colonoscopy (Van Sickle et al., 2011), carotid angioplasty (Willaert et al., 2011) and laparoscopic cholecystectomy (Stefanidis et al., 2007). Educators expand limits of any single simulation tool by combining two or more models to create hybrid simulations. Overly and colleagues (2007) used standardised patients along with high-fidelity mannequins to teach an approach to dealing with difficult scenarios in paediatric emergency medicine (EM), including medication errors and sudden infant death. Girzadas and others (2009) enhanced their ectopic pregnancy simulation by using a hybrid set-up incorporating a high-fidelity mannequin and an endovaginal ultrasound task trainer. Both residents and faculty

evaluators highly rated the hybrid simulation as an educational activity compared with a high-fidelity simulator with ultrasound images.

5.5 Virtual Reality (VR) Simulators

Although virtual technology was deficient in tactile control-haptic until 1990, great efforts were made in the technology of medicine (Bajura et al., 1992). One of the earliest virtual reality operation environments was developed by Savata and Lanier in 1993, however the graphics, interaction and surgical interventions were extremely basic (Lanier J, 2011). The main principle of virtual reality is simply to simulate the experiences of computer games using multi-sensory data that combines sight, sound and touch alongside its interactive capabilities to give users greater control. This is broadly defined as a human–computer interface that simulates realistic environments of a three-dimensional (3D) digital world whilst enabling user interaction (Ahmed et al., 1996). Like with many disciplines, the virtual reality ultrasound simulator generates images representative of objects or environments with which the user interacts and thereby responds to those actions, e.g. in sonography scans, ultrasound images displaying female pelvic structures such as the uterus and ovaries as reconstructed from real scans and displayed virtually. In 1992, ultrasonic images of a fetus were superimposed onto video images of a pregnant woman's abdomen to provide an accurate and unique perspective for guiding physicians as they inserted probes into the body (Lanier J, 2011).

Addressing the underlying educational principles and practice of the virtual reality simulator, this demonstrates that virtual reality application and the haptic system are a combination of two learning approaches to provide training in sophisticated skills along with kinaesthetic and tactile sensations. Some examples of medical applications are endoscopic (Garuda et al., 2002), laparoscopic (Burden et al., 2011), endovascular procedures (Macmillan and Cushieri, 1999; Shah et al., 2001) and this system enables the generation of user data to be presented subsequently as detailed feedback on performance and maintained as on-going records.

On the other hand, virtual reality simulations are widely used in different applications such as aviation, engineering, the legal profession and games. With aviation, flight virtual reality training significantly impacts on improving pilots' skills and recently

flight simulators have not just been used for training but also are widely operated for private industries, cargo and military services (FAA, 1990). In engineering application, many companies have facilities for virtual reality simulation for designers to communicate with other people by offering computer-aided design (CAD) representation of the product or the design (Kutz et al., 2006). While the legal profession benefits from virtual reality technology in investigating courtroom procedures, allowing law students to interact with individuals in a fabricated courtroom, and replay accident scenes and other critical events to be reconstructed in courtrooms (Lanier J, 2011).

A great benefit of virtual reality simulation technology is that it allows trainees to engage with real-life events as they are simulated on computers, without causing any harm or risk to the patients (Schlectre et al., 1992). As discussed above, various theories of learning psychomotor skills suggest that repeated practice is essential for learning and maximum benefits (Kunkler K, 2006; Shah et al., 2008). Training on virtual reality simulation also offers a convenient learning environment to enable trainees to practise different, rare and complex clinical cases as they take place in reality (Burden et al., 2011). Although there are many publications demonstrating the advantages of virtual reality simulation in learning, the limitations of this technology vary depending on the type of application utilised or may reflect the technical limitations of the simulators (Merz E, 2006; Gurusamy et al., 2008).

6. Medical simulation: a review

The story of simulation is dominated by the efforts of pioneers who have struggled to improve training by using the resources available to them. These individuals have, however, largely worked in isolation and often failed to identify prior, related work that would have provided them with valuable guidance (Cooper et al., 2004). Therefore, it is difficult to write a definitive history of simulation, as any one innovation can rightly be ascribed to multiple groups, and publication of the results has often followed many years after the development of the technique (Rosen K, 2008).

Simulation in medical training has a long history that started with the use of very simple models to enable learners to practise skills and technique. Medical simulation

did not gain widespread use in the following centuries, principally for reasons of cost and reluctance to adopt new teaching methods. Many studies in healthcare demonstrated the usefulness of using simulations in improving skills and performances and addressing trainees' needs with attention to patients' safety (Loftin et al., 2006; Rosen K, 2008; Ahmed et al., 2011; Al-Rasheed et al., 2013). With advances in the technology of medical simulators, evaluation, validation and assessment of the simulators, whether virtual reality or high-fidelity simulators, became a main issue for developers in the medical field to judge and assess the robustness of the healthcare service provided by these devices (Feinstein and Cannon, 2001; Rosen K, 2008). In the past, the first simulators used in medicine were simple models of human patients (Kunkler K, 2006). In other words, medical educators recognised the importance of training using real patients as this is most effective for learning clinical skills, however this may have limitations in terms of legal and ethical viewpoints. Therefore the use of mannequins and simulators in medical education is considered useful (Nara et al., 2009).

In the next part of this review, we are concerned with a review of the published work in different medical specialities where simulators have been used, including anaesthesia, surgery, including laparoscopy and endoscopy, obstetrics and gynaecology, and ultrasonography, including transvaginal ultrasound scan.

6.1 Simulation in Anaesthesia

Several reviews and published works are concerned with the simulation in anaesthesia and some examples of these simulators are: "SimOne", high-fidelity anaesthesia simulators, "CASE™" and "GAS™". American anaesthesiologists have used simulated clinical environment and computer-controlled full-body patient simulators since the mid 1990s, for training procedures on crisis management and teamwork (Good and Gravenstein, 1989). A crisis resource management or inter-professional exercise finds the educational approach most beneficial in a simulated setting, thereby enabling trainees to rehearse skills in a safe environment (Rosenblatt and Abrams, 2002). As with other learning aspects, the continuous development of simulators and mannequins in anaesthesia has offered a wide range of realistic events for training novices and experts (Schulz et al., 2011). In terms of competence, Tuttle and

colleagues (2007) demonstrated improvement of residents' skills in carrying out the mini broncholeovar lavage (mini-BAL) procedure after simulation training. The majority of trained residents felt that their discomfort levels after training fell from 28% to zero while their main scores on performing tasks significantly improved from 49% to 93% ($p<0.01$). In addition, 79% of residents agreed that simulation training allowed integrating basic procedure skills such as in ambidextrous and hand–eye coordination manoeuvres into clinical settings. The findings are relevant to many other anaesthesia studies that investigated the effectiveness of simulation. This also supports the evidence that simulation training provides objective evidence of performance and for interactive and team performance assessment (Chopra et al., 1994; Good M, 2003; Ahmed et al., 2011).

With skill acquisition, simulation training is considered a masterful model that measures the speed of gaining skills or attains experts level. Good (2003) examined the effect of simulation training in anaesthesia on twenty-six novice residents' performance. In a three-week trial, novices were divided into two groups: one received a daily simulator training session while the other group received daily lectures and both groups were assessed in a similar way; theoretical by written test, and clinical test. Although the findings indicated no significant differences between the two groups, either in the written or practical tests, the simulator group's practical scores were greater than those of the other group. The evidence given in this study suggested that the faster improvement in clinical ability occurred when residents used a simulator. Furthermore, and after three months of that trial, all residents' skills were re-assessed and interestingly all residents had gained additional experience and improved to a similar level, which indicated that simulation training helps in skill acquisition and shortens training time in the very onset of learning (Good, 2003). Although there was no definite evidence to prove that the simulation training was able to replace traditional current anaesthetists' training to maintain skills. It has been agreed that simulation has benefits in terms of shortening the time of training, the acquisition of skills and is perceived as less stressful for trainees in anaesthesia induction in comparison to the real workplace (Schulz et al., 2011).

The simulation is increasingly linked with the evaluation, testing, and validation of anaesthetic techniques and equipment (Robertson and Bandali, 2008; Ross et al.,

2013). The best evidence provided in Ross et al (2013) review suggested simulation training leads to higher clinical and non-technical skill levels than didactic methods of teaching. Although that review filled a gap in the anaesthetic literature, the authors believed longitudinal work with evidence of transferability to clinical practice remains elusive and this is likely to be the case for other specialties.

Boet et al (2014) systematic review in transfer of learning and patient outcome in simulated crisis resource management (CRM), a small number of reviewed studies, found that CRM skills learned at the simulation centre were transferred to clinical settings (Kirkpatrick's level 3), and the acquired CRM skills may translate to improved patient outcomes (Kirkpatrick's level 4), including a decrease in mortality. In terms of transfer of learning to the workplace, the review results of all included studies but one, found a significant effectiveness of simulation-enhanced CRM training, including when compared with didactic teaching alone. In terms of skill preservation, there were conflicting results among studies. For example, in the study by Miller et al. (2012), transfer of CRM skills in the workplace was not retained after a month, while transfer was retained for at least five weeks in another study (Bruppacher et al., 2010). This review led to conclude that a few studies were able to examine transfer of learning to the workplace by healthcare providers and/or changes in patient outcome after simulation-based CRM training. Whereas majority were limited to lower-level outcomes, such as reaction of participants and learning that has been measured using further simulation scenarios. This approach leaves the studies open to the criticism that learners may have been taught to perform well only in the simulator and not necessarily in real life.

6.2 Simulation in Surgery

There are many similarities that can be drawn between pilots and surgeons; these specialist professionals have to learn to manage stressful and life-threatening situations that may be unpredictable and subject to change at any moment. Therefore, the benefits of simulation in the aviation industry have inspired attempts to bring simulators into surgical training (Shah et al., 2001). The discipline of surgery requires a significant amount of cognitive analysis and integration to gain surgical experience through traditional hands-on (see one-do one-teach one) in supervised clinical training (Moore and Bennett, 1995). Nevertheless, a massive amount of studies in the

literature about laparoscopic surgery were conducted to investigate the possible value in diagnostic accuracy, cognitive processes and procedural completion of the use of simulators (Derossis et al., 1998; Fried et al., 2004; Beyer et al., 2011; Kolozsvari et al., 2011). Existing surgical simulators in literature vary according to the purpose of the procedure, i.e. laparoscopy or endoscopy, all ranging from cadaver, physical trainer, computer-based or virtual reality simulators. Some examples of these are invasive surgical trainer virtual reality (MIST VR), daVinci skills simulator, Mimic dV-Trainer, Surgical Simulated Systems' RoSS', and Simbionix Robotix Mentor (Alzahrani et al., 2013; Tanaka et al., 2015).

King et al. (2016) carried out a recent review of simulation training in therapeutic endoscopic techniques in training and assessing competency. One reviewed study found that trainees who underwent procedural simulation before performing the procedure, such as central venous catheter insertion on an actual patient, had significantly improved complication rates and patient safety. The overall conclusion drawn was that simulation has been shown to increase the skill and learning curve of trainees. In terms of construct validity, other published work found no difference was indicated when experts performed the skills, however significant difference showed in novices' performance. This led to assent that the novice or less skilled subjects benefited more from the simulation in comparison to the highly skilled or experts. Moreover, it also suggested that trainers and instructors also gained advantages from simulation training, as these tools could play their role in teaching basic skills to trainees that usually required prolonged time to be learnt (Feldman et al., 2004; Moglia et al., 2015).

As a traditional approach to educating surgeons has been increasingly investigated. It is widely arguable that 'Learning by Doing' has failed to offer acquisition of skills in an organised and systematic approach. This may refer to several factors: limited training opportunities offered, random cases of patients that flow through clinics or operating rooms, and also increased concerns about the safety of patients and students (Fried et al., 2004; Stefanidis et al., 2012; Schreuder et al., 2014). In addition, traditional surgical training requires various degrees of supervision and clinical procedures for trainees to learn and have repeated practice on real patients until they become competent (Feldman et al., 2004; Kundhal and Grantcharov, 2009).

Competency of the performing surgeon is crucial and with limited availability of a systematic learning approach in the traditional method, training for surgical procedures has become a challenge (Kumar et al., 2015). Therefore, a number of training modalities have been developed to overcome this problem (Kneebone R, 2003; York et al., 2016).

Measuring the competency of trainees using the simulator, Ericsson (2007) defined the experts' performance with the simulator as the highest level of skill acquisition and a final plateau of performance. This finding has been similarly demonstrated in other studies that evaluated the learning curves of trainees and determined the competence approached by attaining experts' score/level (Eversbusch and Grantcharov, 2004; Hogle et al., 2007; Ward et al., 2014). Several published works outlined the main concepts of practice in terms of skill acquisition, suggesting trainees should motivate and strive to improve a specific aspect of performance. Trainees need valid, thorough and immediate feedback on their performance, also the opportunity to practise freely and repeatedly within a controlled environment, and to sufficient training sessions offered (Kneebone R, 2003; Ericsson K, 2007; Bayona et al., 2009; Kazemi et al., 2010). Although there is a substantial body of evidence on the effect of simulation based training on skill acquisition and improved learning curve in surgical procedures, little has previously been published on the transferability of skills to the operating room (Sturm et al., 2008; Burden et al., 2011).

The practical requirements for designing studies that examine improvements in patient outcome can be difficult due to the need for large sample sizes and a control group. In Boet et al (2014)'s systematic review, all of the randomised controlled trials (RTC) studies included in the review involved one time-limited intervention on a small number of subjects. It is possible that modification of patient outcome requires a whole series of interventions on many subjects. Nevertheless, in comparison with other high-stake industries, like aviation, despite several studies showing an improvement in pilots' behaviour in the cockpit, some studies showed the benefit of crisis resource management (CRM) pilot training on client safety were lacking. In simulation-based education, it has been suggested that larger sample sizes, more multicentre studies, and studies with less risk of bias are required to provide a precise measure of the effect that simulation-based education has on healthcare provider skills

in the workplace and patient outcome (Boet et al., 2014). Other systematic reviews showed that there is no need for more Kirkpatrick Level 1 (reaction) and Level 2 (learning) studies, since learners are virtually constantly positive toward simulation training and learning occurs when measured in a simulated environment (Gordon et al., 2012). Frequency of retraining, skill retention, and instructional design remain research priorities in studies investigating Kirkpatrick Level 3 (transfer of learning at the workplace) and Level 4 (patient outcome) outcomes.

6.3 Simulations in Obstetrics and Gynaecology

Physicians in obstetrics and gynaecology specialities recognised the importance of using simulators and mannequins to teach junior doctors obstetrics procedures and offer hands-on practice in a free learning environment (Al-Rasheed et al., 2013; Chalouhi et al., 2015a). With work-hour restrictions and time in the operation rooms becoming increasingly more expensive, there are limited opportunities available for trainees to gain basic knowledge and skills (Ennen and Stain, 2010). In order to fill this gap, there is a greater need for simulation training to help assure the best clinical outcomes for patients (Loveless et al., 2011). For many years ago, a doll and bony pelvis have been used to teach cardinal movements of labour and delivery in obstetrics while the E-pelvis mannequin was introduced for pelvic examination skills in gynaecology (Shain et al., 1982; Cox et al., 1994). At a present, there are a variety of simulation modalities to assess core competencies, teach proper techniques, and achieve and maintain proficiency in technical skills (Bradly P, 2006; Loveless et al., 2011; Haerizadeh and Frappell, 2014).

A few years ago, there was increasing interest in using obstetrics models not only for teaching students new skills but also to refresh experts' manual skills and decision-making skills under stress (Hay et al., 2015). The main concern was to teach basic performance in obstetrics as the basic knowledge that every obstetrician should know, especially in how to manage and perform such procedures, e.g. postpartum haemorrhage (PPH), vaginal breech delivery (VBD) and obstetrics performance in shoulder dystocia delivery (SDD), eclampsia (Letterie G, 2002). The immediate answer for these requirements is structured and well-designed simulation programmes in the speciality to provide the necessary learning elements to teach trainees, test their knowledge, assess their skills and manoeuvres for the best outcome (Ennen and Stain,

2010).

Croft and colleagues (2006) assessed trainees' performance on simulated SDD using the test–retest method. The findings revealed an improvement rate in trainees' ability to deliver fetuses after three weeks of simulation training of 82%, and after six months 84%, then twelve months 85%, when compared to pre-training rate which was 49%. That led to the conclusion that the performance learnt in the systematic approach was easier for the trainee to memorise and retain. Another study also demonstrated a significant rate in reduction of neonatal injury from 9.3% to 2.3% after simulation training in shoulder dystocia (Draycott et al., 2008). Similar results were found in other studies when assessing the skills in performing PPH and VBD, and there was improvement indicated after simulation training (Deering et al., 2009).

In terms of evaluating the learning curve of trainees' performance, authors have argued that the learning curve for performing in-utero stenting procedures such as amniocentesis, chorionic villus sampling (CVS) in traditional training, would not plateau until 175–200 procedures had been achieved (Merién et al., 2010). With the advance of technology and simulation, authors have become more interested in addressing the length of time needed for trainees to reach competency after simulation training compared to the traditional training. A large amount of studies concerned this issue, however few studies were able to prove significant findings. Thereby further investigations are needed to determine the specific number of procedures needed when considering training with simulation (Ennen and Stain, 2010). Pittini and others (2002) believe that simulation based training is able to shorten the time of learning in-utero stent procedures (Wijnberger et al., 2000; Nitsche et al., 2009; Nitsche and Brost, 2013).

Simulation in obstetrics and gynaecology used as an educational tool to assist in the transfer of knowledge, practising diagnostic and simple practical skills, surgical skills training, emergency drill training, and team training. Whereas simulation should not be perceived as a replacement for training with real patients, educators should embrace the opportunities that simulation provides and integrate it into current training programmes to maximise training opportunities and patient safety (Deering et al., 2013; Haerizadeh et al., 2014).

6.4 Simulation in sonography

Simulation was introduced to ultrasonography twenty years ago, unlike other specialties that used simulation centuries ago. Since then, rapid development of technology in computers has led to a faster growth of modalities in ultrasonography simulators (Blum et al., 2013; Mema and Harris, 2016). In 1995, Meunier and Bertrand stated that the first sonography simulator had been introduced by Grunst for echocardiography applications as they used a puppet as a physical model of a patient together with a dummy probe to be operated by the trainee. However, the limitation of the image database was identified and reconstructed later with a developed version of the simulator (Blum et al., 2013).

Ultrasound practice encompasses a wide variety of diagnostic and interventional procedures such as prenatal diagnosis or invasive procedures (Pittini et al., 2002; Terkamp et al., 2003; Maul et al., 2004; Evans et al., 2010). A large amount of published work has explored the effectiveness of ultrasound simulation training and proved its usefulness in terms of improving trainees' knowledge, acquisition of skills, proficiency and patient safety. However, skill transferability to a clinical setting and measuring trainees' competency has not yet been addressed (Sidhu et al., 2012; Blum et al., 2013). A range of ultrasound simulators have been reported in the literature, such as UltraSim, VirUS, EchoComJ, SONOSim3D, SonoTrainer, SonoSimulator, pelvic US task trainer Blue Phantom™ ScanTrainer® (Ehricke, 1998; Knudson and Sisley, 2000; Maul et al., 2004; Magee et al., 2007; Chung et al., 2011; Platts et al., 2012; Sidhu et al., 2012; Tolsgaard et al., 2015a).

A preliminary review in the literature about ultrasound simulators was undertaken by Maul (2004) to summarise the potential benefits of simulation-based ultrasound training. This review briefly described the properties of a variety of ultrasound simulators that have been developed for various applications including prenatal diagnosis and presented the SonoTrainer sonography simulation system. This system made it possible to run a real-time simulation of a complete prenatal ultrasound examination. It found that the simulation-based training enabled physicians to diagnose rare fetal anomalies in the second trimester with a sensitivity of 86% and a specificity of 100%. The author suggested that simulation-based training would provide an ideal educational tool to test and monitor a physician's or technician's

ultrasound skills in detecting fetal anomalies. In addition, that review also discussed the development of ultrasound simulators such as the simulator for gynaecology which was developed later for ultrasound mammography, for radiology, prenatal diagnosis, transvaginal ultrasound and intravascular ultrasound.

The first systematic review of ultrasound simulators was reported by Sidhu et al (2012). The review addressed the role of simulation-based education in ultrasound practice training and whether ultrasound and/or ultrasound procedural simulation leads to improvement in ultrasound competence, particularly in the clinical setting. The authors reviewed relevant studies in terms of sample size, study population, study design, ultrasound simulators used and measured outcomes. Most studies demonstrated acquisition of skills and improve knowledge and skills while others examined validity of the simulator. Agreeably, ultrasound simulation training has positive impacts on trainees' knowledge and performance.

According to a review of computer-based simulation for ultrasound training that was conducted by Blum et al (2013) to classify simulators according to the image, simulation method, user interactions and medical applications. A key advantage over traditional training was that simulators enable novel training concepts for advanced visualisation, case databases and automatically generated feedback. Conversely, an experienced trainer or doctor must be present to provide feedback in conventional training. In addition, some procedures are uncomfortable and painful for the patient, or even more problematic for some ultrasound-guided procedures. Therefore, Simulation training is valuable in offering safe settings to perform procedures with no harmful effects for patients. Furthermore, computer-based simulation has virtual scenes that enable many cases with different pathology and different difficulty levels to be provided simply by a software update, unlike with physical phantom which presents one specific pathology or with patients and coincidence pathology presented in clinical training. In contrast, some critical aspects have been identified about the simulator that make traditional training more valuable. The realism, artifacts, quality of the images generated and the unrealistic sensation of the haptic device that simulates interaction between probe and patient are reduced the reliability of the simulator (Bø et al., 2010). In order to overcome these limitations, a computer-based simulation system should be evaluated and there are different aspects of the simulator

such as face, content and construct validity. Face validity judges the degree of resemblance between the simulator and reality whereas content validity shows how appropriate a system is for teaching and to what extent it covers the real activity. Most studies investigated face and content by carrying out a questionnaire. Construct validity is the ability of a simulator to discriminate among subjects of different experience and to demonstrate cut-off or borderline of performances between experts' and novices' scores.

The evaluation of simulation in obstetrical and gynaecological ultrasound remains mainly at level 1 and 2 of Kirkpatrick until now. Most studies evaluate reaction, satisfaction or learning. Although several currently available ultrasound simulators include measurement of time to complete tasks, and accuracy of measurement, most studies have not yet evaluated the transfer of knowledge acquired during simulation training to clinical practice (Chalouhi et al, 2015a).

Konge et al. (2013) identified level of competence in performing transbronchial needle aspiration (EBUS-TBNA) based on objective assessment using a virtual reality EBUS simulator. Twenty-two respiratory physicians were divided into three groups depending on their EBUS clinical experience. Each physician performed a standard simulated EBUS procedure and contrasting-groups method used to determine the pass/fail level of standard setting of performance. It showed acceptable reliability of 0.8 in four selected simulation metrics and significance indicated among three groups' performances. The authors suggested that the simulator is essential in terms of credentialing prior to entering a supervised clinical practice. Authors in different studies arguably outlined that cognitive skills and knowledge in ultrasound practice are easily measured by written or oral exams, however competence is difficult to assess and measure (Sheehan et al., 2013; Chalouhi et al., 2015b). It is hypothesised in these studies that using the Echo simulator would used for testing competence in echocardiography ultrasound practice. The objective structured assessment technical skills (OSATS) checklist was used to assess skills of trainees with different levels of experience. The finding indicated a significant difference between novices and experts, thus construct validity was demonstrated for the Echo simulator.

Experts also benefited from simulation training, particularly in diagnosing the complex anatomy of congenital heart disease. Xue and colleagues (2010) introduced a novel method known as the 3D echocardiographic intra-cardiac endoscopic simulation system (3DE IESS) for the non-invasive imaging of intra-cardiac structures. The method is utilised to assist cardiologists and cardiac surgeons examining heart malformation through a 'virtual eye' that is flexibly positioned to any point inside the heart for better understanding of congenital heart disease, specifically in the pre-operative learning stage (Bose et al., 2009). A feasibility study carried out for 3DE IESS showed a reliable tool in terms of assessing congenital heart disease.

An up-to-date review of ultrasound simulation in obstetrics and gynaecology reported by Chalouhi et al (2015b), reviewed the existing literature to provide an overview of uses of simulation. The simulators in this field appeared at the beginning of this century, thus few studies have been published so far. Therefore, ultrasound simulation in obstetrics and gynaecology is still developing and could bring several benefits into teaching, training and evaluation of ultrasound competency and proficiency. Additionally, a review addressed a number of criteria and procedures that are required to perform obstetrics and gynaecology ultrasound in clinical settings according to international guidelines and recommendations such as the RCOG. However, the authors argued that theoretical knowledge was not sufficient, and practical training either achieved on real patients or on volunteers was still not satisfactory. This altogether suggests the conventional patient-centred approach puts trainees in an uncomfortable situation, especially during the initial phase of training. It is widely accepted by many different works in other specialities that simulation training is most valuable at this particular learning phase in order to gain and understand the performance (Lucas et al., 2006). Although interaction with the patient is essential in conventional training, this may distract trainees from becoming familiarised with the handling and manipulation of the ultrasound probe and with the interpretation of ultrasound images. The systematic learning approach through simulation would enhance the knowledge as well as the skills.

A challenge of ultrasound training is being an operator-dependent with sufficient skills due to obtaining high quality images. Simulation training has become a popular

method of learning especially for mastery of learning in an educational and professional context. Together, these studies have mainly shown positive learning effects with computer-based, virtual reality and other high fidelity models of simulators. However, some evidence is provided about the long-term effects on performance but not on measured transferable skills to actual clinical practice.

6.5 Transvaginal ultrasound simulator: ScanTrainer®

A number of available transvaginal ultrasound (TVUS) simulators are reported in the literature, such as pelvic US task trainer Blue Phantom®, SimMan Freehand 3D US device (GE EchoTech), UltraSim (MedSim), Schallware system, with the latest being virtual reality ScanTrainer®. Several reviews have been published with the aim of demonstrating the effect of simulation training. Although there is an increasing amount of data indicating positive effects of simulation training, research in ultrasound using the ScanTrainer® simulator is still in its infancy.

ScanTrainer® ultrasound simulator is at the heart of our research and is increasingly recognised as a new virtual reality simulator for learning TVUS skills. In 2010, MedaPhor Ltd (Cardiff Medicenter, Wales) unveiled a new haptic virtual reality TVUS simulator using modules with 'easy-to-follow tutorials and assignments' allowing the operator free practice in learning TVUS skills. Despite very limited published work has evaluated the acquisition of skills and the learning curve, the current existing findings are encouraging. Recently, researchers have shown an increased interest in ScanTrainer®. The amount of published work about the ScanTrainer® so far does not exceed six published studies (Williams et al., 2013; Gibbs et al., 2014; Madsen et al., 2014; Tolsgaard et al., 2015a,b; Carolan-Rees et al., 2015), two abstracts (Preshaw et al., 2012; O'Brien et al., 2015) and being referenced in two systematic reviews (Sidhu et al., 2012; Blum et al., 2013). The grey literature and unpublished work were excluded from this search.

Williams et al. (2013) conducted a pilot study to compare the use of ScanTrainer® and clinical training in improving TVUS skills and self-confidence for nine doctors in radiology and gynaecology specialities. The experiment was designed to test and re-test subjects' skills after ten hours of training. The assessment was under direct supervision for those in control and with simulation practice for intervention, and

interviewed the two groups to assess self-confidence. Although no significant difference was indicated between the two groups, the authors agreed that the simulation training enabled the teaching of basic TVUS similar to the way with a physical trainer. Thereby it concluded that it could replace initial clinical training. With regard to the self-confidence outcome, subjects felt more confident in scanning real patients after simulation training. Simulation training had an advantage over clinical training for being in a less stressful setting, allowing mistakes to be made and learning through unlimited repetition. Moreover, TVUS scanning is particularly difficult to teach within a clinical setting. This would include extensive training time, experienced tutors with plenty of available time and a wide variety of compliant patients, with normal and pathological findings, willing to be examined on multiple occasions. Despite the small sample size in Williams et al.'s (2013) study, and it has been suggested to investigate the effect with large scale study.

Another study was carried out by Madsen et al. (2014) to assess learning curve using ScanTrainer®. Due to limited evidence available to guide educators on how to assess simulated performance or how much time, amount of practice, training elements are needed to gain basic TVUS, the authors demonstrated valid and reliable performance measures and established performance standards to measure levels of competence. Developing competences in ultrasound is largely dependent upon the variety and the number of cases encountered during clinical practice and there is also a great diversity of opportunities among trainees. The study findings showed that novices were able to attain experts level and be competent in performing TVUS after three to four hours of practice. Although this did not indicate that novices had become proficient practitioners yet, they may be ready to enter supervised clinical practice.

As predicted, some trainees reach a level of competency that is suitable for clinical practice after a few scans, while others need more time to reach the same level. Thus, an acceptable way of assessing skills should be based on practical tests, during which the trainee demonstrates his/her abilities (Chalouhi et al., 2015b). Likewise several European training programmes in obstetrics and gynaecology rely on time spent in specialised ultrasound units for trainees to attain basic skills. On the other hand, Tolsgaard et al. (2014b) suggested that a minimum of 12–24 days of practice in specialised ultrasound units is highly associated with confidence of performing TVUS

scans independently. Agreeably, ultrasound simulation training has positive impacts on trainees' knowledge and performance. There was a subsequent study by Tolsgaard et al. (2015a) had shown that simulation-based ultrasound training leads to substantial improvement in clinical performance and that is sustained after 2 months of clinical training. Moreover, another study carried out by Tolsgaard et al. (2015b), assessed the effectiveness of ScanTrainer® training in pairs (dyad practice) as compared to training alone measured on subsequent clinical performance on patients. Findings showed lack of significant interactions between the two training types, learning outcome and skills transfer. The authors suggested that the dyad group did not tolerate having only half the hands-on time compared to that of the single group, however they were able to increase their efficiency in terms of number of attempts needed to achieve a certain simulator score (Tolsgaard et al., 2015b).

Further studies investigated the benefits of ScanTrainer® in improving trainees' TVUS skills and offering alternative educational strategy to enhance trainee competence. One study proposed a new clinical assessment framework for diagnostic medical ultrasound students and incorporating simulation training to standardise assessment of trainees' skills (Gibbs et al., 2014). While another study determined the cost viability of replacing clinical training with simulation training and found that ScanTrainer® is cost-saving for clinics (Carolan-Rees et al., 2015). With regard to two other abstracts, one study investigated whether obstetrics and gynaecology trainees with simulation training perform more scans than those without (O'Brien et al., 2015), while the other conducted construct validity for ScanTrainer® (Preshaw et al., 2012). Although findings were encouraging, further on-going investigation has been suggested.

Notwithstanding, improvement in knowledge and better recognition of clinical scenarios after training sessions on the simulator has been established, the necessity of supervised training even in simulation is still debated. In view of all that has been discussed, all studies provided important insights about simulation training in medical education, however the long-term effects in clinical settings should necessarily be examined.

2. Study hypothesis

2.1 Study hypothesis

Based on the literature review, evidence suggests that using simulation training supplemental to clinical practice is effective in terms of skill acquisition, enhancing trainee knowledge and practical skills, thereby shortening the training period compared to conventional training. Although few studies have been conducted into ultrasound simulation with encouraging findings, further investigations for simulation training in TVUS in non-controlled learning environments to assess learning curves is essential. Our hypothesis was to determine the length of time required for novice trainees to acquire the skills necessary to perform TVUS when using an ultrasound simulator, supplemental to their conventional training in a randomised controlled trial. The null-hypothesis was that no difference between the simulation training supplemental to conventional training and clinical training alone in skill acquisition. Other approaches were to investigate reliability and validity of an ultrasound simulator which were carried out through a number of experiments to demonstrate face, content and construct validity.

2.2 Overall aim and objectives of the study

The aim of this research was to determine the length of time required for trainees to acquire the skills necessary to perform TVUS, with the simulation training supplemental to clinical training in addition to determining reliability and validity of the ultrasound simulator ScanTrainer®.

The objectives of study were determined accordingly to goal of each experiment;
Experiment 1: Face and content validity
Experiment 2: Intra- and inter-observer reliability of scoring systems for ultrasound skills assessment
Experiment 3: Validation of simulator metrics

Experiment 4: Validation of subjects' performance on the Simulator: Construct validity
Experiment 5: Assessment of learning curves: Randomised controlled trial
Experiment 6: Participants' perceptions of the effectiveness of simulation practice: End trial survey

2.2.1 Experiment 1: Face and content validity (chapter three)

The aim was to determine face and content validity of the TVUS ScanTrainer®.
The objectives were: (1) to recruit practitioners with varying levels of ultrasound experience from attendees of an international conference, and (2) for study volunteers to undertake relevant simulator tutorials and complete a structured questionnaire including statements on face and content validity.

2.2.2 Experiment 2: Intra- and inter-observer reliability of scoring systems for ultrasound skills assessment (chapter four)

The aim was to test the reliability of scoring systems developed for the assessment of obstetrics and gynaecology ultrasound skills.
The objectives were to use scored video-recordings of ultrasound scans to (1) determine intra-observer (test and re-test) absolute agreement of the scoring systems for each independent observer individually, (2) determine inter-observer reliability between two independent observers' ratings to evaluate the consistency of two scorings and (3) test the level of agreement between the checklists and GRS scores of the two observers.

2.2.3 Experiment 3: Validation of simulator metrics (chapter five)

The aim was to ensure that the two metrics designs (individual skill task IS and full examination task FE) in the ultrasound simulator are consistent in providing identical feedback on TVUS performance by the same subject as well as feedback that is consistent with that given by a human judge. This should determines the reliability of simulation-based assessment and its suitability for reporting the actual performance of trainees and for reflecting their gradual change in TVUS practical level during the six assessments sessions in the RCT.
The objectives were (1) to determine the reliability between the two metrics designs in the simulator: "IS" and "FE", in providing consistent and identical feedback on

TVUS performance of participants; (2) to determine the level of 'absolute' agreement between simulator metrics (IS and FE) as compared with human observer due to select the appropriate simulator metric and use it for assessing participants' TVUS performance in randomised controlled trials and finally (3) to determine absolute agreement between the simulator metric (FE) and the observer during the six assessments in the randomised controlled trial.

2.2.4 Experiment 4: Validation of subjects' performance on the Simulator: Construct validity (chapter 6)

The aim was to assess whether the ScanTrainer® ultrasound simulator can discriminate between novice, intermediate and experienced-level practitioners in performing transvaginal ultrasound skills (TVUS).

The objectives were (1) to test the significance of scores achieved by the three groups in performing three simulation assignments; (2) to determine any significant difference between two scores obtained by the simulator and human observer in assessing subjects' performances; (3) to standardise the performance of contrasting-groups 'pass/fail scoring' method in order to measure levels of competence of TVUS practice.

2.2.5 Experiment 5: Assessment of learning curves: Randomised controlled trial (chapter 7)

The aim was to determine the length of time required for trainees to acquire the skills necessary to perform transvaginal ultrasound (TVUS).

The objectives were (1) to determine the trainees' speed of acquisition of ultrasound skills whether simulation-supported or conventional training; (2) to explore the potential factors that influence learning curves for two study groups, and (3) to explore the factors associated with each point on the learning curve; for example number of training sessions received, engagement to simulation training.

2.2.6 Experiment 6: Participants' perceptions of the effectiveness of simulation practice: End-of-trial survey (chapter 8)

The aim was to explore trainees' perceptions of simulation training as supplemental to their clinical training.

The objectives of this end-of-trial survey were to investigate current ultrasound training delivered to obstetrics and gynaecology trainees to determine (1) the benefits and limitations of ScanTrainer® training compared to a physical model i.e. mannequin Blue Phantom™, (2) the barriers and obstacles that have contributed to the gap in learning transvaginal ultrasound and (3) to clarify the potential solutions that might help in enhancing current ultrasound training.

Chapter two

Outline of methods

CHAPTER 2

Outline of methods

2.1 Approach

The methods used in the thesis to assess simulation and acquisition of ultrasound skills and performance in gynaecology and early pregnancy varied according to the objectives of each experiment/project. These methods used in the six exploratory experiments/projects included quantitative and semi-qualitative data whereby TVUS performance was evaluated by an observer and the ultrasound simulator. Methods used chapters three, four, five and six were applied to validate the simulator's metrics and to estimate the reliability of a set of simulation assignments and tasks that represent core transvaginal ultrasound (TVUS) skills and performances. Methods used in chapters seven and eight, of the thesis evaluated the learning curves of trainees in relation to acquisition of TVUS skills and performance in gynaecology and early pregnancy. This enables the researcher to understand trainees' needs while they were learning basic TVUS skills.

2.2 Contributions

The experiments/projects listed below were designed to achieve the overall aims of the PhD research and to demonstrate the acquisition of skills that trainees needed to perform basic TVUS procedures. The experiments/projects are discussed thoroughly in the next six chapters, which are:

1. **Chapter three:** Face and content validity of the virtual reality ultrasound simulator ScanTrainer®
2. **Chapter four:** Intra- and inter-observer reliability of scoring systems for ultrasound skills
3. **Chapter five:** Validation of simulation metrics
4. **Chapter six:** Validation of subjects' performance on the simulator: Construct validity
5. **Chapter seven:** Assessment of learning curves: Randomised controlled trial

6. **Chapter eight:** Participants' perceptions of the effectiveness of simulation practice: "End of trial" survey

2.3 Samples and methods:

Sample size calculation, methods/protocols, recruitment, inclusion/exclusion criteria of research subjects, and randomisation strategies for each experiment are discussed and explained independently in each chapter of the thesis.

2.4 The tool used in the experiments/projects:

Ultrasound simulator 'ScanTrainer®' (MedaPhor plc, Cardiff, UK) was the tool used for the six experiments/projects included in this research. Further details about the ultrasound simulator ScanTrainer® are found in section (2.7)

2.5 Statistical consideration

Statistical analysis plans and data collection were determined according to each experiment's objectives and are explained in each chapter. The statistical analysis details included testing the normality of the collected data.

2.6 Ethical consideration

- Local Research Ethical approval was obtained from South East Wales Research Ethics Committee (NHS REC Reference 10/WSE02/75) for all study aspects (Appendix I). All participants were required to provide informed consent (appendices II and III). Participants were free to withdraw at any time without giving any reason.

- The study's protocol for the randomised controlled trial has been registered with Control Clinical Trials with the reference number ISRCTN03408765, and public title of 'The influence of a virtual simulator on the acquisition of trainee's ultrasound skills' and the scientific title 'Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach' (Controlled-trials, 2013), Appendix IV.

- This study is an educational training research project that helps subjects to achieve a standardised level of ultrasound skills through simulation and clinical training. The researcher used an ultrasound simulator 'ScanTrainer®' to evaluate the subjects' acquisition of skills and no patients were involved in the assessments at any point in the trial.

- Video information and data will not be used for commercial purposes. Data will be retained for a period of 15 years. This retention period complies with guidelines set out by the Cardiff University Governance Framework. Data will be stored soft and hard copy format in School of Medicine, Cardiff University.

2.7 The ultrasound simulator ScanTrainer®

The ultrasound simulator ScanTrainer® is a simulation learning tool used for training, learning and assessing transvaginal ultrasound (TVUS) skills. The simulator contains learning materials installed by special software in a personal computer connected to a haptic device for free practice (Figure 2.1). The software provides real ultrasound images, which were initially generated from scanning real patients, in addition to virtual anatomy images for learning purposes (Figure 2.2). The image represented on the monitor is related to the probe movement that is attached to the haptic device.

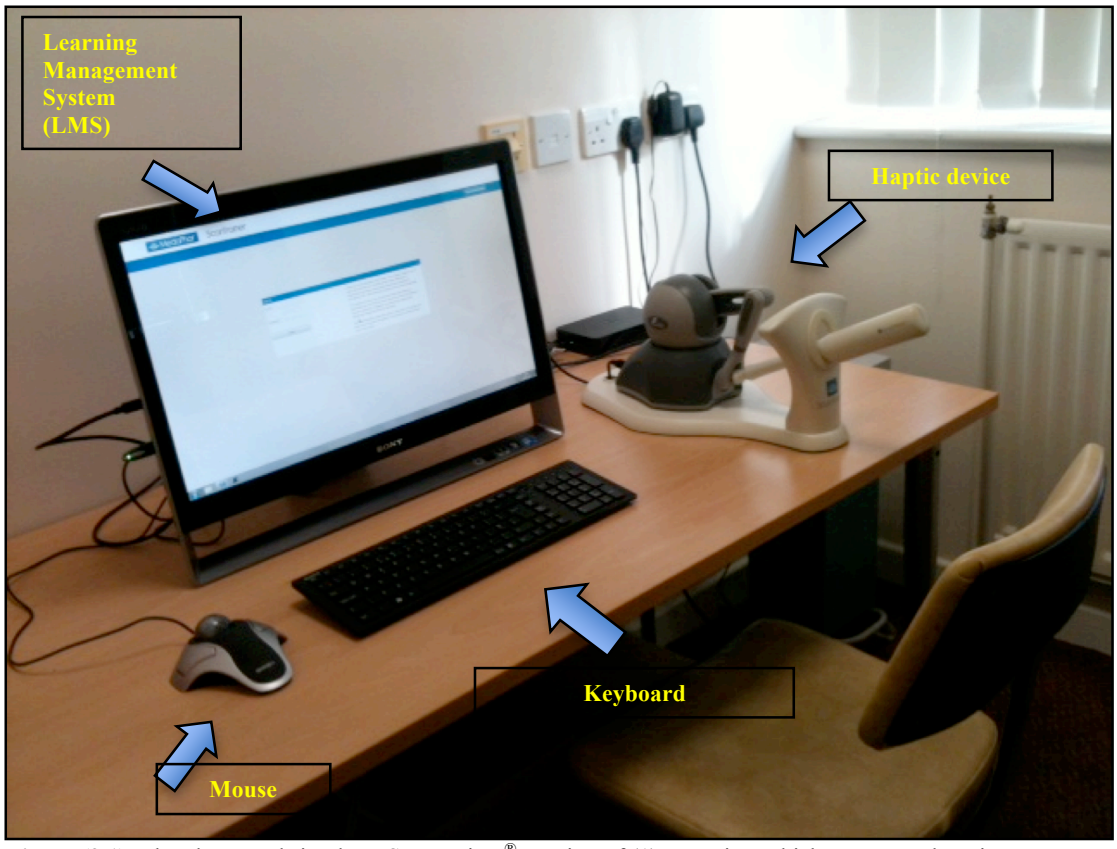


Figure (2.1): The ultrasound simulator ScanTrainer® consists of (1) a monitor which represents learning contents as programmed by specific learning software, and connects to (2) a haptic device, (3) mouse and (4) keyboard.

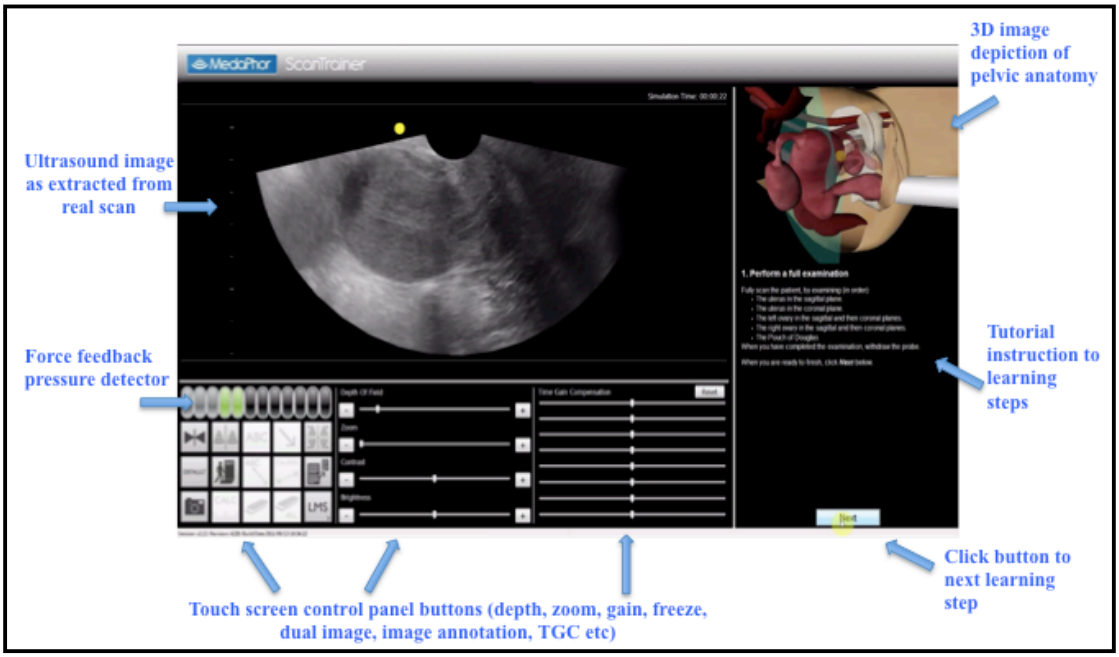


Figure (2.2): Learning contents as (it) operated by the ultrasound simulator ScanTrainer®

2.7.1 Learning tutorials and assignments

The simulation learning system consists of core and advanced skills modules. Each module has one or more tutorials, each of which is subdivided into one or more assignments. Each assignment consists of a series of tasks of basic or advanced TVUS skills in obstetrics and gynaecology (Figure 2.3). The modules are designed to create a structured learning process of transvaginal ultrasound skills through a series of step-wise instructions. The modules used for assessing trainees' acquisition of TVUS skills and performance in this thesis included core ultrasound skills only. The gynaecology modules consist of normal uterus (GYN1) and retroverted uterus (GYN2) assignments, and the obstetrics module is an early pregnancy (11-weeks) assignment. Each of these core skills modules had a final assignment called "full examination" that assessed the acquisition of seven basic ultrasound skills. These skills are listed in Table 2.1.

	Skill	Skill description
GYN1	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas (POD)
GYN 2	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas (POD)
Early pregnancy	**Skill 1**	Examining gestational sac (GS) in sagittal plane
	Skill 2	Examining fetal heart activity
	Skill 3	Examining fetus in sagittal plane
	Skill 4	Viewing yolk sac (YS)
	Skill 5	Labelling yolk sac (YS)
	Skill 6	Optimise image in viewing yolk sac (YS)
	Skill 7	Examining placenta in sagittal plane

Table (2.1): Description of seven skills listed in the checklist for the three assignments (GYN1, GYN2 and early pregnancy)

Figure (2.3): Learning resources of tasks and assignments with computer-generated individualised trainee(s) feedback.

2.7.2 Haptic device

The haptic device enables the user to interact with and modify virtual objects by probe manipulation and provides force feedback sensation that is closer to a real scan. The haptic instrument includes: (1) the base of the device, (2) probe stand, (3) the probe, and (4) the haptic tool which is connected to the computer, Figure (2.4). The haptics most frequently seen in medical simulations and training use technology known as proprioceptive "force feedback," where users hold a tool that pushes back on the user's hand when it makes contact with virtual objects. This haptic technology has developed rapidly in the last decade (Hayward et al., 2004). The basic purpose of the haptic device is to transfer the position of the tool's tip to the computer, and to supply 3D and visual images, including force feedback, on the computer. The ultrasound simulator ScanTrainer® uses a SensAble Technologies PHANTOM® haptic device to provide a realistic, touch-enabled training experience. This training method enables the trainees or students to develop a complex mix of cognitive skills and hand-eye movement coordination without the need for an ultrasound machine, or a patient, or direct supervision by an expert (ScanTrainer®, MedaPhor plc, Cardiff, UK).

The haptic device should be handled with care and the probe is held properly in order to apply minimum pressure only to avoid overheating. This could prevent the generation of realistic scans or accurate feedback. In addition, the “dock and re-dock” of the probe should always be applied with care and should be placed properly into the haptic body (Figure 2.5). The proper practice of handling the probe with the haptic device was clarified to all users as an important aspect of maintenance and to reduce systematic errors that may affect the veracity and accuracy of scans and results.

Figure (2.4): Haptic device

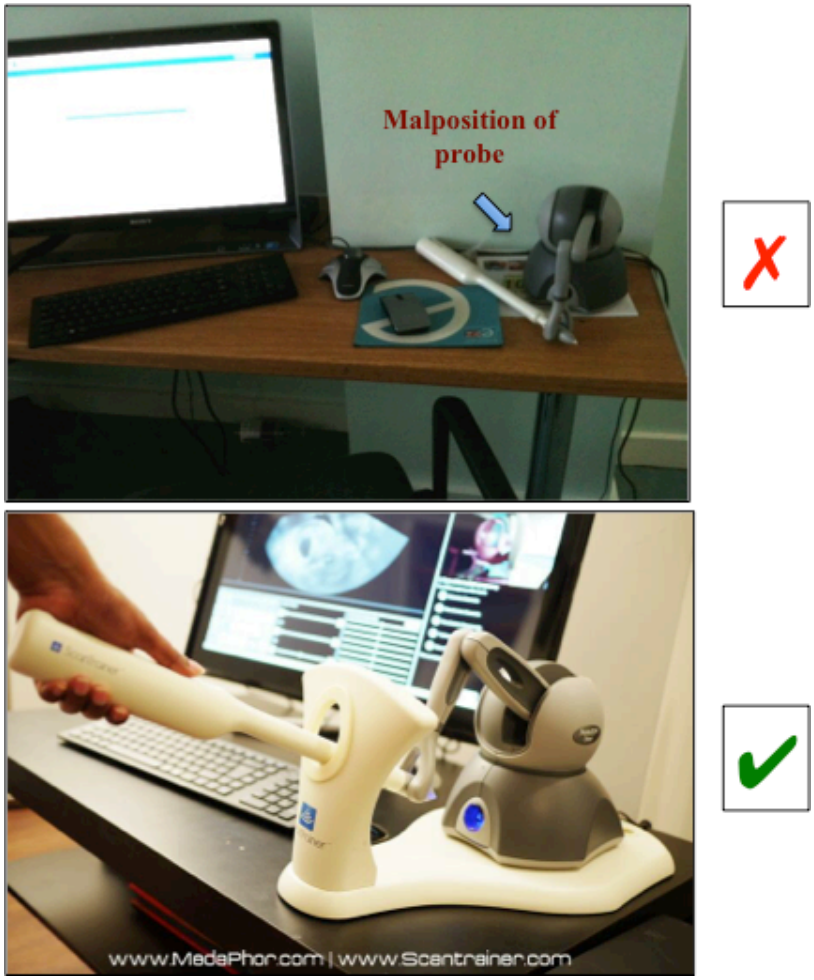


Figure (2.5): The haptic device includes: (1) the base of the device, (2) probe stand, (3) the probe, and (4) the haptic tool which connects to the computer. This figure shows the method of 'dock and re-dock' the probe into and out of the haptic body.

Chapter three

Face and content validity of the virtual reality simulator ScanTrainer®

CHAPTER 3

Face and Content Validity of the Virtual Reality Simulator "ScanTrainer®"

Simulation tools are either simplistic models or complex applications and regardless of the technology used, a simulator must demonstrate validity to be an effective education tool (Weidenbach et al., 2009). This entails gathering evidence from multiple sources to show that the interpretation of image, examination or assessment is sound and sensible (Weidenbach et al., 2009; Markowitz et al., 2011). At the outset, validation will usually attempt to confirm the fundamental reasons that these tools need to exist for learning (Carter et al., 2005; McDougall et al., 2006; Gilliam and Acton, 2007; Wilfong et al., 2011). From an educational perspective, a simulated performance should appear realistic when creating a cognitive-sensory mechanism known as 'sense of presence' because it allows the trainee/operator to interact with the remote environment as if s/he were present in it (Aiello et al., 2012). With regard to the role of simulation in developing ultrasound knowledge and skills, the validity and reliability of a simulator system for educational goals must be proven, through structured face, content and construct validity studies (Weidenbach et al., 2009; Wright et al., 2013; Madsen et al., 2014; Tolsgaard et al., 2015a).

Face validity is defined as the extent of a simulator's realism and appropriateness when compared to the actual task (Byrne and Greaves, 2001; Hung et al., 2011; Alzahrani et al., 2013). Whereas content validity is defined as the extent to which a simulator's content is representative of the knowledge or skills that have to be learnt in the real environment. This is based on detailed examination of the learning resources, tutorials and tasks (Carter et al., 2005; Nicholson et al., 2006; Schreuder et al., 2009; Xiao et al., 2014). Hence, in the context of ultrasound, face validity

addresses the question of how realistic is the simulator, for example, in examining the female pelvis and how realistic is the simulated feel (haptic sensation) experienced during the examination. Similarly, content validity addresses the question of how useful is the ultrasound simulator in learning relevant skills such as measuring endometrial thickness, fetal biometry (Seixas-Mikelus et al., 2011; Dulan et al., 2012; Alzahrani et al., 2013).

According to McDougall and colleagues (2006), Kenney and colleagues (2009) and Xiao and colleagues (2014), face validity is expressed as the assessment of virtual realism by novices while content validity refers to experts' assessment of the suitability of a simulator as a teaching tool. However, reports in the literature are diverse and some authors undertake face validity of a simulator by seeking the opinion of any user including expert and non-expert subjects (Verdaasdonk et al., 2006; Seixas-Mikelus et al., 2010; Hung et al., 2011; Kelly et al., 2012; Schreuder et al., 2009; Alzahrani et al., 2013). Others (Schijven and Jakimowicz, 2002; Sweet et al., 2004; Maithel et al., 2006; Dulan et al., 2012; Aydin et al., 2014), have argued that subjects' experience is required for face validity of any educational instrument. With regard to content validity, it widely refers to experts' judgement towards the learning content and tasks of a simulator (Fisher et al., 2006; Nicholson et al, 2006; Scott et al., 2008; Gould, 2010; Seixas-Mikelus et al., 2011). Nevertheless, many published studies rely on subjects with different levels of experience in evaluating content validity of a simulator (Vick et al., 2007; Gavazzi et al., 2011; Hung et al., 2011; Kelly et al., 2012; Alzahrani et al., 2013; Schreuder et al., 2014).

The ultrasound simulator (ScanTrainer®, MedaPhor plc, Cardiff, Wales, UK) enables the student to acquire transabdominal (TAS) or transvaginal ultrasound scanning (TVUS) skills through a series of simulation tutorials, each with one or more assignments that include specified tasks reflecting real ultrasound practice. Upon completion of the tasks, the simulator provides computer-generated individualised student/trainee feedback. The hypothesis was that the simulator was (1) realistic for the purpose of developing ultrasound skills and reflects real life scanning, and (2) the content of its structured learning approach represents the knowledge and psychomotor skills that must be learnt when scanning patients.

3.1 Aims and Objectives

The aim of this study was to determine face and content validity of TVUS ScanTrainer®. The objectives were; (1) to recruit practitioners with varying levels of ultrasound experience from attendees of an international conference, and (2) for study volunteers to undertake relevant simulator tutorials and complete a structured questionnaire including statements on face and content validity.

3.2 Subjects and method

Subjects were voluntarily recruited from delegates visiting the "ESGE Simulation Island" during the 23rd European Congress of Obstetrics and Gynaecology in Glasgow, Scotland, UK. Each delegate was given a brief, general introduction on the purpose of the study and instructions on how to use the simulator and the relevant tutorials. They gave verbal consent to participate and proceeded to explore specific tasks in three tutorials with the TVUS ScanTrainer®. These were (1) core skills gynaecology which has assignments on assessing the uterus, ovaries & adnexa and measuring the endometrial thickness, (2) core skills early pregnancy which has assignments on assessing the gestational sac, yolk sac as well as evaluating fetal viability and measurements, and (3) advanced skills that consisted of several case studies e.g. ovarian cyst, ectopic pregnancy and twin pregnancy. At the conclusion of the session, subjects completed a short questionnaire.

The structured questionnaire (Appendix 3.1) consisted of two sections; one detailed subjects' demographic information, previous ultrasound experience and any previous experience with VR simulation or ultrasound mannequins. The other section included simulation-related statements. An expert was defined as a subject who had ultrasonography experience of more than two years, conducted daily scanning sessions and considered her/himself as an independent practitioner. A non-expert was defined as having limited experience with ultrasound, had less than two years ultrasound experience, had very limited scanning sessions e.g. once/month or occasionally or considered her/himself as a trainee under supervision.

Fourteen simulation-related statements/parameters were subjectively scored along a 10 cm visual analogue scale (VAS) line by marking the point that subjects felt most appropriate, with (0) at one end (very bad) and (10) at the other (very good).

Statements 1 to 6 assessed face validity, 7 to 12 evaluated the simulator's learning content and 13 and 14 were general statements on the value of the simulator as training and testing tool. Ratings on the scale (10 cm which was equalised to 100 mm) were defined in "mm" as; 0-9 (very strongly disagree), 10-19 (strongly disagree), 20-29 (disagree), 30-39 (moderately disagree), 40-49 (mildly disagree), 50 (undecided), 51-59 (mildly agree), 60-69 (moderately agree), 70-79 (agree), 80-89 (strongly agree), 90-100 (very strongly agree).

The study was conducted in accordance with the general terms and conditions of the South East Wales Research Ethics Committee SEWREC (NHS REC Reference 10/WSE02/75) approval and following approval of the study protocol by the congress organising committee (Appendix I).

3.3 Statistical data analysis

IBM SPSS Statistics software version 20.0 was used for statistical analysis. Median values were chosen in preference to mean values as the data set were not normally distributed. Median scores and box plots were constructed for each statement as rated by non-experts and experts. Face validity and general statements data were stratified by expert and non-expert status, while content validity data were reported for experts only. Differences between experts and non-experts ratings were analysed using the Mann-Whitney U test where the significance indicates p-value ≤ 0.05.

3.4 Results

3.4.1 Demographics

Thirty-six subjects, 24 females (67%) and 12 males (33%) participated in this pilot study. Nine were UK-based and twenty-seven were based in other European countries. Eleven subjects (31% - expert group) rated themselves as skilled with more than two years' experience and practiced independently (n=10) or with one to two years' experience and had daily ultrasound sessions (n=1). Twenty-five subjects (69% - non-expert group) were trainees under supervision and included two with more than two years' experience but had occasional scanning sessions. Median age for expert group was 51 years (range 32-67) and 31 years (range 25-39) for non-expert group. The median ultrasound experience for experts was more than two years and for non-

experts was in the category "six to eleven months". Further breakdown of demographics and years of ultrasound experience are detailed in table (3.1).

3.4.2 Assessment of face validity

Median scores of face validity statements are detailed in table 3.2. In summary, experts and non-experts' ratings ranged between 7.5 and 9.0 and were slightly higher by experts in two statements (2&6) relating to; "*realism of the simulator to simulate the transvaginal scan of female pelvis* and *realism of the simulator to provide actual action of all buttons provided in the control panel"*. Two statements (1&3) were rated lower by experts and related to; *"relevance of the simulator for actual transvaginal ultrasound scanning* and *the realism of the simulator to simulate the movements possibly required to perform in the female pelvic anatomy (uterus, ovaries/adnexa, Pouch of Douglas POD)"*. The remaining two statements (4&5) referring to; *"realism of the ultrasound image generated during the performance* and *force feedback provided on the operator's hand to simulate real scan"* were equally rated. Two general statements (13&14) were also rated lower by experts. However, there were no statistically significant differences between the two groups' ratings in all statements (Table 3.1). Median values and box-plots of the eight statements in the two groups are shown in chart 3.1 and 3.2.

3.4.3 Assessment of content validity

Experts' median scores of content validity statements ranged from 8.4 to 9.0 and are detailed in table 3.3. Median values and box-plots of the six statements are shown in charts 3.3.

3.5 Discussion

In this pilot study, the ScanTrainer® simulator demonstrated high face and content validity and its overall value as a training and testing tool received high ratings as well. To accurately measure participants' level of agreement with relevant statements, VAS method was used in the questionnaire (Jensen et al., 2003). Higher ratings given by non-experts than experts with regard to relevance of the simulator to actual TVUS and its realism to simulate the movements required to perform in the examination of the female pelvis (statements 1 & 3) highlight the fact that such realism is crucial for

non-experts for several reasons. This may be because experts need to develop greater understanding of the strengths and limitations of the simulator compared to trainees (Shanmugan et al., 2014). Alternatively, beginners in the early stages of learning ultrasound skills are able to address their learning needs through simulated learning compared to the experts who expect variety and advanced or more complex performance rather than basic tutorials (Hung et al., 2011).

To the best of the researcher's knowledge, no comparable face and content validity studies addressing virtual reality simulators for TVUS in obstetrics and gynaecology have been published in the literature at the time of submission (10th March 2016). In a face validity study of the dVT robotic surgery simulator (Schreuder et al., 2014), experts rated the simulator as less useful for training experts than for students/juniors and pointed out to the experts' need for more critical and advanced procedures in gynaecological surgery and that simulators specifically designed for learning basic skills are less preferable to experts. Creating simulated scenarios to correspond to real ones is always a challenge (Carter et al., 2005; O'Leary et al., 2008; Gould, 2010; de Vries et al., 2016).

Experts' ratings were higher for two statements relating to the realism of the simulator to simulate the transvaginal scan of a female pelvis and in providing actual action of all buttons in the control panel (statements 2 & 6) This may stem from non-experts' limited knowledge and experience, or they might not be familiar with the measurement possibilities of virtual simulators (Schijven et al., 2002 and Verdaasdonk et al., 2006). Similarly, Weidenbach and colleagues (2009) argued that experts gave a better grading for the realism of the EchoCom echocardiography simulator because they were not distracted to drawbacks such as manikin size and its surface properties, which were harder and more slippery than human skin and that experts scanned more instinctively. The author noted that this mental flexibility seemed to be as yet underdeveloped in beginners.

Non-experts and experts' ratings were similar when evaluating the realism of the ultrasound image generated during the performance and the force feedback provided onto the operator's hand (statements 4 & 5). Force feedback (haptics) scored 7.5 out of 10, the lowest score in this study. Similar to this study, Chalasani and colleagues

(2011) reported low face validity ratings for the haptic force-feedback device of a transrectal ultrasound TRUS-guided prostatic biopsy virtual reality simulator (experts' lifelike rating 64%, and novices' 67%) even though, the author pointed out that haptics, often very difficult to replicate in a simulator environment, were realistic. Haptics will not replace the real-patient scan experience but should enhance the learning approach and improve self-confidence. A further factor is that the ScanTrainer's haptic device can be tailored to three force feedback levels; normal resistance (most realistic), reduced and minimal (lowest) designed to avoid overheating during heavy use and it's likely that a lower force feedback setting might have contributed to the lower scores.

The role of force feedback in laparoscopic surgery is not clear (Verdaasdonk et al., 2006). Improving the realism of the simulator and its anatomical structures increase costs considerably due to increased demands for more complex hardware and software. In contrast, Lin and colleagues (2014) encouraged learning of bone-sawing skills with simulators that provide force feedback rather than not, confirming the importance of force feedback when seeking to enhance hand-eye coordination. With regard to ScanTrainer®, virtual ultrasound and haptics are used instead of a mannequin allowing measurement of the force applied to the probe and provide a somewhat realistic force-feedback during scanning. However, it still has the limitation of allowing a lower range of movements to the probe while lacking a simulated environment exemplified by the absence of a physical mannequin (Chalouhi et al., 2015a).

There are numerous simulator systems in usage particularly in the fields of laparoscopy and endoscopy (Chalasani et al., 2011) and several authors emphasised the importance of evaluating their content, including reviewing each learning task and assessing its overall value to determine whether it is appropriate for the test and whether the test contains several steps and skills for practice (Seixas-Mikelus et al., 2011; Hung et al., 2011; Gavazzi et al., 2011). In this study, experts' data were used to assess content validity. They had adequate time to review the simulator's learning resources, help functionality "ScanTutor", read the task-specific instructions, and undertook specified tasks before going on to the next step in the same tutorial. In addition, participants had the opportunity to review feedback on their performance in

the respective tasks. The results of this study demonstrated that the simulator's content and metrics were appropriate and relevant for ultrasound practice.

There are a number of published content validity studies in ultrasound simulation such as, the educational curriculum for ultrasonic propulsion to treat urinary tract calculi (Hsi et al., 2014), web-based assessment of the extended focused assessment sonography in trauma EFAST (Markowitz et al., 2011), and validating the objective structured assessment of technical skills for duplex assessment of arterial stenosis DUOSATS (Jaffer et al., 2014) which not based on virtual reality simulator devices. Shumard and colleagues (2015) reported on face and content validity of a novel second trimester uterine evacuation task trainer designed to train doctors to perform simulated dilatation and evacuation under ultrasound guidance. Although all respondents were residents with limited ultrasound experience, they rated the task trainer as excellent.

Other studies evaluated the effectiveness of simulation-based training in obstetrics and gynaecology ultrasound, whether to investigate the construct validity of a simulator system (Maul et al., 2004; Merz E, 2006; Chalouhi et al., 2015b; Madsen et al., 2014) or comparing simulation training to conventional methods such as theoretical lectures and hands-on training on patients (Williams et al., 2013; Tolsgaard et al., 2015a).

Feedback that is automatically generated immediately after a practical simulator session should enhance trainees' knowledge and ability to reflect critically on their performance and improve their skills (Cline et al., 2008). However, the big challenge is to determine how accurate, realistic and trusted the feedback is and thus, should also be validated appropriately.

Validation studies at national scientific meetings have been reported previously (Maithel et al., 2006, Stefanidis et al., 2007). They offer researchers a rich environment where subjects from different backgrounds and levels of experience are present in one place at the same time. A potential limitation of the study is that it did not determine in advance the sample size required to obtain a reliable result for face and content validation. In a study validating robotic simulator performance, a sample

size of six participants in each of the expert and pure novice groups was deemed adequate to achieve significance at 80% statistical power on the basis of available literature data (Hung et al., 2011). However, this was in contrast to Alzahrani and colleagues' (2013) pilot study which validated de Vinci Surgical Skills Simulator (dVSSS), as having six experts only was regarded as one of the study's limitations. The number of subjects in this study was higher and the findings are consistent with others (Gavazzi et al., 2011; White et al., 2010; Dulan et al., 2012; Kelly et al., 2012). In addition, many face and content validity studies of simulators were based on smaller sample size compared to the current study (Vick et al., 2007; O'Leary et al., 2008; Kenney et al., 2009; Bright et al., 2012; Shetty et al., 2012; Alzahrani et al., 2013). A larger number of participants in this study might have improved the power of its results (Markowitz et al., 2011). Participants in this study were from different UK and European institutions unlike others who were from single academic institution (Hsi et al., 2014), thus it may be more widely generalisable.

In summary, this study confirms that ScanTrainer® simulator has the feel and look (face validity) and tutorial structure (content validity) to be realistic and relevant for actual TVUS scanning. This study also concurs with the notion that advancing computer technologies have been able to incorporate virtual reality into training to facilitate the practice of basic skills as well as complex procedures that leave little room for error or mistake (Sweet et al., 2004; Carter et al., 2005; Verdaasdonk et al., 2006; Tolsgaard et al., 2015b). Equally, such simulators should be subject to ongoing validation to address trainees' learning needs and improve patient care and safety (Vick et al., 2007; O'Leary et al., 2008; Gavazzi et al., 2011).

Chapter 3: Tables and figures

Table (3.1): Participants' demographics and ultrasonography experience

	Non-expert	**Expert**
No of participants (n=36)	25(69%)	11(31%)
Gender		
Female	17(68%)	7(64%)
Male	8(32%)	4(36%)
Country of practice		
Within UK	6(24%)	3(27%)
Outside UK	19(75%)	8(73%)
Speciality		
Consultant	0	3(27%)
Speciality	2(8%)	4(36%)
Specialist trainee	20(80%)	3(27%) ST7
Medical student	1(4%)	0
Other	2(8%)	1(10%)
Median age	31(25-39)	51(32-67)
Years of ultrasound experience		
Never	3(12%)	-
<6months	5(20%)	-
6-11months	9(36%)	-
1-2 yrs	6(24%)	1(10%)
>2 yrs	2(8%)	10(90%)
Transvaginal ultrasound experience		
Independent practitioner	2(8%)	11(100%)
Trainee under supervision	23(92%)	-
Ultrasound sessions		
Never	4(16%)	0
Daily	1(4%)	5(46%)
Once/week	9(36%)	0
Once/month	3(12%)	2(18%)
Occasionally	5(20%)	2(18%)
Other	3(12%)	2(18%)
Previous experience with the ScanTrainer®		
Yes	3(12%)	3(27%)
No	22(88%)	8(73%)
Previous experience with ultrasound model i.e. blue Phantom™		
Yes	4(16%)	4(36%)
No	21(84%)	7(64%)

Table (3.2): Face validity "median scores" ratings by experts and non-experts (n=36)

Face validity statements	Median score			
	Expert (n=11)	**Non-expert (n=25)**	**Overall**	**p-value**
Statement 1: Relevance of the simulator for actual transvaginal ultrasound scanning	7.5(5.0-10)	9.0(7.0-10)	8.7(5.0-10)	0.1
Statement 2: Realism of the simulator to simulate the transvaginal scan of female pelvis	8.3(5.0-10)	8.0(5.9-10)	8.1(5.0-10)	0.9
Statement 3: Realism of the simulator to simulate the movements possibly required to perform in the female pelvic anatomy (uterus, ovaries/adnexa, POD)	7.7(1.0-10)	9.0(5.0-10)	9.0(1.0-10)	0.1
Statement 4: Realism of the ultrasound image generated during the performance	9.0(1.3-9.8)	9.0(6.0-10)	9.0(1.3-10)	0.2
Statement 5: Force feedback provided on the operator's hand to simulate real scan	7.5(3.0-9.5)	7.5(2.7-10)	7.5(2.7-10)	0.4
Statement 6: Realism of simulator to provide actual action of all buttons provided in the control panel	9.0(1.0-10)	8.7(3.0-10)	9.0(1.0-10)	0.5
General statements				
Statement 13: Overall value of the simulator as a training tool	9.0(5.0-10)	9.3(6.0-10)	9.0(5.0-10)	0.2
Statement 14: Overall value of the simulator as a testing tool	9.0(5.0-10)	9.5(5.6-10)	9.3(5.0-10)	0.2

Table (3.3): Content validity "median scores" ratings by experts (n=11).

Content validity statements	Expert (n=11)
Statement 7: Realism of the simulator to provide the endometrial thickness measurement in gynaecology task	8.6(3.5-10)
Statement 8: Realism of the simulator to provide measurements of the ovary in gynaecology task	8.7(4.5-10)
Statement 9: Ability to test normal gynaecological anatomy: uterus, adnexa and Pouch of Douglas	8.4(4.7-10)
Statement 10: Ability to test early pregnancy structures: fetus, viability and placenta	9.0(5.0-10)
Statement 11: Realism of the simulator to provide the CRL measurement in early pregnancy task	9.0(4.7-10)
Statement 12: Relevance of the simulator's learning resource, videos and ScanTutor function	8.7(5.0-10)

Chart (3.1): Box-plots represented median, first and third quartiles, minimum, maximum and outliers of scores obtained by expert and non-expert ratings of the six face validity statements.

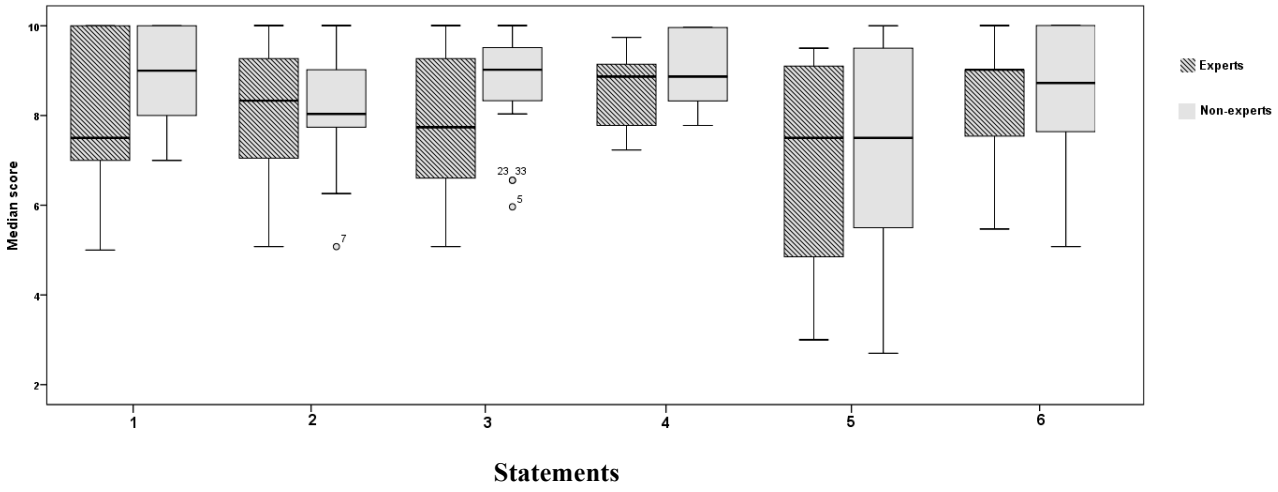


Chart (3.2): Box-plots represented median, first and third quartiles, minimum, maximum and outliers of scores obtained by expert and non-expert ratings of the two general validity statements on the simulator as a training and testing tool.

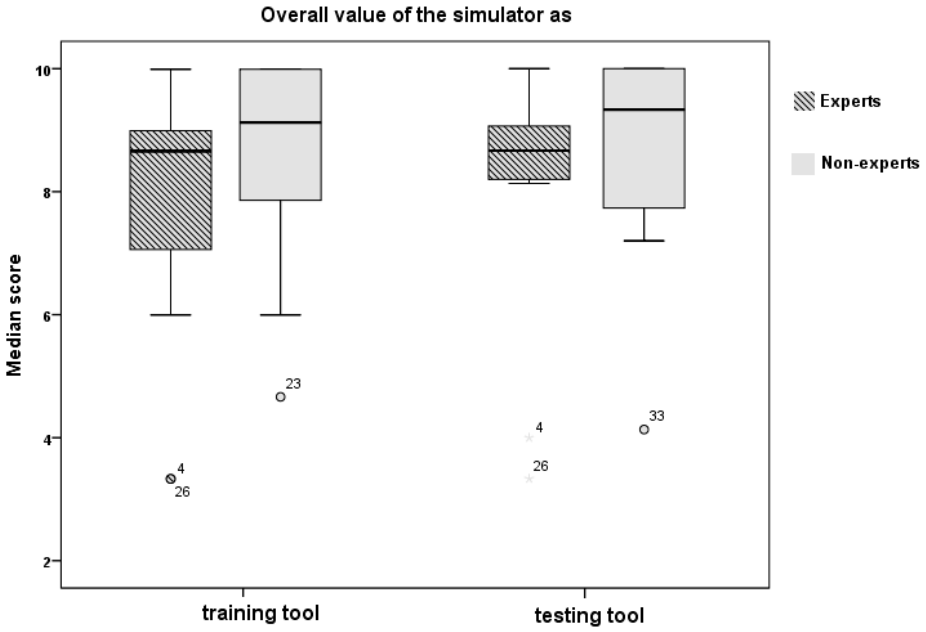
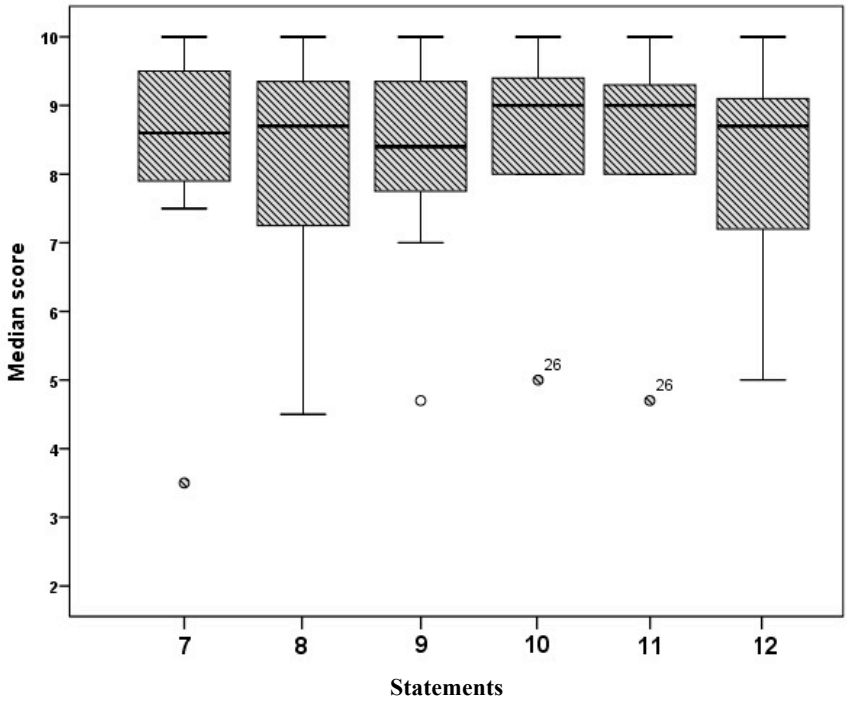


Chart (3.3): Box-plots represented median, first and third quartiles, minimum, maximum and outliers of scores obtained by experts' rating the six content validity statements.



Chapter four

Intra- and inter-observer reliability of scoring systems for ultrasound skills assessment

CHAPTER 4

Intra- and inter-observer reliability of scoring systems for ultrasound skills assessment

---

Medical educators employ different evaluative approaches to assess the performance of medical trainees such as the use of scoring systems based on structured checklists and/or global rating scales GRS (Kim et al., 2009; Scott et al., 2000). Evaluation of ultrasound skills in obstetrics and gynaecology using different scoring systems was also reported previously (Alsalamah et al., 2009; Tolsgaard et al., 2013). However, it is critical to validate such scoring systems within and between evaluators (raters), test the level of agreement to determine the appropriateness of the assessment method. Video-recordings have been frequently used to evaluate trainees' clinical skills in several branches of medicine such as nursing education, surgery, laparoscopy, obstetrics, gynaecology, emergency medicine and anaesthesia (Beard et al., 2005; Edwards and Ragaratnam, 2009; Beyer et al., 2011; Cash et al., 2012).

Graham et al (2012) define observer agreement (inter- or intra-) as the degree to which two or more observers give the same score to an identical observed performance using the same rating scale i.e. checklist or GRS. The measure of agreement is intra-class coefficient correlation (ICC) - Absolute Agreement. Observer agreement (intra- or inter-) is often confused with observer reliability (inter- or intra-) which refers to the similarity in the ranking of scores made by two or more observers. Measures of reliability used in this chapter include intra class coefficient correlation (ICC) - Consistency and Cohen's Kappa, which account for the possibility that observers actually guess on at least some variables due to uncertainty and a number of these would be congruent. Factors that decrease the variability between observers and provide highly reliable outcomes should be identified and considered prior to the evaluation process. These include; selection and training of the observers and identification and calibration of criterion-references for assessment (Bakker N, 2008;

Graham et al., 2012). This evaluation process should help in the establishment of a systematic approach in reviewing recorded videos. Applying these principles in evaluating acquisition of ultrasound skills in obstetrics and gynaecology would enhance our understanding of the learning process. At the time of submission of the thesis (10 March 2016), there were no published manuscripts validating scoring systems assessing TVUS using recorded videos from an ultrasound simulator. The information generated in this study would help in determine the most appropriate scoring system to be used in evaluating trainees' performance of ultrasound skills in a randomised controlled trial (Chapter 7).

4.1 Aims and objectives

The aim of this study was to test the reliability of scoring systems developed for the assessment of obstetrics and gynaecology ultrasound skills. The study objectives were to use scored video-recordings of ultrasound scans to (1) determine intra-observer (test and re-test) absolute agreement of the scoring systems for each independent observer individually, (2) determine inter-observer reliability between two independent observers' ratings to evaluate the consistency of two scorings and (3) test the level of agreement between the checklists and GRS scores of the two observers.

4.2 Subjects and methods

This was an observational experiment, partial (test and re-test) research design. The sample material was video recordings of participants in a randomised controlled trial (RCT) undertaking TVUS procedures in gynaecology and early pregnancy using a virtual reality simulator. The study was conducted in accordance with the general terms and conditions of the South East Wales Research Ethics Committee SEWREC approval (NHS REC Reference 10/WSE02/75).

4.2.1 The tool used to record videos

The ultrasound simulator ScanTrainer® was the tool used to assess and record the participants' performance of predetermined tasks during a randomised controlled trial over a period of six assessment sessions. Tasks included in the assessments were two from gynaecology modules (anteverted uterus GYN1, and retroverted uterus GYN2), and one from an obstetrics module (early pregnancy).

4.2.2 Editing of recorded videos

A free and multifunctional screen recorder software "BB FlashBack Express version 2.8.2" that was installed in the ScanTrainer® ultrasound simulator was used to make all recordings. At each individual session, and specifically when the participant begins the assignment in the simulator, the researcher turns on the video recorder to start capturing a full screen visual overlay including the trainee's name, time, date, assignment as well as all notes taken throughout the performance until the video is terminated. The length of each video ranged from three to seven minutes. Each recorded videos was coded specifically as

traineeID_assignment_code_session_number_date and saved in the simulator. Each video selected for this study was anonymised professionally to conceal participant identifiers and the automated feedback generated by the simulator's "Learning Management System" (LMS) metrics for blind rating. The videos were saved in flash or AVI formats and were easy to review using movie player application with the facility to review images frame-by-frame. The videos were accessible for the addition of notes at the time of recording and for replaying any media player application away from the simulator. Figure (4.1) shows an example of an anonymised video recording obtained from the ultrasound simulator with the anatomy of a female pelvis in concealed (4.1a) and revealed (4.1b) mode.

4.2.3 Selection of observers

Observers were determined prior to the study, both of whom were independent practitioners with experience in obstetrics and gynaecology ultrasonography. The first (observer 1) is the PhD researcher (A.A) and the other (D.A, observer 2) is a specialist with a Masters degree in obstetrics and gynaecology ultrasound.

(A.A.)= Amal Alsalamah, (D.A.)= Dina Albalushi

Observer 1 and 2 underwent a brief training session before the start of the study to familiarise themselves with the scoring systems. Each reviewed and scored six randomly selected videos representing good and poor performance from the three assignments. After which they compared their checklists and GRS scores. Disagreements were resolved through discussion to reach a consensus with regard to what constituted unsatisfactory, borderline or satisfactory performance. Observer 1 rating was blinded to that of observer 2. In the inter-observer reliability study, video

recordings were viewed only twice; the first time to complete the checklist scoring and the second time to evaluate the performance using the GRS. Exceptionally, further reviews were allowed, as this was occasionally essential to ensure a comprehensive assessment.

4.2.4 Checklist and Global Rating Scale (GRS)

For each assignment, there was a specific checklist and a GRS, which were used to assess and score participants' performance. The checklist, developed previously for the applicant's MSc project, was based on a hierarchical task analysis of a transvaginal ultrasound scan procedure (Alsalamah et al., 2009) and was supplemented by additional skills based on published Royal College of Obstetrics and Gynaecology (RCOG) Objective Structured Assessment of Technical Skills (OSATS) scores. These are currently used as a formative assessment tool of essential clinical skills and competencies necessary for learning ultrasound skills in obstetrics and gynaecology (Salvesen et al., 2010). Seven skills were assessed in the checklist of gynaecology assignments and eight early pregnancy one, in which there are the scorings were: pass (1), fail (0) or (N/A). The GRS assessment was based on basic skills listed in the checklist with addition of a parameter assessing a systematic approach to the ultrasound scan. The assessment scale ranged from "not attempted" (NA=0), 1 (very poor) to 5 (excellent) for each performance.

The checklist parameters are described in Table (4.1) and were based on pass/fail or not attempted (NA) outcomes and judged according to the appropriateness of hand movement during each skill. The GRS (Table 4.2) was criteria based and additionally, assessed the systematic approach to scanning (Appendices 4.1 and 4.2).

4.3 Inter- and intra-observer reliability and agreement

Intra-observer reliability was tested by each observer (AA and DA) blindly rating ten randomly selected video recordings five times. The first two ratings took place on the same day six hour apart, the third and fourth ratings took place on the second day on the same basis. The fifth rating was undertaken on the third day with at least a 24-hour gap to avoid any recall of the rating. Inter-observer reliability assessed the level of agreement and reliability between the two observers based on their independently scored video recordings using the checklist and GRS.

4.4 Statistical methods

IBM SPSS (Statistical Package of the Social Science) version 20.0 was used for data entry and analysis. Microsoft Excel 2010 used to generate random numbers for the selection of video recordings. For the inter-observer reliability and agreement study, a representative sample of 144 video recordings; intervention (n=72) and control (n=72) were randomly selected from a total of 1134 video recordings in the RCT (Chapter 7). For the intra-observer reliability study, the ten video recordings were randomly selected from the study sample of 144 videos.

Shapiro-Wilk test was used to test normality of data distribution of the checklist and the GRS. Kruskal-Wallis (for the checklists) and one-way ANOVA (for the GRS) were used for a test-retest statistical analysis with significance indicated at 0.05 of the five repetitions of each observer's scores. Test-retest absolute agreement of the five repetitions by each observer was tested with the intra-class correlation coefficient (ICC) and was interpreted as follows: <0.40 poor, 0.40–0.59 fair, 0.60-0.74 good and ≥0.75 excellent (Fleiss, 1986; Cicchetti, 1994, cited Hallgren K, 2012). The box-plot represented median values of five attempts of checklist and GRS for each observer independently.

Inter-observer agreement was tested with the percentage of absolute agreement, intra-class correlation coefficient (ICC) and inter-observer reliability was tested by Cohen's Kappa (Ķ) (Cohen, 1960 and Fleiss et al., 1991, cited Hallgren K, 2012). Interpretation of Cohen's kappa Ķ values was as follows; < 0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good, and >0.80 very good correlation (Altman, 1991, cited Hallgren K, 2012). The Bland–Altman plots represented the difference of mean values between two observers scores with checklist and GRS. Spearman and Pearson correlation coefficients r used to measure the degree of linear relationship "correlation" between scores obtained by the two observers for checklists and GRS. Spearman r used for non-parametric data (checklist) and Pearson r for parametric data (GRS). Correlation coefficient "r" ranges between -1.0 to +1.0 and the closer it is to +1 or to -1, the more closely the two observers' scorings are related. The generalisability of the variance components used to verify the sample of videos included in this study was representative of the RCT sample as a whole (Ping and Sconing, 2008). Reliability coefficient, which called generalisability G-coefficient

should range of 0 to 1 in order to consider the dependability of differences among individuals “raters, assignments, videos and variance errors” potentially applicable. Closer to 0 is more generalisable.

4.5 Results

4.5.1 Intra-observer agreement

Data in the intra-observer study were not normally distributed, Shapiro-Wilk test $p<0.05$. Test-retest of five repetitions showed no difference in mean scores given by any of the observers at each attempt, either using a checklist or GRS. Table (4.3) showed the overall result of median scores for individual skills across the five repetitions by each observer. The test and re-test Kruskal-Wallis findings indicated no statistically significant difference in the five repetitions outcome for each observer, p-value >0.05, (Table 4.4). Intra-observer agreement (ICC) scores for A.A. and D.A. were 0.80 and 0.72 for checklists and 0.80 and 0.71 for GRS respectively. The absolute agreement of ICC scores obtained by A.A. with checklists and GRS revealed an excellent agreement while good agreement was shown by D.A, (Table 4.5). The boxplots represent the median values across the five repetitions marked by each observer with checklist and GRS; the results of A.A. and D.A. are found in charts (4.1) and (4.2), respectively.

4.5.2 Inter-observer reliability and agreement

The checklist datasets for the three assignments were not normally distributed as tested by Shapiro-Wilk, $p=0.001$, while GRS datasets were normally distributed, $p=0.5$. To estimate variance components in the selected sample, generalisability coefficient for variables was ranged between 0-1 (Table 4.6). Hence, the generalisability effect of using 144 videos out of 1134 was considered to be an acceptable sample size and can be generally applied.

Table (4.7) showed an excellent inter- observer absolute agreement for checklist and GRS (ICC=0.96 and 0.97 respectively) with no statistically significant difference between the two observers’ checklist and GRS scores. Table (4.8) showed good to excellent inter-observer reliability for checklist (Cohen’s kappa Ķ); 0.83, 0.78 and 0.92 for GYN1, GYN2 and early pregnancy assignments respectively. However, the GRS results showed moderate inter-observer reliability GYN1 and GYN2, with Ķ

values 0.56 and 0.60 respectively, and good reliability in the early pregnancy assignment, Ƙ =0.69.

Estimates of inter-observer reliability of individual skills with checklists (Table 4.9) was lowest in (SK7) in GYN2, Ƙ value 0.50 (0.29–0.68) and GYN1, Ƙ= 0.58 (0.35–0.81). With reference to the other skills in the checklists in the gynaecology assignments, the majority of results ranged from good to excellent, Ƙ ≥0.61. There was excellent reliability among all seven skills in the checklists used with the early pregnancy assignment, Ƙ ≥ 0.80. The illustration of Bland-Altman (BA) plots implies a degree of correlation between the two observers' scores, which indicates positive correlation in checklist, Spearman r=0.91, 0.92, 0.92 for GYN1, GYN2 and early pregnancy assignments respectively (Chart 4.3).

Estimates of inter-observer reliability of individual skills in the GRS (Table 4.9) showed that GRS-SK8 has fair agreement in GYN1, Ƙ=0.39 (0.22–0.51), moderate in GYN2, Ƙ=0.50 (0.33–0.74) and very good in early pregnancy Ƙ=0.91 (0.79-1.61). In early pregnancy, moderate agreement was noted in GRS-SK2, Ƙ=0.54(0.33–0.76), while the other seven GRS skills were good to very good (Ƙ ≥ 0.61). In gynaecology assignments, there were two skills in each assignment that revealed good agreement compared with the other skills which revealed fair agreement only. The GRS result is illustrated with BA plots in chart (4.4) and showed positive linear correlation between the two observers with GRS scoring in the three assignments, Pearson r=0.84, 0.86, 0.99 for GYN1, GYN2 and early pregnancy assignments respectively. Despite these results in Ƙ values using checklists and GRS described in table (4.8), the ICC were likely to be correlated more positively with the checklist scoring system than with GRS scores.

4.6 Discussion

In this chapter, estimates of intra-observer agreement and inter-observer reliability for checklists, GRS and individual tasks are reported. The intra-observer absolute agreement estimates were high; good and excellent for the checklists and GRS respectively with no statistically significance difference in the five repetitions' scores. These findings indicate that with appropriate training and clear scoring criteria, observers are consistent in their scores and more so, it is possible that one attempt by

an individual is sufficient to score video recordings of transvaginal ultrasound scanning (TVUS).

Inter-observer absolute agreement was excellent for both checklist and GRS total scores (ICC>0.75) indicating that with appropriate prior training, different observer may exhibit independently consistent total scoring patterns. Similarly, estimates of reliability were good to excellent for the checklists, moderate for GYN1/2 GRS and good for EP GRS. These results imply that scoring by one individual may be a reliable reflection or indicator of skills' attainment.

The use of checklists or GRS in the assessment of ultrasound skills has been reported previously (Alsalamah et al., 2009; RCOG, 2016; Tolsgaard et al., 2014b), however the work by Tolsgaard varies considerably in the way it was developed or evaluated and hence its applicability to the wider ultrasound community is uncertain.

More detailed estimates of reliability for individual tasks in the checklists showed good or excellent ($K \geq 0.61$) inter-observer reliability for 19 out of 21 checklist skills across the three modules. This was not demonstrated in the GRS where it was fair in one skill, moderate in 12 of 24 skills. Good in ten and one was excellent. Some notable examples include moderate reliability achieved in skill 7- *identification of the pouch of Douglas (POD)* for GYN1/2 checklists and GRS. With limited information provided during the review of recorded videos and fixed POD structure, the observer was sometimes unable to judge fully and correctly the trainee's performance based on GRS. However, using the pass/fail checklist protocol, it was easier for the observer to recognise that the POD was correctly identified by the trainee. Other factors affecting GRS include failure to centralise the POD correctly or lack of accurate knowledge of its anatomical location. Skill 8- systematic approach was also more difficult to score and had fair to moderate GRS reliability in GYN1 and 2 respectively.

The use of global scoring systems, which evaluate the quality of performance, can be challenging to observers (Cremers et al., 2005). Graham et al (2012) speculate that this possible lack of agreement may be due to the level of observer experience and motivation to adhere to the evaluation process, rather than producing different sets of scores/ratings using different rating tools. Research has also shown that an observer's

pedagogical beliefs can influence that person's ability to use a rating system as intended and also might conflict with the underlying theoretical foundation of the evaluation system (Henry et al., 2010 and Tarara et al., 2014). Another potential cause of discrepancy between observers is the observer's degree of familiarity with the trainee/person who is being evaluated. Familiarity may encourage bias as noted by several authors (Bretz et al., 1992; Cremers et al., 2005; Kim et al., 2009; Cash et al., 2012; Oremus et al., 2012). Moreover, it has been reported that the use of recorded videos for the evaluation of trainee performance leads to less accurate evaluation than the use of direct observation of the trainees' performance (Graham et al., 2012). Direct contact with the trainees enabled them to explain, clarify and identify some issues during the examination. For instance, on the occasion of poor performance of a trainee, the correct identification of an ovary either right or left sided, became not clear to the observer (D.A) who was blindly evaluating the performance without awareness of trainees' lack of knowledge. Possible solutions to improve reliability include clear instructions given to the trainee at the onset and to add audio to the video recording, so the trainee explains her/his actions.

Henry et al (2010), Cash et al and Oremus et al (2012) pointed out the importance of maximising potential agreement between observers by prior training to develop a common understanding in order to apply the rating system as consistently as possible. In this study, prior training had a positive link with increased intra-observer agreement. A number of researchers suggested that observer training was important for consistent results (Cremers et al., 2005 and Oremus et al., 2012), while others (Cremers et al., 2005; Henry et al., 2010; Gulgin and Hoogenboom, 2014; Tarara et al., 2014; Yanes et al., 2016) found that even extensive training would not ensure that every observer agreed with the standards set or with other observers. The authors referred to the importance of considering the observer expertise to improve rating accuracy. In contrast, Haywood et al. (2004) demonstrated that high agreement among expert and student observers had no influence on results after the pilot phase training in which the source of disagreement between observers was identified. In this study, variation in the level of agreement in some tasks may have been due to observers' expertise, trainees' instruction prior to or during video recording sessions, or to technical aspects of video recording. One limitation was the use of video recordings that either revealed or did not reveal the anatomy of the female pelvis.

Those that did so enabled the observer to monitor the movement of the probe during the examination easily. However, this difficulty is unavoidable because video recordings of trainee performances with revealed and concealed anatomy were recorded for the randomised controlled study.

Designing a choice of rating scales for assessing ultrasound performance can present the need for a trade-off between observer agreement and reliability. Pass/fail checklists are likely to produce high rates of agreement because they give little room for comparison (Barry et al., 2013). Thus, in this project, checklists produced higher ratings agreement than the use of GRS. In contrast, Larsen et al (2008) reported that the GRS was more effective than yes/no-based checklists when using video recordings of laparoscopic gynaecological procedures to improve quality assurance. Several authors (Penny et al., 2000; Scott et al., 2000; Graham et al., 2012; Barry et al., 2013; Tolsgaard et al., 2013) had argued that despite the fact that the checklist is more objective and rules out partiality, the range of scores available in a GRS could improve reliability by allowing more variation in ratings but reducing the exact agreement on a particular score. The subjectivity of a GRS scoring system enables it to give more feedback on performance, whereas the checklist scoring system is limited to deciding if the performance is a pass or a fail. Given all the above factors, the checklist was considered to be more appropriate method and hence, used for assessing ultrasound performance in subsequent studies of this PhD research.

The ScanTrainer® simulator as a self-directed learning tool provides automated feedback on trainee performance and records this performance for later revision or for use for educational and teaching purposes. Thus, it aids trainees in drawing experiences that help in learning; facilitates continuing professional development; and assist trainers in monitoring and assessing their trainees. Several authors (Williams et al., 2013; Madsen et al., 2014; Tolsgraad et al., 2015b) have argued that its value went beyond being a training tool, to being a valid and reliable assessor for the evaluation of practitioners' competence in ultrasonography. The provision of an ongoing quality assurance platform through one trustworthy tool, make the use of video recordings of significant value to medical education. This study demonstrated high agreement and reliability between two observers, and thus one (A.A) was considered as a standard human judge against simulation automated feedback in subsequent studies.

Chapter 4: Tables and figures

Table (4.1): Checklist

Module	Ultrasound skills	Observer 1 Skill correctly done		Observer 2 Skill is correctly done	
		Yes	**No**	**Yes**	**No**
Gynaecology GYN1/GYN2	1. Uterus correctly examined in the sagittal plane				
	2. Uterus correctly examined in the coronal plane				
	3. Left ovary correctly examined in the sagittal plane				
	4. Left ovary correctly examined in the coronal plane				
	5. Right ovary correctly examined in the sagittal plane				
	6. Right ovary correctly examined in the coronal plane				
	7. Pouch of Douglas correctly examined the sagittal plane				
Early pregnancy	1. Gestational sac correctly examined in the sagittal plane				
	2. Fetal heart correctly examined in the sagittal plane				
	3. Fetus correctly examined in the sagittal plane				
	4. Labelling the Yolk sac				
	5. Yolk sac correctly viewed				
	6. Yolk sac correctly magnified				
	7. Placenta correctly examined in the sagittal plane				

Table (4.2): Global Rating Scale (GRS)

Module	**Performance general evaluation**	**NA 0**	**Very Poor 1**	**Poor 2**	**Fair 3**	**Good 4**	**Very good 5**
Gynaecology GYN1/GYN2	1. Uterus seen in sagittal plane						
	2. Uterus seen in transverse plane						
	3. Left ovary seen in sagittal plane						
	4. Left ovary seen in transverse plane						
	5. Right ovary seen in sagittal plane						
	6. Right ovary seen in transverse plane						
	7. POD visualised						
	8. Perform systematic scan						
Early pregnancy	1. Scan GS in sagittal plane						
	2. Confirm fetal viability						
	3. Scan fetus						
	4. Identify Yolk Sac (YS)						
	5. Correctly viewing YS						
	6. Correctly magnifying YS						
	7. Identify placenta						
	8. Perform systematic scan						

POD= Pouch of Douglas, GS = gestational sac, YS = yolk sac

Table (4.3): Intra-observer statistical significance Kruskal-Wallis test and median scores of checklist and GRS obtained by the two observers individually, in the three assignments; GYN1, GYN2 and early pregnancy

	Observer 1 (A.A.): Five repetitions						
		GYN1		GYN2		Early pregnancy	
		Median	p-value	Median	p-value	Median	p-value
Checklist	Skill1	1	1	1	1	1	1
	Skill2	1	1	1	1	1	1
	Skill3	1	0.9	1	1	0.5	1
	Skill4	1	1	1	1	1	1
	Skill5	1	0.9	1	1	0.5	1
	Skill6	1	0.9	1	0.9	1	1
	Skill7	1	0.9	0	0.9	0	1
GRS	Skill1	5	1	3	1	4	1
	Skill2	5	1	3	1	3.5	1
	Skill3	3	0.9	3	1	2.5	0.9
	Skill4	3	0.9	3	1	3	1
	Skill5	3	0.9	3	1	2.5	1
	Skill6	3	0.9	3	1	3.5	1
	Skill7	3	0.9	2	1	1	0.9
	Skill8	5	0.9	5	1	5	1

	Observer 2 (D.A.): Five repetitions						
		GYN1		GYN2		Early pregnancy	
		Median	p-value	Median	p-value	Median	p-value
Checklist	Skill1	1	1	1	1	1	1
	Skill2	1	1	1	1	1	1
	Skill3	1	1	1	1	1	0.9
	Skill4	1	1	1	1	1	1
	Skill5	1	1	1	1	0.5	1
	Skill6	0	0.9	1	0.9	1	1
	Skill7	0	0.9	0	0.9	0	1
GRS	Skill1	4	0.9	4	1	3	0.9
	Skill2	4	0.9	4	0.9	3	1
	Skill3	3	0.9	3	1	3	0.9
	Skill4	3	0.9	3	1	4	1
	Skill5	3	0.9	3	1	2.5	1
	Skill6	2	1	3	0.9	4	1
	Skill7	2	1	1	0.9	2	0.9
	Skill8	5	1	5	1	5	1

Table (4.4): Intra-observer statistical significance Kruskal-Wallis test of overall score of five repetitions of ten videos measured by each observer and marked in checklist and GRS for the three assignments; GYN1, GYN2 and early pregnancy.

		Significance (p-value) of overall score obtained in five repetitions		
		GYN1	**GYN2**	**Early Pregnancy**
Observer (A.A)	Overall checklist	0.9	0.9	1.0
	Overall GRS	0.9	0.9	1.0
Observer (D.A)	Overall Checklist	0.9	0.9	0.9
	Overall GRS	0.9	1.0	0.9

Table (4.5): Intra-observer agreement statistics: intra-class correlation ICC absolute agreement among five repetitions as scored independently by observer A.A. and D.A. (n=50)

	ICC absolute agreement		
	Checklist	**GRS**	**Agreement**
Observer (A.A.)	0.80	0.80	Excellent
Observer (D.A.)	0.72	0.71	Good

ICC, intra-class correlation coefficient; <0.40 = poor, 0.40–0.75 = fair to good, and >0.75 = excellent agreement.

Table (4.6): Variance component analysis for testing generalisability of 144 videos out of 1134 videos in the randomised controlled trial.

Component	Total checklist Var (R)	Total GRS Var (R)
Raters	0.013	0.039
Raters and assignments	0.056	0.288
Raters and videos	0.008	0.167
Var (error)	0.000	0.000

The result applied for mixed model analysis

Table (4.7): Inter-observer reliability: intra-class correlation coefficient ICC to measure absolute agreement between two observers' ratings using checklist and GRS to assess ultrasound performance of 144 videos.

Scale	ICC (95% CI)
Checklist (overall scores)	0.96 (0.95 – 0.97)
GRS (overall scores)	0.97 (0.95 – 0.98)

CI = Confidence interval
ICC, intra-class correlation coefficient; <0.40 = poor, 0.40–0.59 = fair, 0.60 -0.74 = good, and ≥0.75 = excellent agreement.
There is no statistical significant difference between the two observers' scores with: (1) checklist as tested by Mann-Whitney U, p=0.7, 0.2, 0.2, or (2) with GRS as tested by one-way ANOVA, p = 0.9, 0.4 and 0.9 for the GYN1, GYN2 and EP assignments respectively

Table (4.8): Inter-observer reliability using kappa **Ƙ** values of checklist and GRS in the three assignments; GYN1, GYN2 in gynaecology assignment and early pregnancy in obstetrics assignment, (n=48).

	GYN1 Ƙ (95 % CI)	**GYN2 Ƙ (95 % CI)**	**Early Pregnancy Ƙ (95 % CI)**
Total Checklist	0.83(0.35-1.79)	0.78(0.29-1.74)	0.92(0.42-1.96)
Total GRS	0.56(0.22-1.04)	0.60(0.33-1.24)	0.69(0.33-1.61)

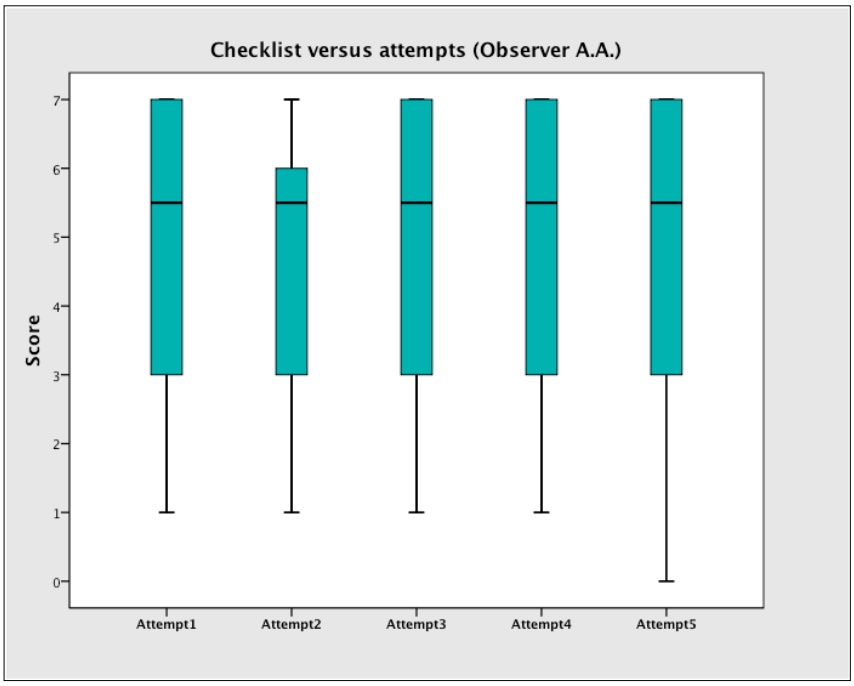
CI = Confidence interval
Kappa **Ƙ** values; < 0.20 = poor, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 good, and > 0.80 = very good

Table (4.9): Inter-observer reliability (kappa) for individual skills in checklist and GRS ratings arranged by assignments, (n=48).

Skills	GYN1 Ƙ (95 % CI)	GYN2 Ƙ (95 % CI)	Early pregnancy Ƙ (95 % CI)
Chk SK1	0.74 (0.55 – 1.17)	0.79 (0.62 – 1.30)	1.00 (1.00 – 1.96)
Chk SK2	0.92 (0.80 – 1.63)	0.83 (0.68 – 1.41)	0.93 (0.79 – 1.62)
Chk SK3	0.96 (0.88 – 1.76)	0.95 (0.86 – 1.74)	0.83 (0.67 – 1.40)
Chk SK4	0.92 (0.80 – 1.63)	0.87 (0.72 – 1.49)	0.95 (0.86 – 1.73)
Chk SK5	0.79 (0.62 – 1.31)	0.83 (0.68 – 1.40)	0.86 (0.71 – 1.47)
Chk SK6	0.88 (0.74 – 1.52)	0.74 (0.55 – 1.18)	0.96 (0.88 – 1.76)
Chk SK7	0.58 (0.35 – 0.81)	0.50 (0.29 – 0.68)	0.90 (0.77 – 1.58)
GRS SK1	0.54 (0.37 – 0.82)	0.57 (0.42 – 0.89)	0.62 (0.42 – 0.92)
GRS SK2	0.56 (0.40 – 0.86)	0.70 (0.55 – 1.15)	0.54 (0.33 – 0.76)
GRS SK3	0.62 (0.46 – 0.98)	0.74 (0.59 – 1.24)	0.64 (0.47 – 1.00)
GRS SK4	0.65 (0.49 – 1.04)	0.57 (0.39 – 0.86)	0.63 (0.46 – 0.99)
GRS SK5	0.59 (0.42 – 0.91)	0.57 (0.40 – 0.87)	0.73 (0.58 – 1.21)
GRS SK6	0.56 (0.40 – 0.87)	0.58 (0.42 – 0.90)	0.73 (0.58 – 1.22)
GRS SK7	0.59 (0.43 – 0.93)	0.58 (0.42 – 0.90)	0.77 (0.63 – 1.31)
GRS SK8	0.39 (0.22 – 0.51)	0.50 (0.33 – 0.74)	0.91 (0.79 – 1.61)

CI = Confidence interval, Chk= checklist
GYN1: anteverted uterus, GYN2: retroverted uterus, EP: early pregnancy CI: confidence interval
Kappa Ƙ values; < 0.20 poor, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 good and > 0.80 very good

Chart (4.1): Intra-observer agreement study: Boxplots represent median, first quartile, third quartile, minimum, maximum, outliers of scores of checklist and GRS versus the five attempts of observer A.A (n=50 per attempt). No statistical significant differences among the five attempts as tested by Kruskal-Wallis, $p \geq 0.05$



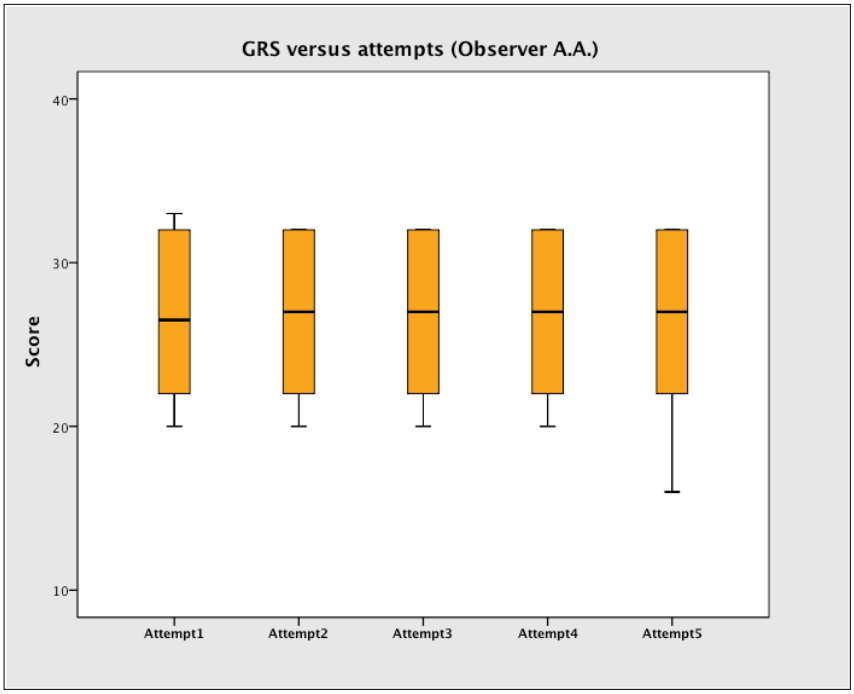
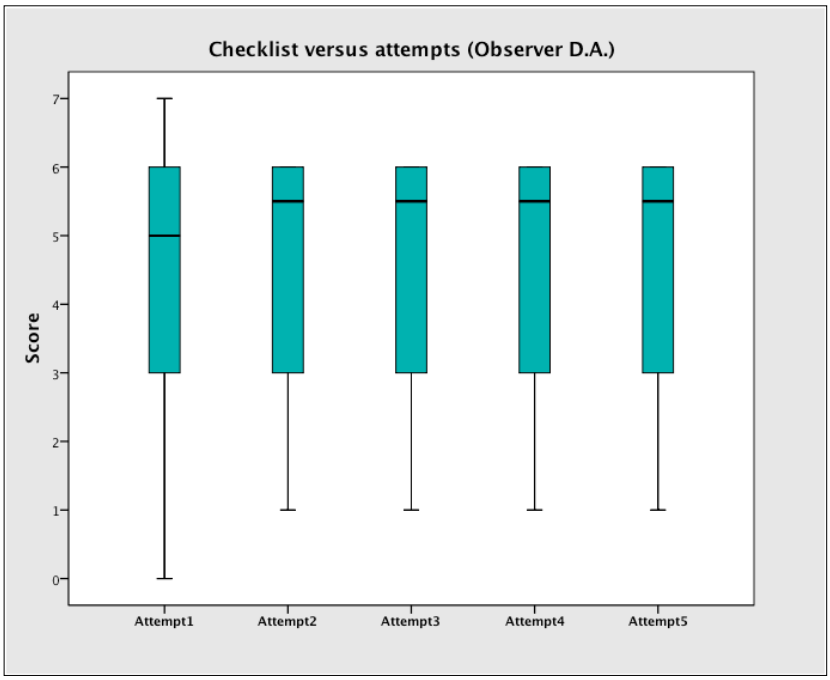


Chart (4.2): Intra-observer agreement study: Boxplots represent median, first quartile, third quartile, minimum, maximum, outliers of scores of checklist and GRS versus the five attempts of observer D.A (n=50). No statistical significant differences among the five attempts as tested by Kruskal-Wallis, $p \geq 0.05$



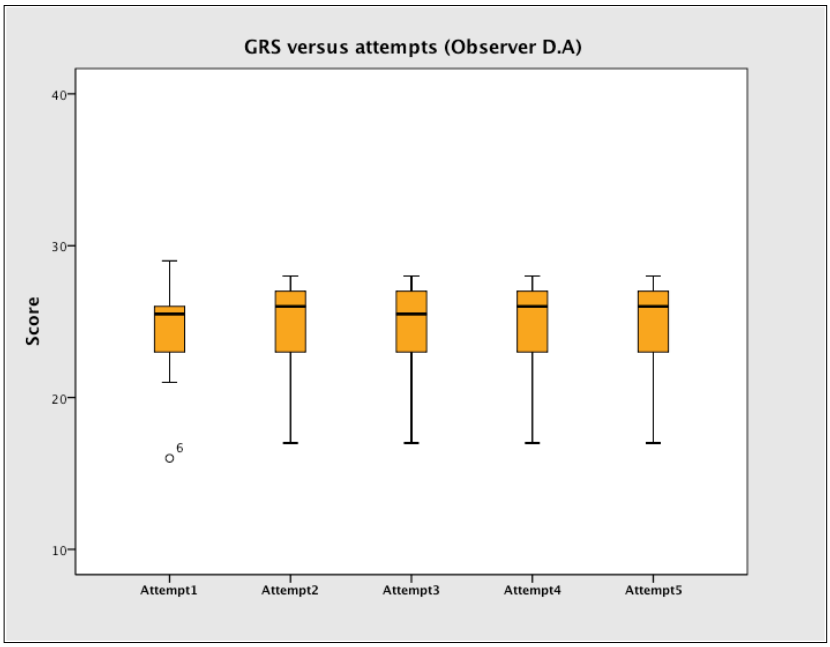


Chart (4.3): Inter-observer reliability: Bland and Altman plots of checklist demonstrate difference in mean scores between observers A.A and D.A in each assignment; GYN1, GYN2 and early pregnancy, (n=48), and tested by Spearman correlation coefficient r

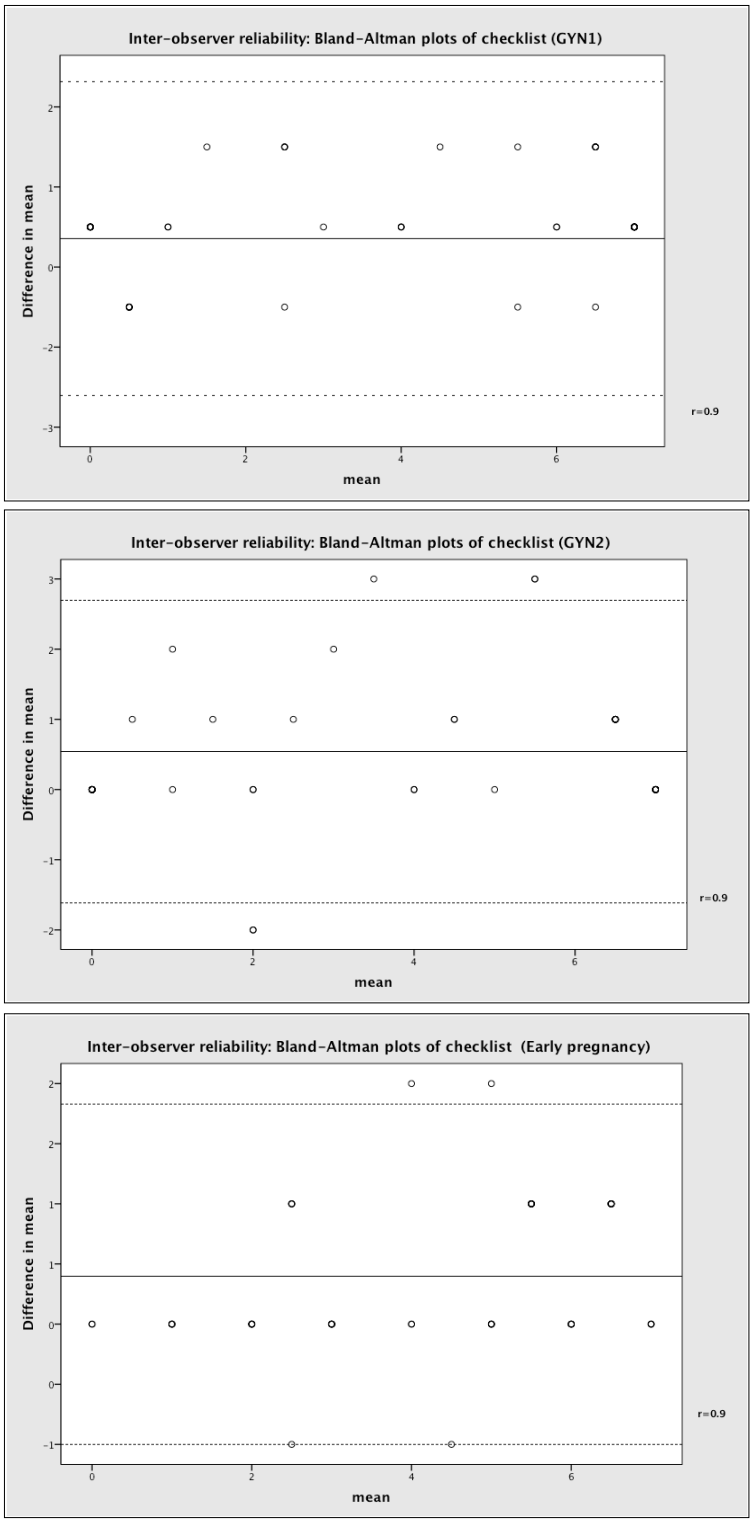
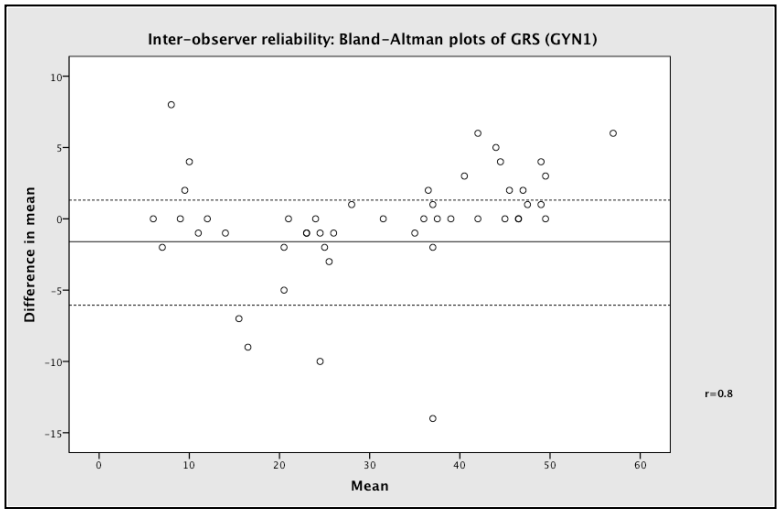
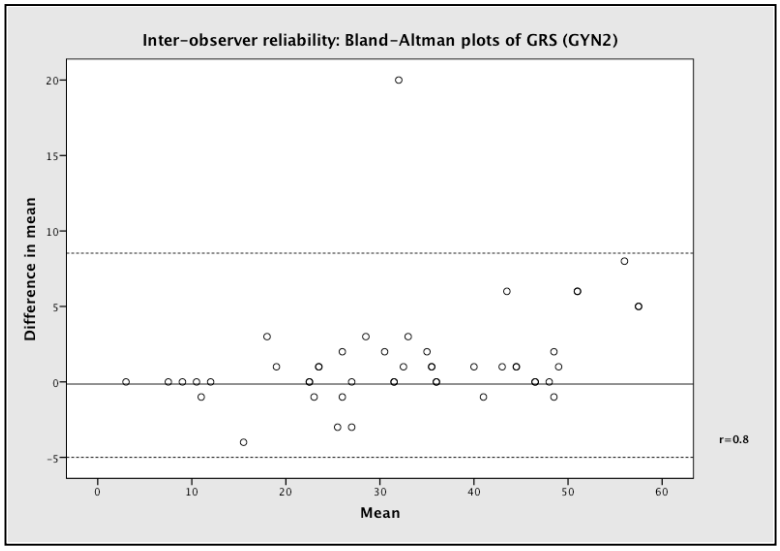


Chart (4.4): Inter-observer reliability: Bland and Altman plots of GRS demonstrate difference in mean scores against the average mean score obtained by the observers A.A and D.A in each assignment; GYN1, GYN2 and early pregnancy, (n=144), and tested by Pearson correlation coefficient r





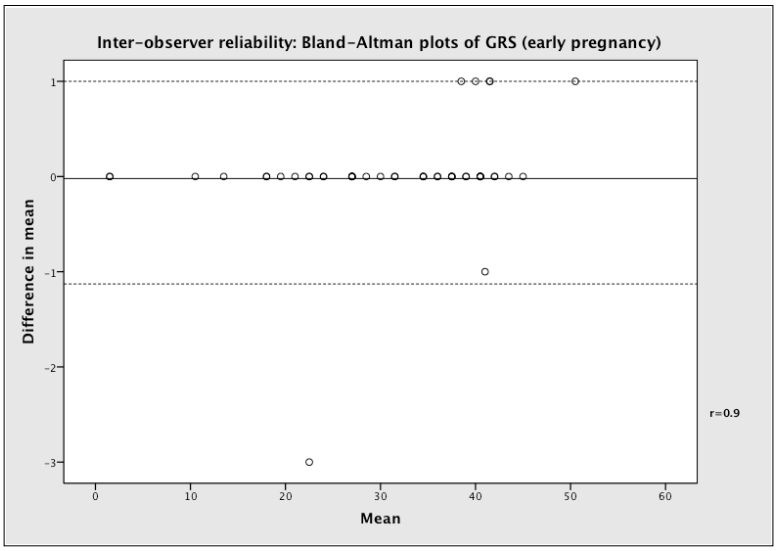
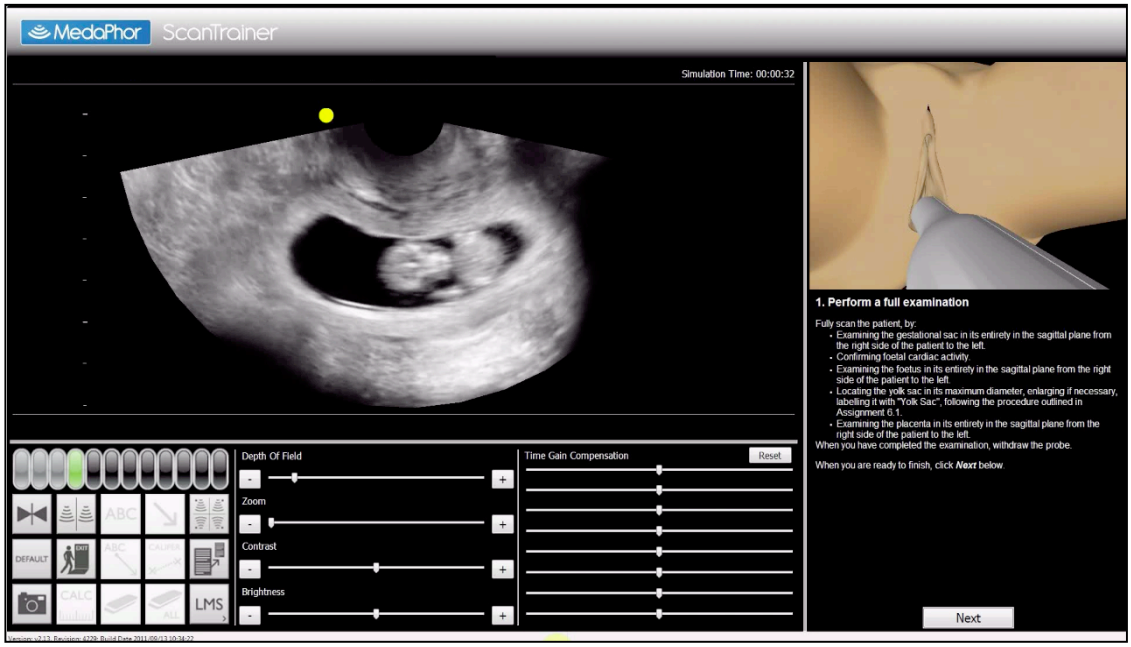
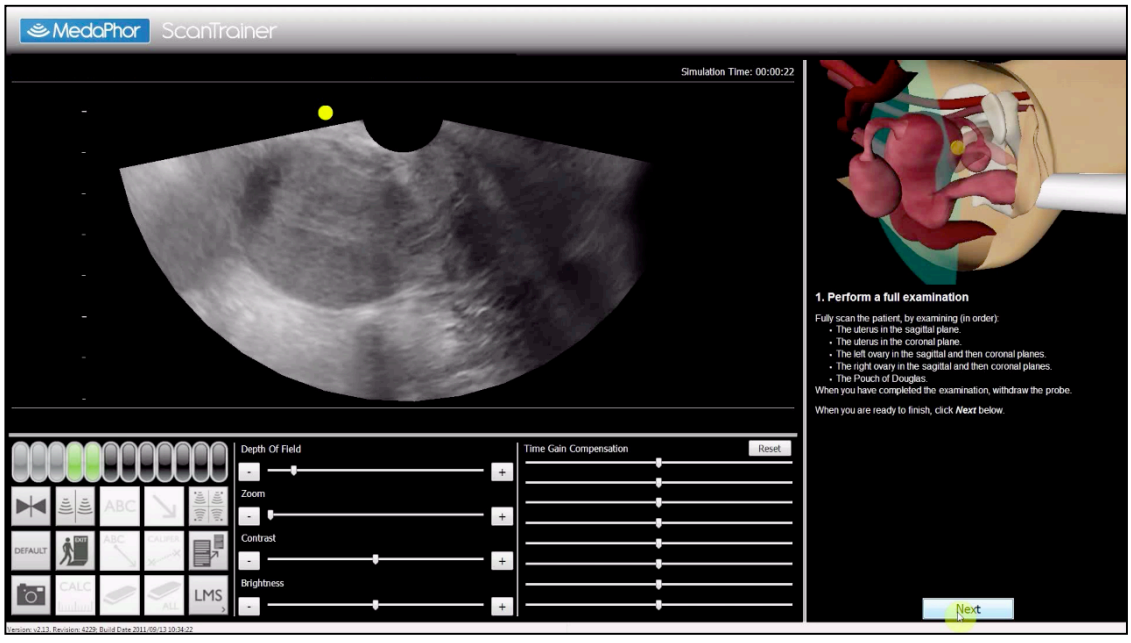


Figure (4.1): BB Flash Back Express video records the PC screen, sound and movies within the ScanTrainer® ultrasound simulator; name of trainee, time, and assignment given including all motions applied to the ultrasound image. Example of concealed anatomy in obstetrics module (4.1a), and revealed anatomy in gynaecology module (4.1b).



(4.1.a): Example of concealed anatomy in obstetrics module



(4.1.b): Example of revealed anatomy in gynaecology module

Chapter five

Validation of simulator metrics

CHAPTER 5

Validation of simulator metrics

The assessment of clinical performance in medicine is crucial and challenging (Good M, 2003; Cook et al., 2011; Kumar et al., 2015). A recent review of medical education brought to the fore new systems for the assessment of competence and performance (Carney et al., 2016; Winkel et al., 2016). Simulation-based assessment is one of these systems and refers to any educational activity that utilises simulation aides to replicate clinical scenarios supplied by timely "automated" feedback (Issenberg et al., 2005; Al-Elq A, 2010; Weller and Zachmann, 2012; Blum et al., 2013). With regard to simulation metrics-based assessment of ultrasound performance, the simulator provides computer-generated assessment of transvaginal ultrasound skills (TVUS) and tasks that the trainee can perform and receive immediate and detailed feedback.

The initial validation studies discussed in chapters three and four reflect the sophisticated and theoretical concepts of simulation metrics and helped to determine that the ultrasound simulator metrics are robust in reflecting the actual trainee's skills level. Yet such simulation metrics require further investigations. One major benefit of simulation metrics-based assessment is its ability to evaluate performance in context and provide its own formative assessment (Mayrose et al., 2007; Cook et al., 2011; Tarara et al., 2014). A focus on the assessment mechanism showing how simulation-based assessment in the simulator can be used for significant practical benefits would be to assess the validity of the two discrete approaches taken by simulator metrics (Persoon et al., 2010; Erdogan et al., 2016). The first approach is the construct validity approach which is appropriate for demonstrating identical feedback on TVUS performance achieved using different assignments/tasks that have identical learning objectives. Those assignments are a selection of individual skill (IS) tasks and full examination (FE) tasks. The second approach is to choose either (IS) or (FE) in the

relevant assignments to employ for the assessment of participants' TVUS performance in the randomised controlled trial (RCT). The virtual reality simulator learning system (metrics) has two designs of metrics based on hierarchical task analysis which consists of the following learning components: assignments, tutorials, tasks and skills.

5.1 Aim and objectives

The aim of this study was to ensure that the two metrics designs (IS and FE) in the ultrasound simulator are consistent in providing identical feedback for defined tasks on TVUS skills and performance by the same subject as well as feedback that is consistent with that given by a human judge. This would determine the reliability of simulation-based assessment and its suitability for reporting the actual performance of trainees and for reflecting their gradual change in TVUS practical level during the six assessments sessions in the RCT.

The objectives were (1) to determine the reliability between the two metrics designs in the simulator: individual skill task "IS" and full examination task "FE", in providing consistent and identical feedback for defined tasks on TVUS performance of participants; (2) to determine the level of 'absolute' agreement between simulator metrics (IS and FE) as compared with the observer, and finally (3) to determine the absolute agreement between the simulator metric (FE) and the observer in assessing three assignments in the six assessment tests.

5.2 Subjects and method: Reliability of simulation metrics: IS & FE tasks: (Objective one)

5.2.1 Sample size

As a sampling strategy, the self-selection method was used to invite participants to take part. The participants were trainees and trainers in obstetrics and gynaecology in the Welsh Deanery of Obstetrics and Gynaecology in Cardiff, UK. This selection method may possibly bring with it inherent bias, as those subjects who responded first to the invitation to participate may be more interested in the project and so decided to help in this research. Inclusion criteria for participation were based on the number of

years of ultrasound experience; type of scans obstetrics or gynaecology, TVUS or abdominal scans and how often the participants scanned. The experts were the subjects who had ultrasonography experience of more than two years and conducted daily scanning sessions. In contrast, the novices group had limited or no experience with ultrasound scanning; had practice for less than two years ultrasound experience; had very limited scanning sessions e.g. once/month or occasionally. All participants consented and filled in the participation form before taking part (Appendix I and II).

5.2.2 Procedure

After a brief introduction to the simulator, the participants were familiarised with the use of the simulator by practising guided sample tasks to understand how the simulation system works. Each participant was asked to perform and complete nine assignments in the simulator; four assignments from the gynaecology module (normal uterus), and five assignments from the obstetrics module (early pregnancy). The first three assignments in the gynaecology module were classified as individual skill (IS) tasks, while the fourth assignment was classified as a full examination (FE) task. With regard to the obstetrics module, there were four (IS) tasks and one (FE) task.

The participants performed nine assignments as instructed by the PhD researcher who scored and assessed the performance using the pass/fail checklist, (Appendix 5.1). Each module had a specific checklist scoring system according to particular skills evaluated in that module. Each checklist sheet included a section for evaluation obtained by the observer, and another section for simulation feedback which was automatically generated by the simulator. There was one assessment session for each participant; which lasted approximately 30 minutes.

5.2.3 Skills checklist used for assessment

The checklist consists of a number of skills: six from the gynaecology module and seven from the obstetrics (early pregnancy) module. Each of these skills scores (1) for correct and (0) for incorrect or not attempt (NA) performance.

5.2.4 Assignments and tasks included in the assessment

The list of TVUS skills is described in detail in Table 5.1.

5.2.4.1 Core gynaecology module:

To achieve consistency with the ScanTrainer® simulator version (2.13-SVN Build4258), assignments coded as (6.1), (7.2) and (7.3) are considered as individual skill tasks (IS), while assignment (8.2) is the full examination task (FE).

Assignment (6.1) examination of the uterus in sagittal and coronal planes
Assignment (7.2) examination of the right ovary in sagittal and coronal planes
Assignment (7.3) examination of the left ovary in sagittal and coronal planes
Assignment (8.2) final examination of the uterus and ovaries

5.2.4.2 Core obstetrics module:

Assignments coded as (4.1), (5.1), (6.1) and (6.2) are considered as (IS), while assignment (7.2) is (FE).

Assignment (4.1) examination of gestational sac (GS) in sagittal plane
Assignment (5.1) examination of fetus in sagittal plane
Assignment (6.1) examination of yolk sac (YS)
Assignment (6.2) examination of the placenta in sagittal plane
Assignment (7.2) final examination of GS, fetus, viability, YS and placenta

5.2.5 Simulation metrics

The ultrasound simulator learning management system (LMS) offers metrics-based feedback on important aspects of skill performance in transvaginal ultrasound, in gynaecology and early pregnancy. The metrics were developed from analysis of the tasks in a skill or performance to be learned, and the outcome of the analysis shaped how the simulation looked and behaved, in order to generate feedback on that particular skill. The skill in the simulation metrics is robustly defined as *what the skilled individual does.* The task analysis separates skilled task performance into components (metrics). Hence these performance units must be unambiguously defined so that they can be scored as occurring correctly, incorrectly or not occurring. These metric units capture the essence of performance and include the steps and order of steps that the performance should be performed in. Crucially, the metrics should describe for each performance step what should and what should not be done, thus characterising performance that deviates from optimal performance. Some

assignments in the simulation metrics are individual skill tasks (IS) which comprise those several performance steps required to complete a performance skilfully. In other words, these specific tasks analyse the performance steps and evaluate them independently at the end of the learning procedure by providing detailed feedback on each skill performed correctly or incorrectly. The metrics-based assessment (IS) deals with each specific skill individually and analyses it separately from other skills included in the same tutorial/assignment. With regard to full examination tasks (FE), the operational definition of performance tasks/skills is a life-like examination where all skills should be performed in a systematic order, and the task analysis for all skills is provided after completion of the examination without the need for a task-analysis hierarchy for the performance of each skill. This approach ensures a homogenous skill-set in evaluating subjects with different levels of ultrasound experience.

5.2.6 Outcome measure

The primary outcome was to determine the ability of simulation metrics-based assessment with IS and FE in providing consistent feedback on ultrasound performance with the ultrasound simulator.

5.3 Subjects and method: agreement between simulator metric (FE) tasks and the observer: (Objective two)

5.3.1 Sample size:

This study included subjects enrolled in randomised controlled trial (RCT). The subjects were primarily selected from specialty trainees (ST) in Obstetrics and Gynaecology in the Welsh Deanery, in addition to other NHS staff and students of the MSc programme at Cardiff University who fulfil the inclusion criteria. All those participants were novices and had a very limited ultrasound background. The participants were randomly assigned to intervention and control groups in the RCT and completed the six assessment sessions.

5.3.2 Procedure:

During the randomised controlled trial (RCT), the baseline assessment was the first session that the participants undertook, to assess their ultrasound skills in transvaginal ultrasound (TVUS) performance in gynaecology and early pregnancy. Each

participant was assessed once, every 4–6 weeks six times to complete a total of six sessions/tests. At each session, the participants - intervention and control - performed TVUS in the ultrasound simulator, and their skills were assessed using three different assignments: two in gynaecology (normal uterus GYN1 and retroverted uterus GYN2) and one in early pregnancy. Each participant was evaluated simultaneously by the observer and the simulator metrics (full examination task FE) for the same three assignments throughout the trial. Each performance was video recorded as well.

The checklist, presented in chapter five was used to evaluate TVUS simulated performance, Table (5.6). In accordance with that, the checklist consists of seven skills of each assignment in which there are two scorings: (score 1) if the skill correctly performed, (score zero) if it is incorrectly performed or not attempted (N/A). The outcome measure in this cohort study was to test agreement between the observer's ratings with the simulator metric feedback using (FE) tasks in evaluating gradual changes in trainees' performance at six different stages of the trial.

5.4 Statistical data analysis

IBM SPSS Statistics, version 20.0 was used for statistical analysis. For the normality test, two data in each cohort study were not normally distributed when tested by Shapiro-Wilk, p-value ≤ 0.05. Descriptive data of participants' demographic and ultrasound experience was detailed. Median values for: novices and experts performances as scored by simulator metrics (IS and FE) tasks and the observer. The statistical significance between two groups was tested by Mann-Whitney U and considered significant at p-value ≤ 0.05. Reliability and agreement were determined by using an intra-class coefficient (ICC).

For achieving objective one, and in order to estimate reliability between the two simulator metrics (IS and FE), and level of agreement between the simulator metrics and the observer, ICC was used with interpretations: <0.40 poor, 0.40–0.59 fair, 0.60-0.74 good and ≥0.75 excellent. Box-plots were represented median, first and third quartiles, minimum, maximum and outliers of scores obtained by the simulator metrics (IS and FE) and the observer in gynaecology and early pregnancy modules.

For achieving objective two, also ICC used to estimate level of agreement between the observer and the simulator's metric (FE) in the three assignments employed to assess TVUS performance in the RCT. Box-plots represented median, first and third quartiles, minimum, maximum and outliers of scores obtained by the simulator metric (FE) and the observer at each point of the six assessment sessions (tests) in RCT. For non-parametric data, the statistical significance between two scorings (the simulator's and the observer's) tested by Mann-Whitney U. Bland-Altman plots illustrated differences in mean value between simulator metric (FE) and the observer' ratings in the three assignments. Spearman correlation coefficient "r" used to measure the degree of linear relationship "correlation" between FE and the observer's scores. Correlation coefficient "r" ranges between -1.0 to +1.0 and the closer it is to +1 or to -1, the more closely the two ratings are related.

5.5 Result

5.5.1 Reliability of simulation metrics: IS and FE tasks: (Objective one)

5.5.1.1 Demographic (n=11)

A total of eleven subjects were recruited for this study. Six were assigned as experts and five as novices. Nine of the subjects were females and two males. The subjects included one consultant, four radiographers and one associate specialist (as experts), and three specialty trainees in obstetrics and gynaecology, one nurse and one postgraduate student (as novices). All the novices had less than six months TVUS ultrasound experience. Three novices were assigned to daily ultrasound scans while the remaining two subjects stated their scanning as 'occasional'. All experts had more than two years TVUS experience. Experts' scored were higher than novices in both (IS and FE) simulator metrics. Median scores for novices and experts are found in Table 5.2.

5.5.1.2 Reliability and agreement

Median values for individual skill tasks (IS) and full examination tasks (FE) are found in Table 5.3. There was no significant difference between (IS) and (FE) in evaluating the same performance and skill, Mann-Whitney U test, $p \geq 0.05$. The intra-class coefficient (ICC) revealed high reliability between (IS) and (FE) tasks: ICC values were 0.95 and 0.89 for gynaecology and early pregnancy modules respectively (Table 5.4). The variation of ICC values in rating individual skills included in the checklist

of gynaecology and early pregnancy modules as rated by simulator metrics IS and FE ranged from excellent to fairly good (Table 5.4). The boxplots illustrated in chart 5.1 represented outliers, minimum, first quartile, median, third quartile and maximum of scores for IS, FE and the observer.

As detailed in Table 5.5, the absolute agreement between the observer scores and the two tasks (IS) and (FE) were high: ICC values were 0.87 and 0.96 for the early pregnancy module and 0.92 and 0.77 for the gynaecology module for (IS) and (FE) respectively. The ICC values of individual skills rated by the observer, (IS) and (FE) ranged from excellent to fairly good.

5.5.2 Agreement: simulator metric (FE) tasks and the observer: (Objective two)

5.5.2.1 Sample size (n=1134)

The number of participants recruited in the randomised controlled trial (RCT) was sixty-three novices. A total of 18 assessment sessions (tests) were arranged for each participant during the trial and there were three assignments used to evaluate the participants' performance at each session. Thus a total of 378 performances were completed for each assignment and the total sample size was (n=1134) performances achieved and scored by the observer and the simulator metric (FE) in this study.

5.5.2.2 Agreement: simulator metrics (FE) and the observer

The intra-class correlation coefficient ICC values revealed excellent agreement between the observer's ratings and the simulator metric (FE) tasks in the three assignments: ICC values were 0.96, 0.83, and 0.86 for GYN1, GYN2 and early pregnancy assignments respectively (Table 5.7). The ICC values for each individual TVUS skill were the highest in the GYN1 as compared with GYN2 and early pregnancy assignments. All seven skills in GYN1 showed excellent agreement between the observer and the simulator metrics while there were only two skills (SK3&6) in GYN2 and three (SK1, 4&5) in early pregnancy assignment that showed excellent agreement. The level of agreement ranged between fair to good in four skills (SK1, 2, 4&5) in GYN2 while poor agreement was found in one skill (SK7). In addition, poor agreement was found in skill (SK3) the early pregnancy assignment. The box-plots in charts 5.2-5.4 represented outliers, minimum, first quartile, median, third quartile and maximum of scores of simulator (FE) and the observer at the six

tests in the RCT. There was no statistically significant difference between the simulator and the observer scores when rating GYN1 (Mann Whitney U test, $p \geq 0.05$). On the other hands, the statistically significant difference was noted at baseline and tests 2 and 3 in GYN2, and at baseline and test 2 in the early pregnancy assignment. Bland-Altman plots in chart 5.5 represent the correlation between the scores of the observer and the simulator metric (FE) positive correlation ($r = 0.9$, 0.8 and 0.8) were noted for GYN1, GYN2 and early pregnancy assignments respectively.

5.6 Discussion

The study's findings demonstrated high reliability between the simulator's metrics when used to assess trainees TVUS performance and equally excellent overall agreement between the FE and observer ratings.

The assessment of clinical performance in medicine is important but challenging (Al-Rasheed et al., 2013; Kumar et al., 2015). Although there are many forms of assessments used to demonstrate trainees' knowledge and/or competence, there is evidence that assessment-based competence does not reliably predict performance in clinical practice (Rethans et al., 2002; Carney et al., 2016). With the development of technology, there is increased need to facilitate standardisation through the use of feedback tools as assessment for learning rather than solely as assessment of learning (Stefanidis et al., 2009). Assignments and tasks chosen in this study were organised according to a principle of cognitive apprenticeship in a systematic learning approach that were applied through simulation technology to develop TVUS skills (ScanTrainer®, Medaphor plc, Cardiff, UK). These assignments included two methods of assessment to reflect the evaluation of a skill-by-skill checklist *as if provided by individual skill task (IS)* and also to reflect a life-like scan scenario by utilising a *full examination task (FE)*. Goodyear (1999) pointed out that part of the value of having a pedagogical framework is to help with the analysis of an educational and learning assignment into components in order to help with tasks and skills evaluation and learning. Therefore, tasks-analysis in simulation metrics-based assessment comprises a representative set of outcome measures of ultrasound skills reflecting what occurs in real practice. More reliable metrics-based feedback should offer improved opportunities to change clinical performance and enhance trainees'

practical knowledge about ultrasound scanning (Slagle et al., 2002; Chalouhi et al., 2015a; Tolsgaard et al., 2014a).

The ultrasound simulator metrics are learning contexts that are clinically relevant and their face and content validity were demonstrated in chapter three. The simulator metrics are also consistent and accurate as demonstrated by the observer. The (FE) task and metrics, which closely represented real-life scanning, were found to be fairly well correlated with the observer's assessment compared to the other metrics: individual skill (IS) tasks. Although IS had higher reliability, the FE tasks were chosen as assignments to measure the skill development in the randomised controlled trial (RCT). There were two reasons for choosing the FE tasks. The first was that the FE assignments provided a real life scan with homogenous skill-sets, which made it possible to have an identical assessment of the control and intervention groups. The second reason was that by using the FE assignments it was possible to hide the instructional 3D virtual anatomy screen of pelvic structures. On the other hand, while training IS tasks in simulation metrics provide details of the steps needed for a particular performance, and this approach helps the trainees to develop their skills and the trainers to monitor trainees' progress. The FE tasks had poor agreement in one skill (SK 7: identifying the POD) in GYN2 and two (SK 2 and 3: examine the fetus and viability) in early pregnancy. In order to perform these skills, optimisation of the image should be applied by maximise the magnification to a certain extent and centralise the image prior to scanning through the area longitudinally from side to side. Despite the fixed structure of the POD in retroverted uterus (GYN2), This may indicate that the POD skill in GYN2 metrics was not as robust as the other assignments and thus the observer was more accurate in assessing TVUS performance. Another possible reason was that the retroverted uterus in the simulator may have been a more difficult assignment especially that the uterus was rotated along its axis and hence required a much higher level of technical skills and knowledge. Similarly with the early pregnancy assignment, examining the fetus and detecting fetal heartbeat required optimisation and centralisation of the image and thus had poor agreement. It is important to note that (100%) of skills in GYN1, (71%) in early pregnancy and (28%) in GYN2 had excellent agreement between FE and observer's evaluations.

In most educational settings, teachers provide students with information on the level of performance against the goal but fail to go further. By not doing so teachers do not engage in action which closes the gap (Conaghan and Lockey, 2009; Slagle et al. 2002). In the simulator, the set of instructions teach the skill and the metrics evaluate trainees' performance while the learning management system keeps the record of trainees' performance over several attempts to monitor their progress. In addition, simulation metrics feedback is a core part of the learning process where the trainees are able to compare their actual level of performance with the goal or standard. In comparison to traditional clinical assessment, the observer's feedback is usually delivered verbally to trainees and not recorded as it is with the simulator (van der Vleuten and Schuwirth, 2005). One major advantage of simulation-based assessment is its ability to evaluate performance in context and help in minimising chances of observer biases that may be related to personal reasons or to a lack of standardised assessment criteria among assessors /examiners (Newble D, 2004; Norcini J, 2005). Additionally, it has been pointed out that feedback on personal characteristics can have a negative effect on individuals (Conaghan and Lockey, 2009). Despite the importance of instructor feedback, it can remain limited to identifying the gaps in students' knowledge without clear instructions on how to bridge the gap (Robertson and Bandali, 2008).

In other medical fields, simulation training has proven its ability to offer a useful environment for standardising learning and assessing methods to evaluate trainees' performance (Boulet et al., 2003; Erdogan et al., 2016). In order to encourage trainees/students to adopt a deep approach to learning, it is important that simulation metrics-based assessment has a sufficiently high quality of information that focuses on individuals and provides them with detailed feedback on the gap between their performance and what is regarded as optimal performance. Pelvic ultrasound phantoms e.g. Blue phantom™ are very realistic and enable a trainee using an ultrasound machine to acquire the relevant skills. However, it requires the presence of a trainer/instructor to teach and assist (Alsalamah et al., 2009; Bahner et al., 2013; Moak et al., 2014) and the use of an ultrasound machine. Despite the benefits of the ScanTrainer® being a self-directed learning tool, it is important to understand the limitation of metrics-based assessment which sometimes leads to flawed conclusions of correct or optimal practice, as the metric is inflexible and has limited ability in

comparison to human assessors (Gallagher et al., 2012; Graham et al., 2012). In other words, sub-optimal performance of ultrasound skills in the simulator may be considered as failure while the same performance would be accepted if it occurred in a real scan. Hence, these limitations should always be considered even though, the simulator is a very powerful training tool for the delivery of deliberate practice coupled with formative and summative metrics-based feedback. In the absence of simulation, formative feedback on training performance needs to be delivered by a trainer who is very experienced in performance assessment. Therefore, metrics-based performance has a clear end point. For example, the testing of competency and proficiency level must provide the facilities and opportunities for skill acquisition (Hsu et al., 2016). Consequently, such a learning approach should be structured to integrate the concept of constructive feedback to improve the learning outcomes.

As in many trials, the small sample size of the first cohort study (n=11) was one of the study's limitations. Arguably, although the self-selection method used to invite participants was widely accepted, this may reflect inherent bias as participation inevitably required participants' interest in taking part in the pilot study. A clear focus of the first cohort study was on the evaluation of the reliability of simulation metrics and feedback on a series of assignments performed by novices and experts. Although novices scored lower than experts, there were no statistically significant differences between the two groups in any of the assignments. Further validation studies on a larger scale were conducted and will be explained in detail in chapter six. The second cohort study had an adequate sample size and the power calculation was considered and explained further in chapter seven. Furthermore, the statistically significant differences between the simulator and the observer ratings of TVUS performances were demonstrated in the early assessments of novices' skills in the RCT but not afterward.

In summary, the assessment of ultrasound skills is essential for quality assurance and the most appropriate method should be identified. Simulation metrics-based performance enhances and reinforces trainees' self-confidence by providing detailed feedback and identifying areas in need of improvement. It also provides the opportunities for repeated practice before entering the real clinical situation (Madsen et al., 2014; Chalouhi et al., 2015a).

Chapter 5: Tables and figures

Table (5.1): Skill task-specific checklist description of gynaecology (GYN1) and early pregnancy modules (used for Objective one)

Simulator skill task	Description
Gynaecology module	
Skill 1	Examining the uterus in sagittal plane
Skill 2	Examining the uterus in coronal plane
Skill 3	Examining the left ovary in coronal plane
Skill 4	Examining the left ovary in sagittal plane
Skill 5	Examining the right ovary in coronal plane
Skill 6	Examining the right ovary in sagittal plane
Early pregnancy module	
Skill 1	Examining gestational sac in sagittal plane
Skill 2	Examining fetus in sagittal plane
Skill 3	Examining fetus heart activity
Skill 4	Viewing yolk sac
Skill 5	Labelling yolk sac
Skill 6	Optimise image while viewing yolk sac
Skill 7	Examining the placenta in sagittal plane

Table (5.2): Median scores of novices (n=5) and experts (n=6) as rated by the observer and the simulation metrics in gynaecology module (GYN1) (total of six skills) and early pregnancy module (total of seven skills). Significant difference between novices and experts was tested by Mann-Whitney U test.

		Median			
		Gynaecology module		**Early pregnancy module**	
		IS	**FE**	**IS**	**FE**
Observer	*Novices*	0	0	3	2
	Experts	4.5	6	7	7
Simulation	*Novices*	2	1	4	4
	Experts	4	5	6.5	7

No significance difference between two groups, $p>0.05$

Table (5.3): Median and significance values of simulator and the observer rating the individual skill (IS) and full examination (FE) tasks in gynaecology and early pregnancy modules. Mann-Whitney U test was used to test the statistical significance difference between the two scores obtained by the simulator and the observer

Gynaecology module (GYN1)	**Median**		***p*-value**
	Simulator (IS)	**Observer (IS)**	
	3	4	0.9
	Simulator (FE)	**Observer (FE)**	
	5	5	0.7
	Simulator (IS & FE), p=0.7		
Early pregnancy module			
	Simulator (IS)	**Observer (IS)**	
	6	5	0.9
	Simulator (FE)	**Observer (FE)**	
	6	6	0.7
	Simulator (IS & FE), p=0.6		

Table (5.4): Intra-class correlation coefficient (ICC) reliability of simulator metrics; (IS) and (FE) in the gynaecology (GYN1) and early pregnancy modules.

SKILLS	Reliability between simulator metrics IS vs FE	
	Gynaecology module	Early pregnancy module
	ICC	
Skill 1	0.58	0.38
Skill 2	1	0.64
Skill 3	0.55	1
Skill 4	1	0.62
Skill 5	0.80	0.89
Skill 6	0.62	0.80
Skill 7		0.55
Reliability (overall)	**0.95**	**0.89**

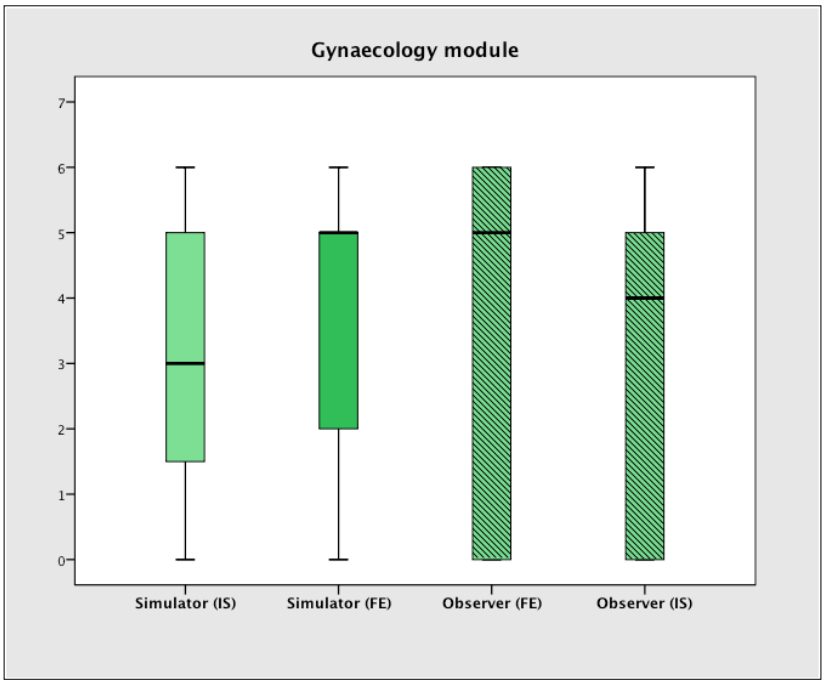
IS= individual skill, FE=full examination
ICC = (poor,<0.40), (fair, 0.40-0.59), (good, 0.60-0.74), (excellent, 0.75-1)

Table (5.5): Intra-class correlation coefficient (ICC) absolute agreement between the observer and simulator metrics; (IS) and (FE) in the gynaecology (GYN1) and early pregnancy modules.

SKILLS	Gynaecology module		Early pregnancy module	
	Observer vs IS	Observer vs FE	Observer vs IS	Observer vs FE
	ICC			
Skill 1	0.77	0.66	0.38	0.86
Skill 2	0.55	0.41	0.89	0.54
Skill 3	0.75	0.80	1	1
Skill 4	0.89	0.80	0.62	0.77
Skill 5	0.89	0.77	0.89	1
Skill 6	0.90	0.63	0.80	1
Skill 7			0.81	0.89
Absolute agreement (overall)	**0.92**	**0.77**	**0.87**	**0.96**

IS= individual skill, FE=full examination.
ICC = (poor, <0.40), (fair, 0.40-0.59), (good, 0.60-0.74), (excellent, 0.75-1)

Chart (5.1): Box-plots represent median, first quartile, third quartile, minimum, maximum and outliers of scores obtained by the simulator metrics (IS and FE) and the observer in Gynaecology and early pregnancy modules. No statistically significant difference between the scores obtained by the simulator and the observer as tested by Mann-Whitney U test.



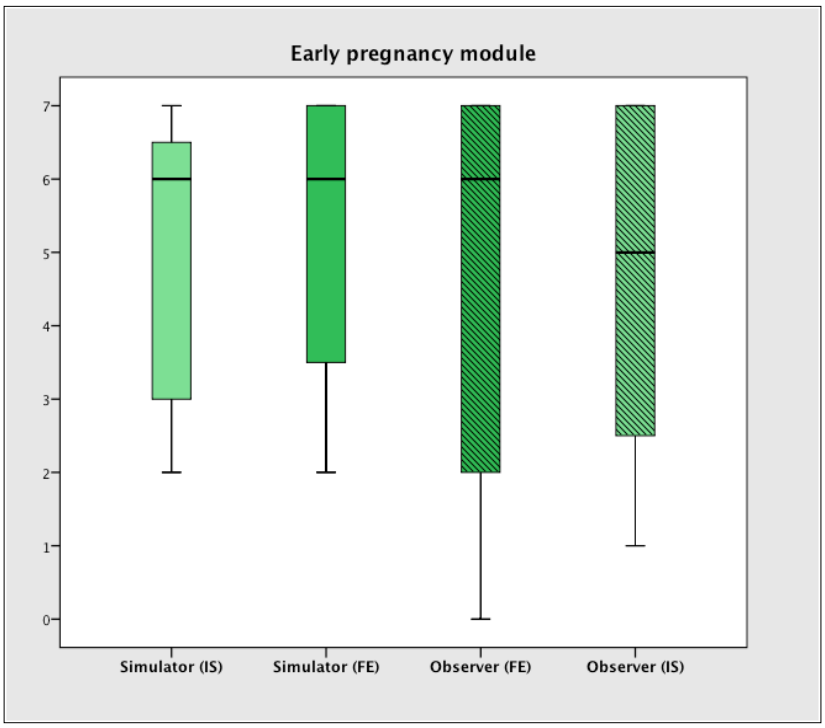


Table (5.6): Description of seven skills in the checklist for the three assignments: GYN1, GYN2 and early pregnancy (used for objective two).

	Skill	Skill description
GYN1	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas
GYN 2	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas
Early pregnancy	**Skill 1**	Examining gestational sac in sagittal plane
	Skill 2	Examining fetal heart activity
	Skill 3	Examining fetus in sagittal plane
	Skill 4	Viewing yolk sac
	Skill 5	Labelling yolk sac
	Skill 6	Optimise image in viewing yolk sac
	Skill 7	Examining placenta in sagittal plane

Table (5.7): Intra-class correlation coefficient ICC (absolute agreement) between the observer's and the simulator metric (FE) scoring seven skills in each of the three assignments: gynaecology GYN1, GYN2 and early pregnancy.

	Assignments		
Skills	**GYN1**	**GYN2**	**Early pregnancy**
	ICC	**ICC**	**ICC**
Skill 1	0.89	0.66	0.78
Skill 2	0.87	0.73	0.32
Skill 3	0.84	0.80	0.14
Skill 4	0.88	0.62	0.92
Skill 5	0.87	0.60	0.79
Skill 6	0.85	0.77	0.74
Skill 7	0.86	0.27	0.70
Overall	**0.96**	**0.83**	**0.86**

ICC = (poor,<0.40), (fair, 0.40-0.59), (good, 0.60-0.74), (excellent, 0.75-1)

Chart (5.2): Box-plots represent median, first quartile, third quartile, minimum, maximum, and outliers of scores performed by 63 trainees in GYN1 assignment, and rated by the observer (red box) and the simulator (FE) (blue box). No statistically significant differences were shown between the observer and the simulators scores at any point in the six test during the trial, Mann-Whitney U test, p>0.05.

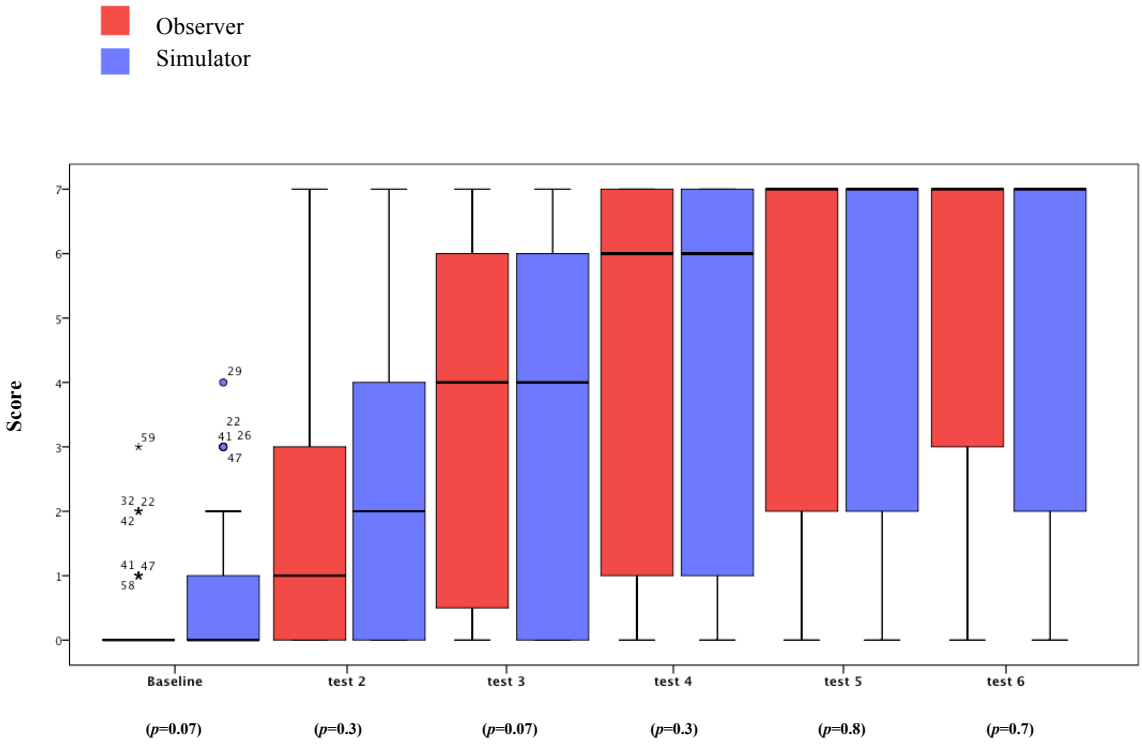


Chart (5.3): Box-plots represent median, first quartile, third quartile, minimum, maximum, and outliers of scores performed by 63 trainees in GYN2 assignment, and rated by the observer (red box) and the simulator (FE) (blue box). The statistically significant differences between the observer and the simulator ratings were demonstrated only at three points: the baseline, tests 2&3, Mann-Whitney U test, $p<0.05$.

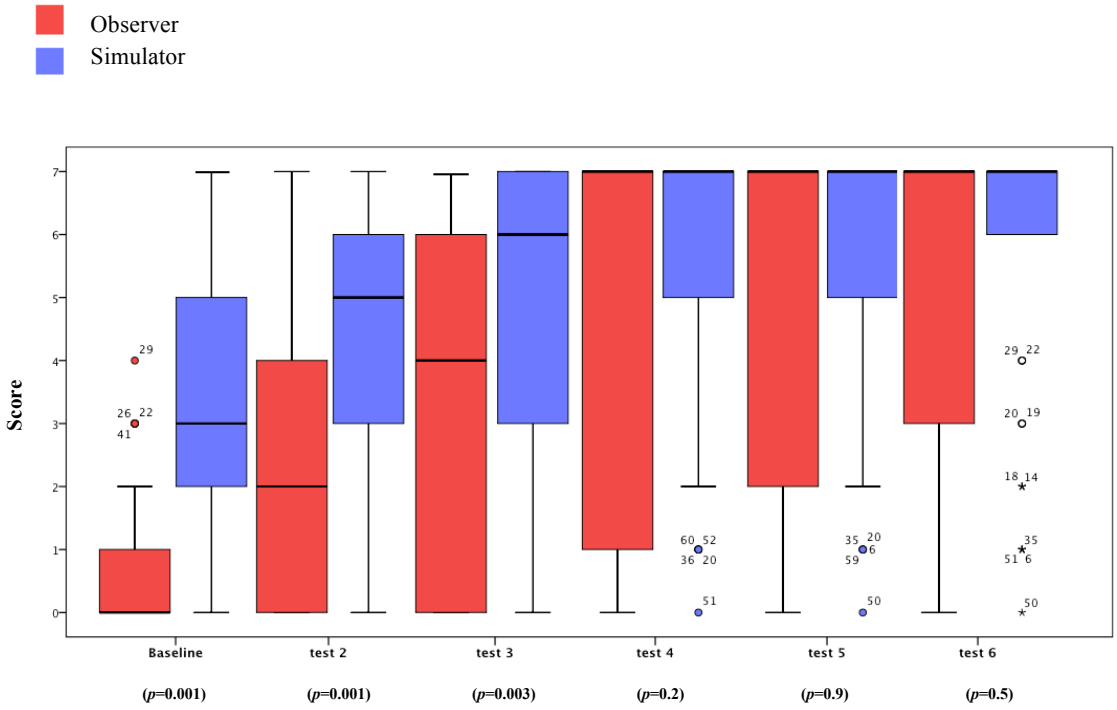


Chart (5.4): Box-plots represent outliers, minimum, first quartile, median, third quartile and maximum of scores performed by 63 trainees in early pregnancy assignment, and rated by the observer (red box) and the simulator (FE) (blue box). The statistically significant differences between the observer and the simulator ratings were demonstrated at the baseline and test 2 only, Mann-Whitney U test, $p<0.05$.

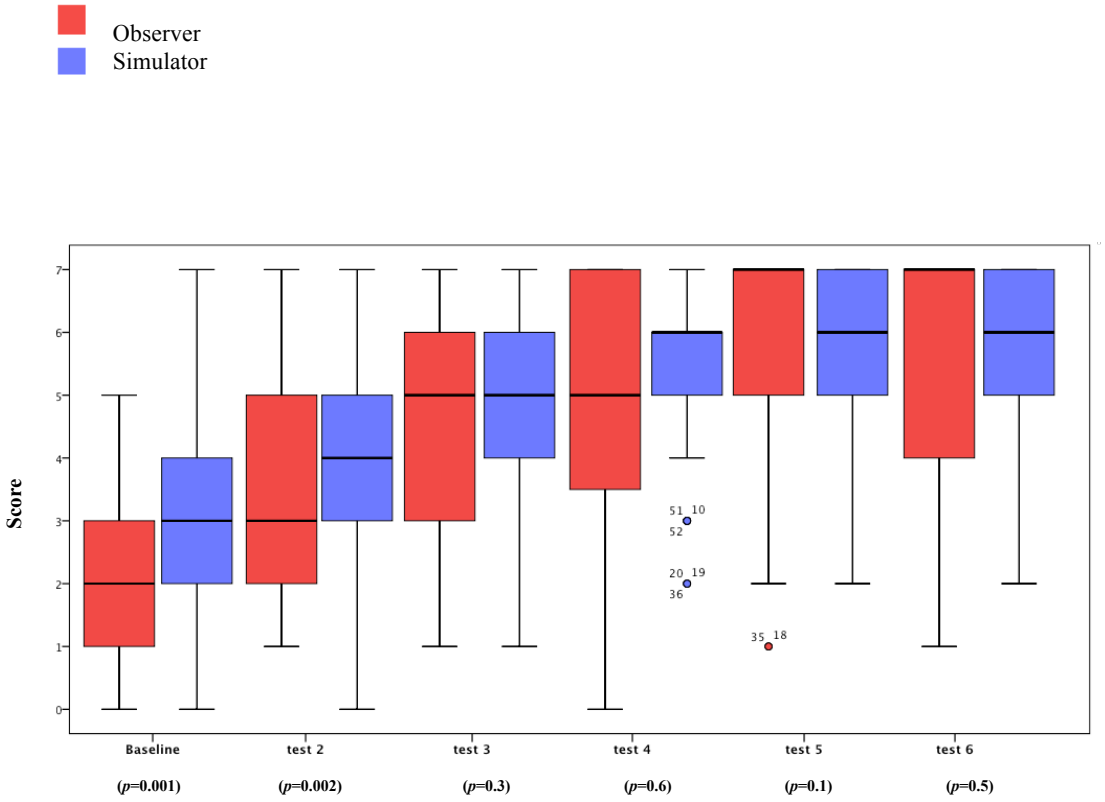
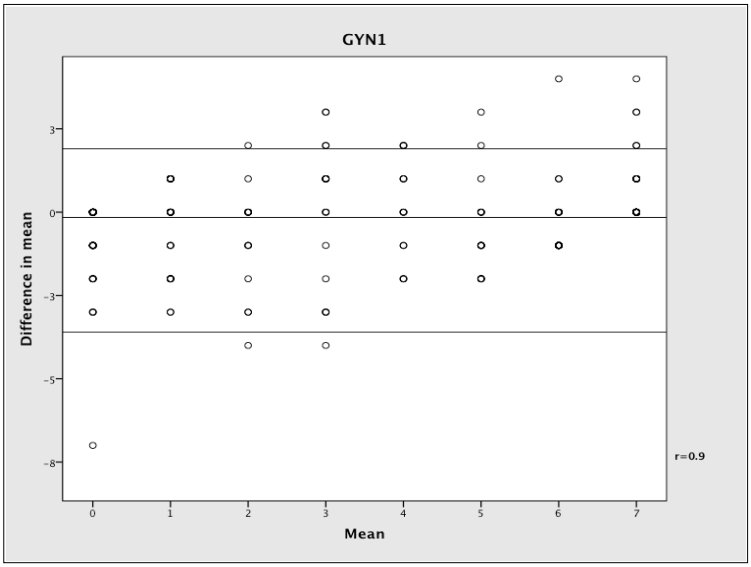
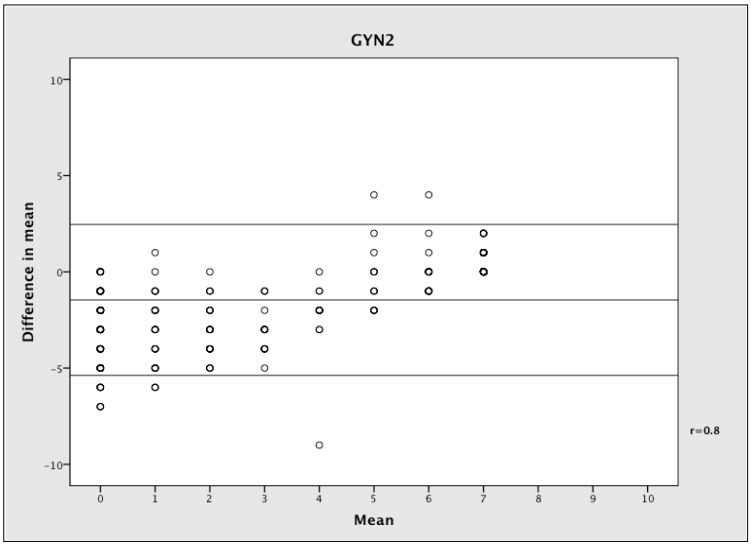
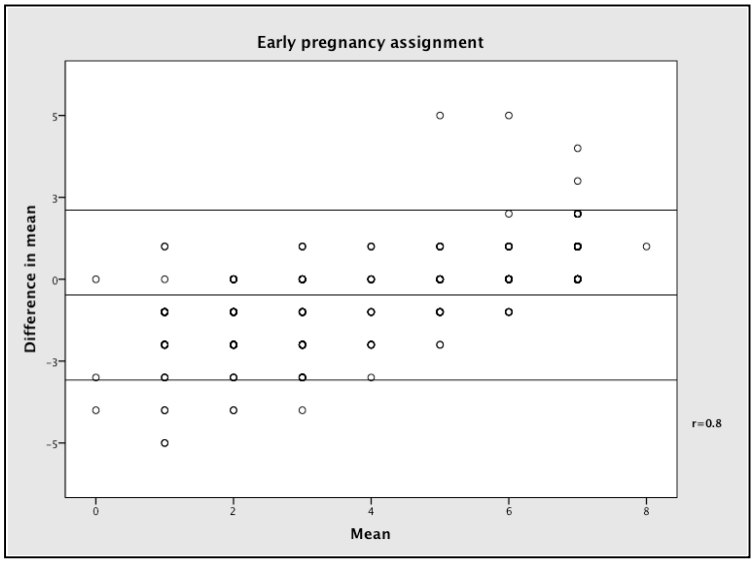


Chart (5.5): Bland-Altman plots represent difference in the means between the simulator's FE metric and the observer scores against the mean of the observer and the simulator, in each assignment: GYN1, GYN2 and early pregnancy.







Chapter six

Validation of subjects' performance on the simulator: Construct validity

CHAPTER 6

Validation of subjects' performance on the Simulator: Construct validity

---

Simulators may be beneficial learning tools and in the assessment of performance away from practising on patients. It also has the potential to expose both novices and experts to situations that occur infrequently (Ven Dongen et al., 2007; Erdogan et al., 2016). Despite simulators being promoted as means of assessing subjects' skills, there is very limited evidence in literature establishing a standard method of measuring levels of competence in ultrasound practice for students or junior doctors using an ultrasound simulator ScanTrainer® (Madsen et al., 2014; Tolsgaard et al., 2015a). Construct validity is defined as the degree to which a test measures what it claims to measure (Byrne and Greaves, 2001). There is limited evidence to support the premise that performance on the ultrasound simulator correlates with and reflects the subject's actual performance and determining its ability to differentiate between novice and expert users is vital to testing its construct validity.

6.1 Aim and objectives

The aim was to assess whether the ScanTrainer® ultrasound simulator could discriminate between novice, intermediate and experienced TVUS practitioners. The objectives were (1) to determine the difference in scores achieved by the three groups in performing three simulation assignments; (2) to determine the difference between scores obtained by the simulator and a human observer in assessing subjects' performances and (3) to set 'pass/fail' performance standards in basic TVUS practice utilising contrasting-groups scoring' method.

6.2 Subjects and method

This was a comparative study in which subjects with different levels of ultrasound experience were recruited from Obstetrics and Gynaecology in the Welsh Deanery. The subjects were categorised as experts, intermediates and novices depending on their ultrasound experience. An expert is defined as an independent practitioner who has more than two years ultrasonography experience and conducts daily scanning sessions. An intermediate practitioner is defined as competent but not yet an expert with six months to two years ultrasound experience and who has conducted regular or occasional scan sessions. A novice is defined as a trainee under supervision and has limited ultrasound experience, namely, less than six months with very infrequent scanning sessions e.g. once/month or occasionally.

Participants completed written consent (Appendix II) and the enrolment form (Appendix III). They were asked to perform three different assignments (two in the gynaecology module and one in early pregnancy) on the simulator. There were seven TVUS skills (listed in Table 6.1) included in each assignment and three checklists were used accordingly (Appendix 5.2). The task-specific objective structured assessment of technical skills (OSATS) checklist is based on pass/fail scores where a pass score of (1) is awarded when the skill is correctly undertaken and a fail score of (zero) if the skill is incorrectly undertaken or not attempted (N/A). The score was based on the appropriateness of the hand movement during scanning of each skill regardless of image optimisation. All subjects were briefly introduced to the ScanTrainer® simulator in order to familiarise themselves prior to participation and also informed that the participation would take only 30 minutes to complete.

The contrasting-groups method, an examinee-centred standard setting method was used to determine pass/fail levels (or borderline pass score) of novices (Downing et al., 2006). The standard is the score that best discriminates between the two groups. One of the advantages of this method is that the standard can easily be adjusted to minimise errors in either direction. The performance of novices and experts was recorded and the distribution of the two groups was plotted. The passing score was set at the intersection of the two distributions. The intersection was adjusted to minimise the error of greater concern, i.e. mistakenly categorising a trainee as a “pass” when

they should have failed. The borderline score represents the threshold at which a novice must cross to become competent ((Zieky and Perie, 2004, van Nijlen and Janssens, 2008; Madsen et al., 2014; Konge et al., 2015, Tolsgaard et al., 2015a). The borderline performance is that of minimally competent candidates and their performance is considered safe but not yet independent (GMC, 2016).

The outcome measures consisted of scores generated by the simulator and the observer of the subjects (expert, intermediate and novice) performing basic TVUS as described above. The study was conducted in accordance with the general terms and conditions of the South East Wales Research Ethics Committee SEWREC (NHS REC Reference 10/WSE02/75) approval.

6.3 Statistical data analysis

IBM SPSS Statistics version 20.0 was used for statistical analysis. Data was not normally distributed when tested by Shapiro-Wilk, $p \leq 0.05$. Median scores were calculated for each group. The median values of TVUS performances calculated for comparison between the three study groups' (experts, intermediate and novices) as rated by the simulator and the observer. Statistically significant difference between the two scores obtained by the simulator and the observer was considered at $p \leq 0.05$. Non-parametric tests: Mann-Whitney U used to test the statistical significance between two groups, whereas Kruskal-Wallis test between three groups performing three assignments, with $p \leq 0.05$ being considered significant. The box-plots represented median, first and third quartiles, minimum, maximum and outliers of scores performed by the three study groups (experts, intermediate and novices) in the three assignments: GYN1, GYN2 and early pregnancy. The pass/fail level or borderline pass score was determined by the contrasting-groups method as described above. The overall borderline pass score was also determined by combining the three assignments in order to represent an overall performance of TVUS skills.

6.4 Result

Thirty subjects participated in this study: eight experts, ten intermediate and twelve novices. Participants' demographic and ultrasound experience is detailed in Table (6.2). The median scores as obtained by the observer were 7, 4.5 and zero in GYN1,

and 7, 6.5 and zero in GYN2 and 7, 6, and 1.5 in the early pregnancy assignment for experts, intermediate and novices respectively, $p \leq 0.05$ (Table 6.3). The box-plots in chart (6.1) median, first and third quartiles, minimum, maximum and outliers of scores performed by three groups as rated by the simulator and the observer in the three assignments. The box-plots results of skilled practitioners' (experts and intermediate) performances were similar regardless of whether they were rated by the simulator or the observer. In contrast, the box-plot results for novices showed a significant difference between the simulator's and the observer's ratings. Therefore, the statistical significant difference between the simulator and the observer scores only indicated for novice group, Table (6.4).

Interestingly, the findings, either scored by the observer or the simulator were statistically significant differences between novices and skilled practitioners (experts and intermediate) in the three assignments, $p<0.05$. In the overall analysis of the results there was a significant difference between the three groups. However, there was no statistically significant difference between the experts and intermediate groups, Table (6.4). The results from the total 21 skills included in the three assignments: GYN1/2 and early pregnancy, indicated that experts scored more highly than intermediates and novices. However, the subjects in the three groups scored equally in one skill; detect fetal viability, in early pregnancy assignment. In addition, skilled subjects in experts and intermediate groups scored equally in two different skills in the gynaecology assignments, Table (6.4). The significant difference indicated in 17(81%) skills between novices and intermediate while 19(91%) between novices and experts, as rated by the observer. In terms of simulation rating, there were 12(57%) skills significantly indicated between novices and skilled subjects (ten skills with experts and two with intermediate). Boxplot chart (6.1) revealed that the discrepancy between the observer and the simulator scores was indicated in novices performance only, Mann-Whitney U $p \leq 0.05$.

The performance of novices and experts was recorded and the distribution of the two groups was plotted. The passing score was set at the intersection of the two distributions. The intersection was adjusted to minimise the error of greater concern, i.e. mistakenly categorising a trainee as a "pass" when they should have failed. Chart (6.2) demonstrates the two distribution graphs and cut-off scores of 4, 4 and 5, which

represented the borderline pass was calculated as described above. Chart (6.3) represented the overall skill performance of the combined three assignments to obtain an overall borderline score for performing basic TVUS in early pregnancy and gynaecology.

6.5 Discussion

This study demonstrated that the ultrasound simulator "ScanTrainer®" was able to distinguish between subjects with different levels of TVUS experience. It was also possible to establish a pass/fail or "border-line pass" score for basic TVUS performance using contrasting groups standard-setting method.

Pedersen et al (2014) pointed out that the first step towards widespread use of any given simulator was to demonstrate its true validity when novices and experts performed differently. When assessing the construct validity of the ScanTrainer®, it was necessary to adopt an approach where the tasks reflected the real-life TVUS of the pelvis in gynaecology and early pregnancy. This multi-dimensional approach was taken to strengthen the study findings and demonstrate the standard of performance required from novices to attain competency in basic TVUS.

Experts' scores were higher than intermediate and novice trainees, with a statistically significant difference between the three groups in three assignments. Such differences for individual skills were found between novices and experts and between novices and intermediates but not between experts and intermediates. This variation may reflect the simulator's high discrimination between skilled and non-skilled subjects performing basic TVUS, regardless of the level of experience of skilled practitioners. The inclusion of advanced modules in the assessment may demonstrate significant difference between experts and intermediate subjects as stated in Sketty et al's (2012) study. The authors argued that inclusion of complex modules in laparoscopic surgery instead of than basic ones showed statistically significant differences among the different level of training, specifically between experts and intermediates (Sketty et al., 2012). Similar results were found when assessing the construct validity of Transurethral Resection of the Prostate (TURP) surgical simulator (Bright et al., 2012). However, Bright et al (2012) study showed no statistically significant difference between experts and intermediates in performing capsular resections:

scores were the lowest for the expert participants, which indicated a level of accuracy and control. Verbal comments from the experts suggested that it was easier to cause capsular resection on the simulator than in a real-life TURP (Bright et al., 2012). The lack of significant difference between intermediate and experts on all skills included in the assessment was one of the study's limitations. Alzahrani et al (2013) suggested that if the intermediates and experts were allowed to practice prior to collecting data, their scores could have been significantly different in more tasks. In addition, the author pointed out that the small sample size of experts (n=6) may lead to absence of significant difference. Sample size calculation was based on data from previous studies where thirty participants were considered adequate for testing construct validity between subjects with three different levels of ultrasound experience (Tolsgaard et al., 2012; Tolsgaard et al., 2015a). In addition, several construct validity studies with similar sample sizes have been conducted in different medical specialities, however the level of significance of outcomes was variable (White et al., 2010; Kelly et al., 2012; Aydin et al., 2014; Erdogan et al., 2016).

In this study, there were a number of skills where the intermediate and expert performed similarly such as in skill7: assessing Pouch of Douglas (POD) in gynaecology assignments, which may refer to POD's anatomically fixed location and also to a failure to centralise the POD correctly or to a technical difficulty with the simulator metric. Another task, the assessment of the left ovary in the coronal plane (skill 4) in GYN2, had similar scores by the intermediate group and experts. A possible reason for this similarity lies in the ovary's location, which makes this task easy to pass for skilled practitioners. In early pregnancy assignment, novices, intermediate group and experts correctly performed and scored fetal cardiac activity. The similarity may reflect that the metric is simple and easy to pass or may be caused by the lack of any specific technique for visualising the fetal heartbeat or even specific hand movement to orient and examine the fetal heart, as it is in a fixed location. On the other hand, intermediate subjects outperformed experts in assessing the yolk sac (YS). This is possibly because experts may not apply image optimisation or use adequate magnification to view the yolk sac properly in the centre of the image. It is also possible that the experts were less serious than the intermediate subjects about their participation in the study, given its artificial nature leading to unexpected results. Another possibility as Choudhry et al. (2005) suggest was that

consultants who have been in practice longer may show a decrease in skills in terms of optimisation of the image.

Assessment of trainees' competency using performance standards is challenging and requires devising credible, defensible, and acceptable passing scores for performance-type examinations in real-world settings. Tolsgaard et al. (2015a) noted that previous studies on learning ultrasound have focused on the number of examinations needed for competence, but this may be an unreliable measure of competence, as a result of the wide individual variations in learning curves. Therefore, it is important to address differences in performance standards between novices (non-competent) and experts (competent) by using one of the recognised standard setting methods used to assess practical skills, such as the contrasting-groups method. Performance standard was demonstrated previously by Madsen et al (2014) where median scores of experts and novices determined the pass/fail scores when performing TVUS skills with the ScanTrainer®. The selected simulation assignments in this study were not included in the previous work of Madsen et al (2014) and thus it was critical to determine performance standards specifically to assess competency of novice trainees in the RCT that is reported in chapter seven. The benefit of performance standard in learning TVUS may encourage novices to objectively attain competency using the simulator and work towards a higher level of practice with real scanning. The scores obtained by the simulator were higher than the observer's scores in determining pass/fail level in the three assignments. Therefore, it was considered that the observer' scores were more robust assessment for setting performance standards in evaluating learning curves of novices enrolled in the RCT.

Recently, several studies concerning the validation of basic skills for training TVUS in obstetrics and gynaecology have been published and have assessed construct validity for the ScanTrainer® ultrasound simulator. Madsen et al (2014) reported on the construct validity of 153 simulator metrics for discriminating between different levels of competence in performing TVUS gynaecological scans. The findings showed that only 48 metrics reliably discriminated between different levels of TVUS experience. However, one-third of examined metrics were valid markers but the author argued that even the remaining metrics should not be considered useless. Indeed they may attract attention to important elements of learning procedure namely

that metrics may unduly make assignments easy or difficult to pass and provide only limited information.

Another construct validity study on the use of ultrasound simulators supports this theory and addresses recent calls for valid and reliable assessment instruments to ensure the quality of scans provided by trainees in obstetrics and gynaecology (Tolsgaard et al., 2015a). The authors demonstrated that their Objective Structured Assessment of Ultrasound Skills (OSAUS) scale discriminated between trainees with different level of competence and established credible pass/fail scores for two types of ultrasound examination. Nevertheless, in that study the intermediate subjects gained higher scores in TVUS scans compared to transabdominal (TAS) scans. That may suggest that differences referring to proficiency can be attained rapidly with TVUS rather than TAS. According to Madsen et al (2014) trainees may need to attend three to four hours of simulation training to be fit for supervised clinical training whilst not yet proficient. However, assessment of the learning curves between control and intervention in the RCT reported in the next chapter showed that trainees required longer time to reach competency.

Our study findings support existing theories of simulation construct validity and the development of competency by establishing performance standards in obstetrics and gynaecology ultrasound. Much is still to be learned about the assessment of ultrasound simulation training and thus further validation research is essential.

Chapter 6: Tables and charts

Table (6.1): Description of seven skills in the checklist for the three assignments (GYN1, GYN2 and early pregnancy).

	Task	Task description
GYN1	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas
GYN 2	**Skill 1**	Examining uterus in sagittal plane
	Skill 2	Examining uterus in coronal plane
	Skill 3	Examining left ovary in sagittal plane
	Skill 4	Examining let ovary in coronal plane
	Skill 5	Examining right ovary in sagittal plane
	Skill 6	Examining right ovary in coronal plane
	Skill 7	Examining pouch of Douglas
Early pregnancy	**Skill 1**	Examining gestational sac in sagittal plane
	Skill 2	Examining fetal heart activity
	Skill 3	Examining fetus in sagittal plane
	Skill 4	Viewing yolk sac
	Skill 5	Labelling yolk sac
	Skill 6	Optimise image in viewing yolk sac
	Skill 7	Examining placenta in sagittal plane

Table (6.2): Participants' demographics and ultrasonography experience including: groups, grade, sex, ultrasound experience, and ultrasound scanning sessions.

	No of participants (n=30)
Groups	
Expert	8(27%)
Intermediate	10(33%)
Novice	12(40%)
Gender	
Female	23(77%)
Male	7(23%)
Speciality	
Consultant	4(13%)
Clinical assistant	1(4%)
Radiographer	7(23%)
Midwife	2(7%)
Speciality trainee	9(30%)
Associate specialist	4(13%)
Other	3(10%)
Years of ultrasound experience	
Never	7(23%)
<6months	5(17%)
6-11months	3(10%)
1-2 yrs	7(23%)
>2 yrs	8(27%)
Frequency of ultrasound sessions attended	
Daily	8(23%)
Once/week	4(17%)
Once/month	6(20%)
Never	12(40%)

Table (6.3): Median of the three groups' TVUS performance (expert, intermediate and novice) in the three assignments (GYN1, GYN2 and early pregnancy) as scored by the simulator and the observer. Kruskal- Wallis test shows statistically significant differences between experts, intermediates and novices in the three assignments, $p<0.05$.

	Median		
Observer	**Expert**	**Intermediate**	**Novice**
GYN1	7	4.5	0
GYN2	7	6	0
Early pregnancy	7	6	1.5
Simulator			
GYN1	6	3	2
GYN2	6	5.5	3.5
Early pregnancy	6	5.5	3.5

Table (6.4): The statistical significance between three groups' TVUS performance: expert (n=8), intermediate (n=10) and novice (n=12), in the three assignments (GYN1, GYN2 and early pregnancy) as scored by the observer and the simulator was tested by Kruskal-Wallis, while Mann-Whitney U test used between two different groups.

		Observer				**Simulator**			
		Mann-Whitney U test			Kruskal -Wallis	Mann-Whitney U test			Kruskal -Wallis
		Expert vs Novice	**Expert vs Intermediate**	**Intermediate vs Novice**	**p-value**	**Expert vs Novice**	**Expert vs Intermediate**	**Intermediate vs Novice**	**p-value**
GYN1	Skill 1	0.001	0.3	0.01	0.001	0.09	0.5	0.2	0.1
	Skill 2	0.001	0.1	0.1	0.001	0.004	0.2	0.1	0.002
	Skill 3	0.004	0.8	0.001	0.001	0.02	0.5	0.001	0.001
	Skill 4	0.001	0.1	0.01	0.001	0.001	0.03	0.3	0.001
	Skill 5	0.04	0.6	0.01	0.03	0.3	0.5	0.6	0.2
	Skill 6	0.001	0.3	0.1	0.001	0.04	0.2	0.5	0.6
	Skill 7	0.01	1.0	0.01	0.001	0.3	0.7	0.5	0.2
	Overall	**0.001**	0.1	**0.001**	**0.001**	**0.001**	0.1	**0.004**	**0.001**
GYN2	Skill 1	0.001	0.5	0.001		0.3	1.0	0.3	0.08
	Skill 2	0.001	0.7	0.001	0.001	0.01	0.5	0.06	0.005
	Skill 3	0.01	0.6	0.02	0.001	0.1	0.8	0.08	0.08
	Skill 4	0.001	1.0	0.001	0.007	0.06	0.5	0.2	0.04
	Skill 5	0.001	0.3	0.009	0.001	0.1	0.4	0.5	0.2
	Skill 6	0.001	0.1	0.009	0.001	0.002	0.2	0.1	0.002
	Skill 7	0.04	1.0	0.002	0.001	0.3	1.0	0.3	0.08
	Overall	**0.001**	0.1	**0.001**	**0.001**	**0.001**	0.1	**0.006**	**0.001**
Early pregnancy	Skill 1	0.01	0.7	0.02	0.002	0.03	0.7	0.05	0.006
	Skill 2	0.5	1.0	0.5	0.2	1.0	1.0	1.0	1.0
	Skill 3	0.004	1.0	0.002	0.001	1.0	1.0	1.0	1.0
	Skill 4	0.02	0.6	0.002	0.001	0.02	0.6	0.002	0.001
	Skill 5	0.004	0.1	0.2	0.002	0.1	0.3	0.6	0.1
	Skill 6	0.01	0.8	0.003	0.001	0.01	0.4	0.1	0.01
	Skill 7	0.001	0.7	0.001	0.001	0.02	0.5	0.08	0.01
	Overall	**0.001**	0.1	**0.001**	**0.001**	**0.001**	0.1	**0.001**	**0.001**

Table (6.5): Contrasting-groups method determined by the experts (n=8) and novices (n=12) median value of each assignment. The interaction point between distributions of median for experts and novices indicates the borderline.

		Median		
		GYN1	GYN2	Early pregnancy
The observer	Experts score	7	7	7
	Novice score	0	0	1.5
	Borderline	4	4	5
	Overall skill performance	4		

Chart (6.1): Box-plots represent median, first and third quartiles, minimum, maximum and outliers of scores performed by the three study groups (experts, intermediate and novices) in the three assignments (GYN1, GYN2 and early pregnancy) and rated by the observer (light grey box) and the simulator (dark-stripes grey box). The statistical significance difference between the observer and the simulator indicated only with novice scores.

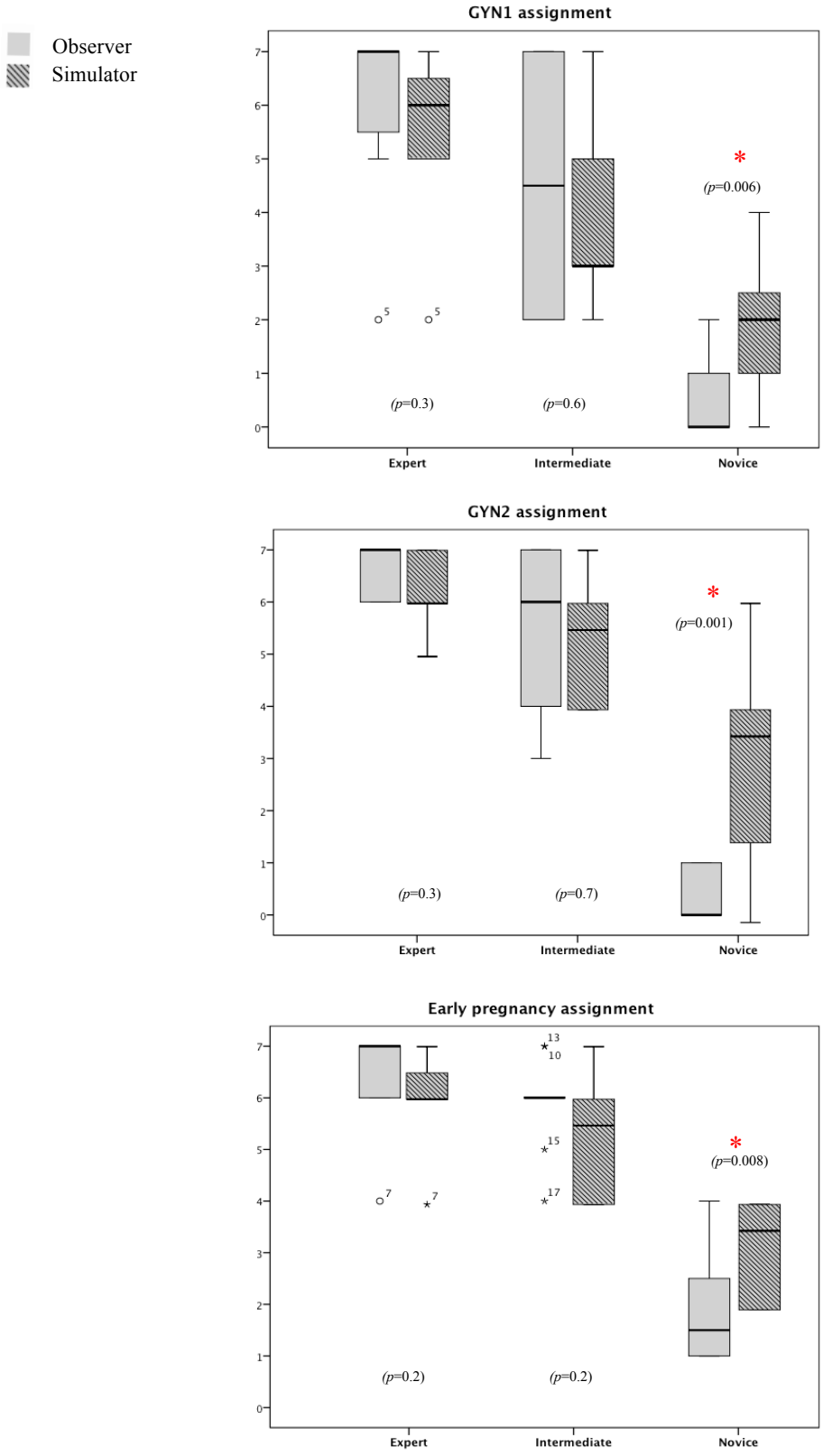
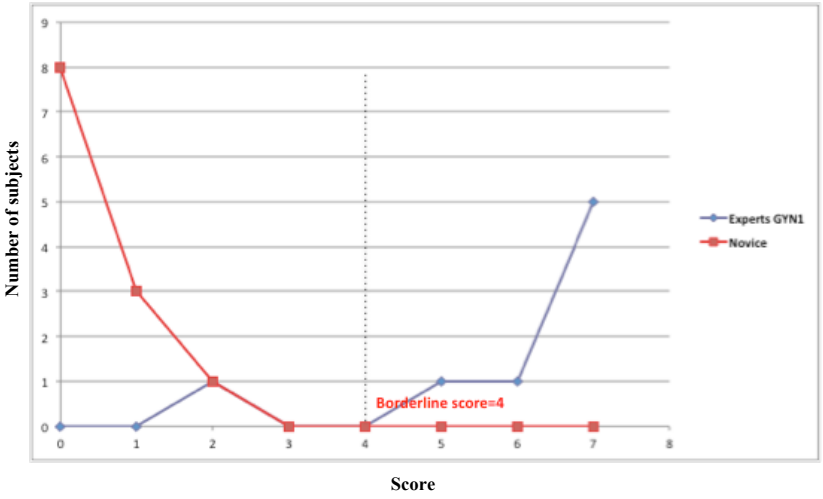
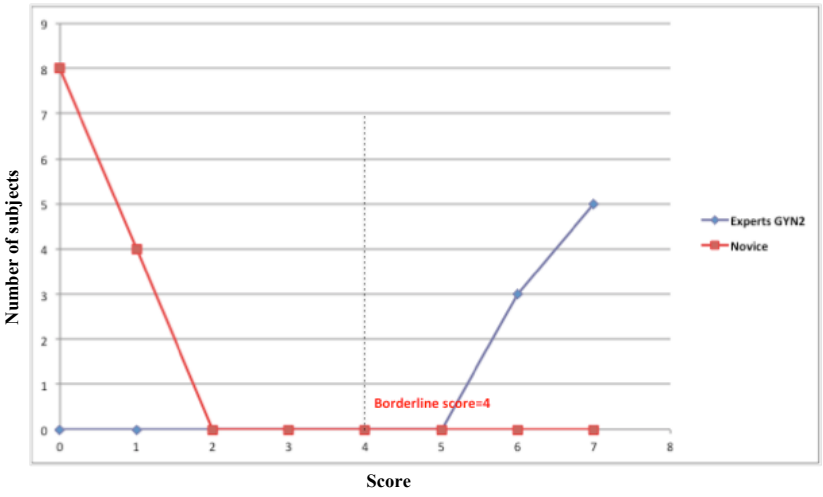


Chart (6.2): Pass/fail level (borderline pass score) for each assignment: GYN1, GYN2 and early pregnancy, was determined by contrasting-groups method. The intersection point between the experts and novices scores represented the borderline score as illustrated by dotted grey line.





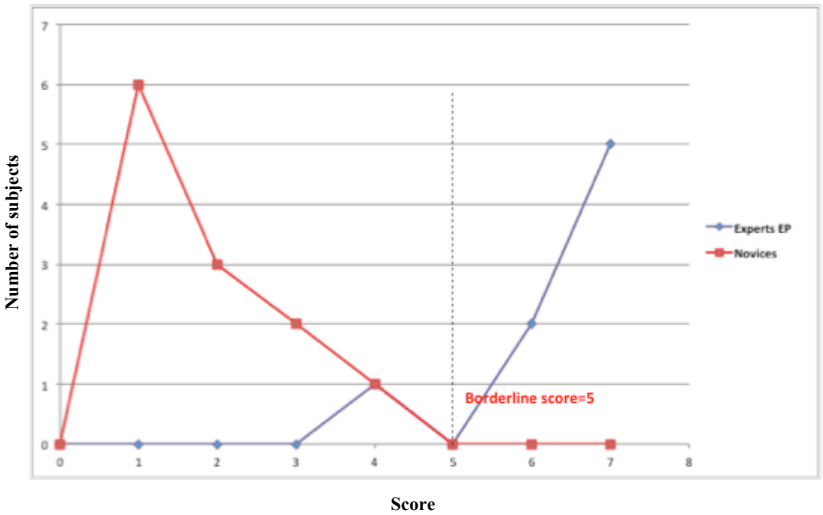
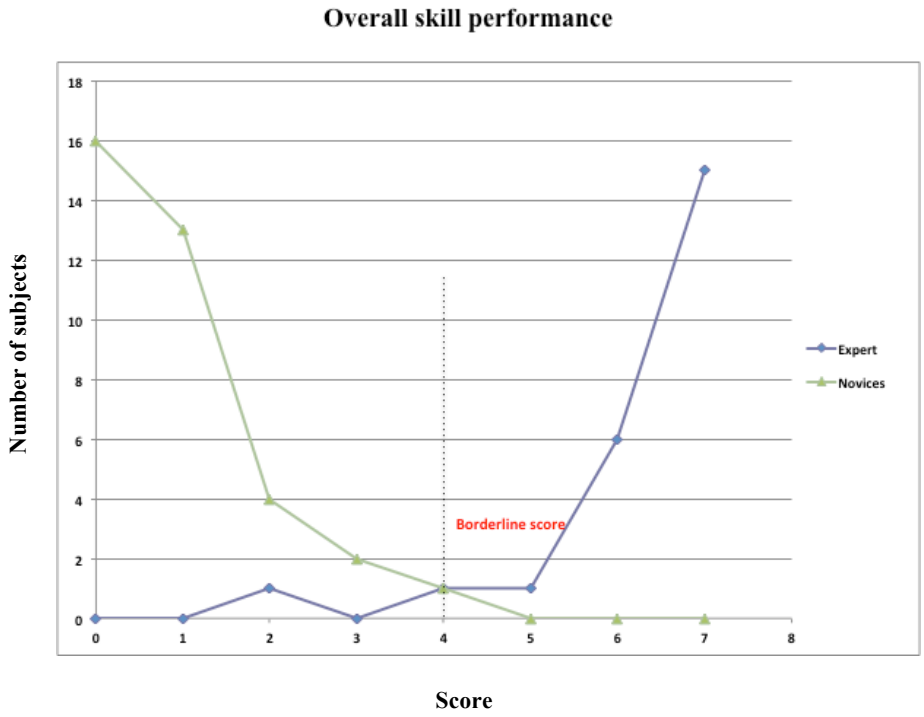


Chart (6.3): Pass/fail level (borderline pass score) for overall skill performance as a result of combined three assignments result. The intersection point between the experts and novices scores represented the borderline as illustrated by dotted grey line.



Chapter seven

Assessment of learning curves: Randomised controlled trial

CHAPTER 7

Assessment of learning curves: Randomised controlled trial

Simulation-based learning plays a significant role in curricula, to enhance the educational process and provide for a better, more efficient learning environment (Modell et al., 2006; Agarwal et al., 2007; Tolsgaard et al., 2015a). In 2009, the Royal College of Obstetricians and Gynaecologists (RCOG) published a comprehensive ultrasound curriculum for trainees in obstetrics and gynaecology, which it envisaged would be delivered through opportunistic conventional training. The ScanTrainer® Ultrasound Simulator was designed and validated by experts in the field as well as three beta test centres in England, to enable the trainees to acquire the core and essential skills outlined in several curricula such as the RCOG and the European Federation of Societies for Ultrasound in Medicine and Biology (EFSUMB), through a series of tasks, assessments and feedback (Madsen et al., 2014, Tolsgaard et al., 2015a). It has been considered as a potentially valid assessment tool that assesses the trainee's incremental acquisition of ultrasound skills in obstetrics and gynaecology (Madsen et al., 2014).

Theoretically, novices require a longer time to attain basic skills properly, compared to skilled/advanced trainees (Kim et al., 2016). With regard to TVUS simulation training, recent publications evaluated learning curves and assessed trainees' skills specifically within a controlled learning environment (Williams et al., 2013; Madsen et al., 2014; Tolsgaard et al., 2015a). However, the long–term effects on learning curves associated with an un-controlled learning environment remains poorly investigated. In this randomised controlled trial (RCT), the emphasis was on evaluating the length of time required for novice trainees to be minimally competent in performing basic TVUS skills when receiving simulation training supplemental to clinical training.

7.1 Aim and objectives

The aim of this study was to determine the length of time required for trainees to acquire the skills necessary to perform TVUS. Trainees were divided into a control group receiving only conventional ultrasound training in their clinical environment and an intervention group, receiving structured simulation training supplemental to the conventional ultrasound training. The objectives were (1) to determine the trainees' duration for acquisition of ultrasound skills with either simulation-supported or conventional training; (2) to explore confounding variables that might have affected the learning curves and (3) to explore the potential factors associated with each point on the learning curve; for example number of training sessions attended and/or engagement with simulation training.

7.2 Subjects and method

In the RCT study subjects were recruited primarily from specialty trainees (STs) in Obstetrics and Gynaecology in the Welsh Deanery, and included trainees at ST1–ST3 level, other NHS staff, and students of the MSc programme at Cardiff University who fulfilled the inclusion criteria. Those subjects had a very limited ultrasound background and were involved in structured hospital training at their centres. The subjects were recruited from South East Wales (Cardiff), South Wales (Swansea) and North Wales (Wrexham).

The subjects were randomly allocated into conventional ultrasound training alone (control) and simulation training supplemental to conventional ultrasound training (intervention).The randomisation codes were generated by computerised allocation software 'Random Allocation Software' (Saghaei M, 2004). The study was reviewed and approved by the South East Wales Research Ethics Committee SEWREC (NHS REC Reference 10/WSE02/75), (Appendix I). In addition, the protocol of this RCT has been registered with Control Clinical Trials with the reference number ISRCTN03408765 with public title of '*The influence of a virtual simulator on the acquisition of trainee's ultrasound skills'* and Scientific title '*Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach*' (Controlled-trials, 2013), (Appendix IV).

7.2.1 Inclusion and exclusion criteria

The inclusion criteria specified that eligible subjects: (1) had no or very limited ultrasound experience of any kind; (2) had no or limited previous access to TVUS practice; (3) were motivated to learn TVUS; (4) intended to complete the requirements of the RCOG ultrasound training curriculum or a similar structured programme; (5) were based in Wales or within a very short travelling distance of any of the three main training hospitals in North and South Wales (Cardiff, Swansea and Wrexham). Subjects who were excluded from the study had at least one of the following criteria: (1) they had already completed a structured ultrasound training programme and been certified accordingly; (2) they were at consultant level in obstetrics and gynaecology or radiology even if they had no ultrasound experience or, (3) were radiographers on the ultrasound MSc programme with previous experience in TVUS.

7.2.2 Sample size

At the time this trial was designed, there was no published literature that determined the length of time needed to acquire ultrasound skills with the conventional approach, which would have determined the end-point of the trial. Hence, the alternative approach to determine the length of time required (in months) to achieve the RCOG objectives, using the conventional training approach, was to seek consensus expert opinion (Appendix 7.1). By their reckoning, the end-point for the acquisition of skills using conventional training was agreed to be within six months. Simulation supported training was hypothesised to assist in reducing this duration to achieve competency in a shorter timeframe. The sample size and power calculation were determined according to Adamchak et al. (2000) assuming a difference of 30–40% (i.e. a shorter duration for skills acquisition in the intervention arm). On the basis of this assumption the estimated sample size was 58 participants in total (29 participants in each arm) to reject the null hypothesis. Additionally, on the basis of a previous pilot study (Morgan et al., 2010) we speculated a dropout rate of 25% thus adjusting the total sample size to about 80 (40 in each arm). This study constitutes a continuous response variable collected from independent control and experimental subjects, with one control per experimental subject. Hence, the study population means that the experimental and control groups are equal with probability (power) 0.8. The Type I error probability

associated with the testing of this null hypothesis was 0.05. Table (7.1) shows the calculation of sample size in this study.

Range of SD	Sample size/group
0.4	29

α	0.05	Type of error probability for two-sided test. This is the probability that we will falsely reject the null hypothesis
δ	0.3 (30%)	Difference in the population mean
σ	0.4	For independent test with group standard deviation.
m	1	Ratio of control to experimental subject
Power	0.8	Probability of correctly rejecting null hypothesis of equal population means

Table (7.1): The program used for power and sample size calculation is PS, (Dupont and Plummer, 1990)

7.2.3 Participation

Potential study subjects were invited to participate by an email that explained the project's objectives along with study information (Appendix II). Participants indicated their preliminary agreement to take part in the study by completing an online survey and provided details of their ultrasonography background. Afterwards, the PhD researcher (A.A.) met each participant for about 30–45 minutes to explain the objectives of the project, the sequence of the assessment sessions and to answer any questions the participant had regarding the project. The participant read the information sheet then signed the consent form. The PhD researcher (A.A.) then assessed each participant using the ultrasound simulator ScanTrainer® to provide baseline information on their skill level. After which the participant was informed of her/his eligibility to enrol in the project according to their baseline scores. Participants were excluded if it became clear that they were experienced in TVUS scanning. At the end of this meeting, a logbook was given to the participant to keep a record of attendance at their ultrasound sessions during the trial and the level of supervision. A simulator account was set up for each participant so they can access the relevant modules according to their randomisation arm.

The intervention group had access to all core simulation modules in "assessed practice" mode, where they were given structured instructions to undertake the

assignments and tasks and receive feedback on their performance throughout the trial. They were given access to the advanced modules after completion of core modules or in case of withdrawal. As instructed by the ethics committee, the control group was given access to the 'un-assessed practice' mode of the modules, which allowed unlimited TVUS practice but did not give instructions or provide feedback on performance. It was hoped that the "un-assessed practice" mode would offer an equal opportunity for those in the control group to practise ultrasound, in order to familiarise themselves with the simulator as well as to compensate for their limited number of clinical sessions. All modules were permitted to those in the control group after completion of the trial or in case of withdrawal.

7.2.4 Trial phases

The trial was divided into three main phases as shown in flowchart (7.1). It is important to note that all participants in both study groups were novices and had similar baseline levels. It was hoped that this similarity would minimise the Hawthorne effect (Lang and Secic, 2006).

7.2.4.0 Phase I - Baseline Phase

This was to record the baseline level of TVUS skill for participants. The structured checklist had been designed previously for the applicant's MSc project (Alsalamah et al., 2009) and supplemented by additional skills based on the published RCOG Objective Structured Assessment of Technical Skills (OSATS), Appendix (5.2). The checklist for GYN1, GYN2 and early pregnancy contained seven basic skills in TVUS and reflected day-to-day practice. The same checklists were used in subsequent assessments.

7.2.4.1 Phase II – Induction program

This phase was a one-day workshop where trainers and trainees were acquainted with the study's objectives, learning material and instructions. At the end of the day, trainers and participants were aware of (1) the project's background; (2) the scientific basis, principles and practical aspects of image acquisition and optimisation; and (3) practical aspects of technique which include orientation, systematic approach and measurements by practising on the mannequin and the ultrasound simulator.

7.2.4.2 Phase III – Trial phase

The trial phase was intended to be completed in six months during which the participants received conventional ultrasound training along with simulation training according to their randomisation as described previously (7.2.3-Participation). In order to monitor participants' skill acquisition and ultimately assessment of the learning curve all subjects were obliged to (1) attend ultrasound training sessions and keep a record of all sessions attended in the logbook; (2) have their ultrasound performance formally assessed by the researcher every 4 – 6 weeks for a six months interval; (3) determine common obstacles that would conflict with conventional ultrasound training either clinical or with simulation; and (4) in the intervention group alone, participants should have completed all core skills tutorials in the simulator in "assessed practice" mode at the beginning of the trial, preferably in the first four weeks after its start.

The three simulation assignments were full examination tasks (FE) and included (1) anteverted uterus (GYN1), (2) retroverted uterus (GYN2) in the gynaecology module, and (3) early pregnancy (EP) in obstetrics module. Further details about these three assignments are found in chapter five (6.2: Subject and methods). Each checklist consisted of seven pass/fail ultrasound skills that were descriptively assigned to the systematic approach to evaluating the female pelvis. The skills reflected participants' performance and a score was calculated at the end of the session. The six assessments were undertaken using the ultrasound simulator.

7.3 Parameters and outcomes measures

The parameters used in describing learning curve were the rate of achievement, length of learning time and final skill level. The outcome was measured in terms of the length of time (in days) and (number of tests) required for control and intervention subjects to reach competence by attaining the borderline pass score (primary end-point) and to attain the experts' level by reaching the maximum score of 7 (final end-point).

7.4 Statistical data analysis

IBM SPSS Statistics version 20.0 was used to collect the data and to conduct the analysis. The Shapiro-Wilk test revealed that the data distribution was not normal, p-

value ≤0.05 as previously mentioned in chapter five. The descriptive data in median scores were collected from: (1) the experts' survey, (2) the logbook and (3) trainees' accounts in the simulator. This collection took place to determine trainees' degree of engagement with ultrasound training either in clinic or with the ultrasound simulator.

In order to repeatedly measure the gradual changes and significance of trainees' performances at different points, from the baseline to the final assessment (sixth test), the repeated measures analysis was used to test any statistically significant differences between subjects in control and intervention, as well as within subjects for each group at each test individually. In addition, data were analysed according to received or actual training during the trial where the subgroups were categorised accordingly and will be reported later. Non-parametric tests: Mann Whitney U (between two groups) and Kruskal-Wallis (between more than two groups) were used to test the significance of gradual changes in performance for subgroups where significance was determined at p-value ≤0.05. The received training/intervention subgroups are categorised in Table (7.2).

The Kaplan-Meier (KM) cumulative hazard analysis is non-parametric estimator used to evaluate trainees' learning curves or achievement at each assessment session during the six tests in the trial. The cumulative hazard analysis model for KM estimated the improvement on the learning curves on the basis of observed data of control and intervention groups and in the subgroup analysis. The KM probability estimated the primary and final endpoint of curves accordingly to (1) borderline pass score of 4 for gynaecology modules and 5 for early pregnancy module and (2) maximum score of 7. Missing data was treated as right-censored, that is to say, observation and data collected from some sessions and missed thereafter were all treated as completed data. In order to estimate the time required to successfully pass one and/or three assignments together, the median values for the control and intervention were calculated according to two approaches: tests and days. The primary analysis of the observed data was according to intention-to-train as per subjects' randomisation. Secondary analysis was carried out according to actual training/intervention received, which resulted into four initial groups then a total of six subgroups on further sub-analysis (Table 7.2). Comparisons of two KM curves statistically tested the null hypothesis using the log-rank test, where the significant considered at p-value ≤0.05.

7.5 Results

7.5.1 Randomisation

In the recruitment phase, of 172 invitations sent to subjects by email, 103 responded by completing the online survey. From the total of 87 subjects who attended the baseline phase, ten subjects did not meet the trial's inclusion criteria and were thus excluded from the study. The mean (SD), median and range of scores for those excluded were 6.8 (0.4), 7 (6-7) for GYN1 and GYN2, while 6.7 (0.6), 7 (5-7) for early pregnancy assignment. Fourteen trainees withdrew from the study after the baseline scan, representing a dropout rate of 22% and occurred between the second and the third assessment sessions, when the participant verbally notified her/his intention to discontinue the study. Flowchart (7.2) illustrates the recruitment at the three training centres: Cardiff, Swansea and Wrexham. A total of 77 subjects were randomised (control=36 and intervention=41). Seven subjects withdrew from each arm resulting in 63 subjects being analysed (control=29, intervention=34). There were 12(19%) male and 51(80%) female participants. Flowchart (7.3) illustrates the randomisation in the RCT.

7.5.2 Length of assessment (expert survey)

Seven consultants in Obstetrics and Gynaecology: three males and four females completed a voluntary survey on basic TVUS training and reported that an estimated period of six months and six assessment sessions were adequate for trainees to attain the basic skills. The survey results are found in Table (7.3). However, as a result of trainees' clinical commitments outside of this study, the six tests or assessment sessions were conducted but not necessarily completed within the six months as suggested by the consultants. Table (7.4) illustrates the median interval (in days) between each test and the baseline for trainees in the control and intervention groups.

7.5.3 Process evaluation

Participants in the control and intervention groups who received regular ultrasound training, had at least one training session per week and had an average of four to five scans in that session, either at gynaecology clinics or in the early pregnancy assessment units. In the intervention arm, trainees were required to complete seven simulation tutorials (four in the gynaecology modules and three in the obstetrics module (listed in Table 7.5). Participants had one access session to the ultrasound

simulator per two weeks and spent 30-60 minutes of uninterrupted practice on it. Each participant recorded all sessions attended and the number of scans in their logbook, while the simulation sessions were recorded under the trainees' account in the simulator. In the intervention group, only six (17%) trainees successfully completed and passed the simulator's core skill modules and certificates of completion were awarded accordingly (Table 7.6).

Not all research subjects received the intended training and/or intervention according to their randomisation. These were due to factors outside the control of the PhD researcher and were largely due to local factors and individual trainee circumstances.

On the advice of the statistician guiding the analysis of this trial, the primary analysis was to be conducted according to intention-to-train principle, i.e. control and intervention subjects. Secondary analysis was to accommodate the heterogeneity of the data and was conducted in four groups initially then six groups at the final sub-analysis as per the table below.

	Four subgroup
Control (n=29)	Group 1=subjects in control did not receive clinical training, (n=20)
	Group 2= subjects in control received clinical training, (n=9)
Intervention (n=34)	Group 3= subjects in intervention received training (simulation and/or clinical training), (n=21)
	Group 4= subjects in intervention group did not receive training (neither simulation or clinical training), (n=13)

	Six subgroup
Control (n=29)	Group 1=subjects in control did not receive clinical training, (n=20)
	Group 2= subjects in control received clinical training, (n=9)
Intervention (n=34)	Group 3= subjects in intervention did not receive any training (neither simulation, or clinical training), (n=10)
	Group 4= subjects in intervention group who did not receive simulation training but had clinical training, (n=3)
	Group 5= subjects in intervention group who received simulation training but no clinical training, (n=10)
	Group 6= subjects in intervention group who received both training: simulation and clinical training, (n=11)

Table 7.2: Subgroup analysis categories

The primary and secondary data analysis was undertaken according to intention-to-train randomisation and subgroup analysis was according the received/actual intervention or training. Repeated measures and Kaplan Meier cumulative hazard tests were applied to investigate the significance of differences between groups as outlined below (Flowchart 7.4).

7.5.4 Repeated measures analysis

The contrasting-groups method was used to standard set the borderline pass scores as described earlier. These were "four" for the overall skill performance of the combined three modules, "four" for GYN1/2 and "five" for early pregnancy assignments. At the baseline, subjects in the control and intervention scored lower than borderline score. At baseline there were no significant differences in the scores between control and intervention groups ($p>0.05$). The mean (SD), median and range for the control and intervention groups respectively were; 0.1(0.5), 0, (0-2) and 0.3(0.7), 0, (0-3) for GYN1, 0.6(1.1), 0, (0-3) and 0.6(1), 0, (0-3) for GYN2, and 1.8(0.8), 2, (0-3) and 1.7(0.8), 1.5, (0-3) for early pregnancy assignment. Their scores gradually increased thereafter. The repeated measures primary analysis estimated training effect on learning curves on the basis of observed data for control and intervention groups, while secondary subgroup analysis estimated training effect for received training/intervention. The results for change from baseline to the sixth test are presented in chart (7.1) for overall skill performance of combined three modules, and charts (7.2-7.4) for GYN1/2 and early pregnancy assignments. In the primary analysis, there were no statistically significant differences between the control and intervention arms (chart 7.1 and table 7.1.A). In secondary subgroup analysis, the differences between the four subgroups (G1, G2, G3, G4) were statistically significant in the overall skill performance and in each assignment analysed individually, $p\leq0.05$ (charts 7.1.1, 7.2.1, 7.3.1, 7.4.1 and tables 7.1.B, 7.2.B, 7.3.B, 7.4.B), In reporting the overall skill performance between subgroups G1, G2, G5 and G6 when compared with protocol determined control and intervention groups (Appendix 7.3), we noted significant differences between subjects in G1: control group who did not receive ultrasound training (n=20) with the intervention group (n=34). Additionally, significance was demonstrated between subjects in G5: intervention group who received simulation training alone (n=10) with the control group (n=29) (Appendix 7.3).

7.5.5 Kaplan-Meier analysis

Kaplan-Meier (KM) curves were constructed to determine the length of time required after randomisation for the control and intervention groups to reach the borderline score and subsequently attain the definitive skill level. The final end-point included successfully achieving a maximum score in performing 21 basic TVUS skills in obstetrics and gynaecology. There were no statistically significant differences in attaining the maximum score for any of the assignments individually or combined (chart 7.5). However the time required to reach the borderline score in the intervention group was significantly shorter than that of the control group in the overall combined, GYN1 and EP only p=0.008, p=0.006 and p=0.04 respectively (charts 7.5-7.8).

In subgroup analysis, two subgroups were eliminated from further data analysis: G3 (intervention group who did not receive any training, ultrasound or simulation, n=10) and G4 (intervention group who received ultrasound training only, n=3) because of the small sample size and intervention received which was not different from control group. In subgroup analysis of the combined modules there were significant differences between G1, G2, G5 and G6 when the interval to reach the borderline score or successfully attain the maximum score of 7 were analysed p=0.0001 and p=0.0001 respectively (charts 7.11 to 7.13). In GYN1 assignment, subgroup analysis showed significant differences between G1, G2, G5 and G6 when the interval to reach the borderline score or successfully attain the maximum score of 7 was analysed p=0.0001 for both (chart 7.14). G6: subjects in the intervention group who received simulation training supplemental to clinical training attained the final endpoint at an earlier date (p=0.04) and lesser number of assessments (p=0.041) but not the primary endpoint compared to G5: intervention and simulation only (Appendix 7.4). There was no statistically significant difference between G2 (control who received clinical training) and G5 (intervention who received simulation training alone) in performing GYN1 assignment, the KM curves probability estimated attainment of skills to final endpoint equally between G2 and G5 at test five and were 169 for G2 and 183 for G5 (Appendix 7.4).

In GYN2, there was a significant difference between G2: control with training and G6: simulation with clinical training, in attaining the primary and final end-points by

the number of tests p=0.038 but not by duration in days, p=0.63. However, there was no significant difference in attaining primary and final end-points in GYN1 and EP between G2 and G6, although there was trend for a shorter duration in group 2. The supplemental results of KM analysis are found in appendix 7.4.

7.6 Discussion

This study was set out with the aim of assessing the length of time required for novice trainees to learn basic TVUS skills during their conventional training supported with structured simulation training in comparison with conventional training only. This RCT is the largest ultrasound simulation training study to date and the findings suggest that those engaged in simulation training supplementary to clinical training (intervention arm) attained the primary end-point significantly faster than those who received clinical training alone (control), in overall skills, GYN1 and EP modules. Nevertheless, the outcome of the two groups was similar at the final assessment for the overall skills and three assignments individually.

In reviewing the literature, no data is available to specify the length of time required for novices to master TVUS performance with simulation training. A recent randomised controlled study explored the effects of TVUS simulation training compared to clinical training alone on the technical quality of scans performed by novices after two months of clinical training and found that the intervention group scores were significantly higher than the control group (Tolsgaard et al., 2015a). This difference demonstrated that simulation training had a substantial influence on subsequent clinical performance that was sustained after two months of clinical training. Another study demonstrated the required number of repeated practices within a controlled learning environment and using the ultrasound simulator ScanTrainer® for the novices to attain an expert level in performing TVUS (Madsen et al., 2014). These earlier studies addressed the time required in hours for novices to attain experts' level (Williams et al., 2013; Madsen et al., 2014). Our study, however, was conducted over a long period of time in a real-world uncontrolled clinical setting and evaluated trainees' skill acquisition according to the number of assessment tests as well as the length of time in days from baseline.

The result of the current RCT study showed that the intervention group attained the borderline score in the overall basic TVUS skills significantly earlier than control group. There was no statistically significant difference between the two groups at the final endpoint. Previous unpublished work showed that simulator training can lead to a similar level of practical skill acquisition as one-to-one sessions in conventional training with mannequins conducted in controlled environment (Morgan et al., 2010). However, the simulator group showed a steeper increase in scores for the first two assessment meetings (out of five) compared to those in conventional training with mannequins. The two groups scored similarly at the final endpoint. These RCT trial results suggest that simulation training would expedite skills acquisition in the early stages of learning TVUS skills. Unlike a controlled environment, it is important to consider the long-term impact on skills acquisition when engaging study subjects in an uncontrolled learning environment, which reflects the real world.

In contrast, Williams et al (2013) determined that ten hours of simulation training were adequate for novices to gain basic TVUS skills and be fit for subsequent, supervised clinical practice, while Madsen et al (2014) suggested that three to four hours of simulation training were adequate in order to attain an expert level. Yet, in the current RCT this was a critical issue to be addressed especially in presence of heterogeneity of subjects within control and intervention arms. Seventeen percent of subjects received structured simulation training and completed the simulation modules, however clinical training was not offered to all of them. Moreover, only nine (31%) subjects in the control group received adequate conventional ultrasound training during the trial, while the majority (n=20, 69%) did not receive conventional ultrasound training.

The overall performance of the two groups showed an increase in the scores from one point to another throughout the six assessments in the trial. As each trainee had a unique individual learning trend and had their assessments as anticipated in the RCT at variable intervals from each other, it was relevant to account for this variability by calculating their attainment in days as well as by the number of assessments. This approach is also supported by Gurusamy et al (2008) who argued that “different trainees have different learning curves for learning different tasks and the time period sufficient to attain proficiency in a task in one individual may not be sufficient for

another individual". The learning trends could not be defined appropriately as it was important to understand the amount of practice required to achieve competency (Chenkin et al., 2015; Kim et al., 2016). In addition to this it was essential to address the potential factors that influenced the learning curves and their heterogeneity thus affecting the study results. Subgroup analysis addressed this heterogeneity but these findings should be interpreted with caution.

Although subgroup analysis was determined after randomisation and resulted in small sample sizes being analysed, the findings were encouraging. Trainees who received additional structured and frequent clinical training whether they were assigned to the control or the intervention group attained maximum score in performing TVUS skills chart 7.11-7.12. However, there was no statistically significant difference between control subjects who received clinical training and subjects who received simulation training alone, thus demonstrating that structured simulation training was beneficial and substantial in learning TVUS skills similar to clinical training. Similar findings were reported by Williams et al (2013). When assessing the acquisition of skills among the subgroups, statistically significant differences were only shown when there were enough sample numbers with sufficient differences in scores to compare at different stages of the assessment. For instance, subjects who received simulation training alone were able to reach primary endpoint (borderline pass) and pass the three modules in shorter time compared to other groups.

According to Tolsgaard et al (2015a), although the participants in the intervention group varied in simulator scores and amount of time they required to achieve an expert performance level, there were no significant correlations between performance measures in the simulated setting and the clinical setting. The authors argued that the low predictive validity of simulator metrics may indicate that the sample size was inadequate to establish a correlation between performance in a simulated and clinical setting due to dilution of differences in individual performance after two months of clinical training. One of the limitations of the current RCT, the largest ultrasound simulation training study to date, is the relatively small numbers included in the subgroup analysis. Unfortunately, this is often the case in medical education research due to feasibility issues and scarcity of participants suitable for inclusion. For

example, the only randomised study performed on virtual reality ultrasound simulator included thirty participants only (Tolsgaard et al., 2015a).

Notwithstanding this, the findings suggest that simulation training added to clinical training may have some beneficial effects on skill acquisition in comparison with clinical training alone. Similarly, Williams et al (2013) showed no statistically significant difference between intervention and control groups and that may be due to the lack power due to small sample size of eleven participants. In determining of sample size for this RCT we ensured that the study had adequate power to detect statistical difference between the two study groups, including taking account of a potential drop-out rate of 25%.

Most studies comparing learning curves between novices and experts reported the number of repetitions required for novices to attain experts level (Eversbusch and Grantcharov, 2004; Grober et al., 2010; Madsen et al., 2014; Lucereau et al., 2016). However, task repetition was not the key to mastering competency, as all trainees included in our study were novices. This may be considered to be an interesting step in the global evaluation of a simulator, which could potentially lead to a generalization of skill acquisition (Sánchez- Peralta et al., 2012). Virtual-reality simulator training alone can only replace the initial part of the learning curve and the results confirm that trainees should not be considered fully competent after training on a virtual reality simulator (Konge et al., 2015). The authors proposed a three-step approach consisting of learning the necessary anatomy and theory (step one), simulation-based training (step two), and supervised practice on patients (step three), before performing independent procedures. Testing can ensure basic (early phase) competency and has been shown to accelerate learning and improve performance. Thus, the authors propose that all three steps should end with a test of competence before proceeding to the next step (Konge et al., 2015; Lucereau et al., 2016).

Several European training programs in obstetrics and gynecology rely on time spent in specialized ultrasound units to avail sufficient conventional ultrasound training, for trainees to achieve the skills necessary to practice (Tolsgaard et al., 2014b). Salvesen et al (2010) pointed out that the European Federation Societies for Ultrasound in

Medicine Biology (EFSUMB) minimum training sessions recommended for obstetrics and gynaecological ultrasound practitioners would be twenty-five supervised examinations in three months, for trainees to reach competency. The International Society of Ultrasound in Obstetrics and Gynaecology (ISUOG) and the American Institute of Ultrasound in Medicine (AIUM) suggested that in a conventional ultrasound training environment 300 examinations must be performed under supervision within three years of training for best practice and to enhance trainees' self-confidence in ultrasound scanning (ISUOG, 2014). On the other hands, the Royal College of Radiologists (RCR) reported that that ultrasound practical training required at least one practical session per week over a period of 3-6 months with approximately 5-10 examinations performed under supervision for trainees to be competent (RCR, 2005).

To date, there has been little evidence to guide educators in terms of how much time to allocate to simulation training. Nevertheless, based on Tolsgaard et al (2014b), it was suggested that a minimum of 12–24 days of conventional training in specialized ultrasound units was associated with confidence in performing ultrasound scans independently. Yet, there are no definitive useful parameters to describe level of competency in terms of assessing the rate, length of time and final skill level in ultrasound simulation training.

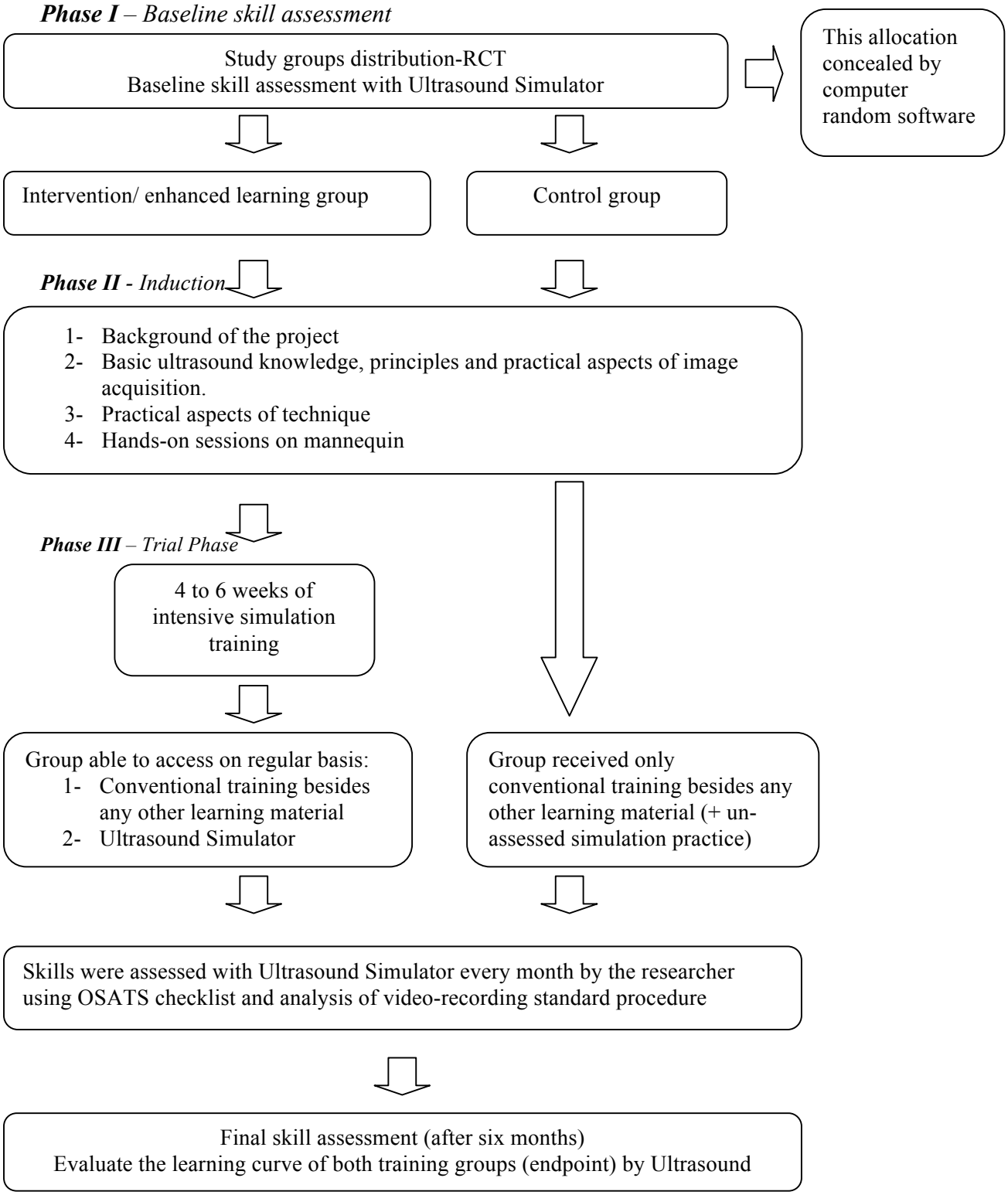
In the current RCT, data collected from trainees regarding regular ultrasound training sessions attended, indicate that the average number of sessions attended was one session every six weeks with 2-3 supervised scans achieved at each particular session. Previous studies in other areas of medicine have consistently shown substantial, immediate effects of simulation-based training when compared with no training (Tolsgaard et al., 2015a). In fact, the clinically trained learner did fewer than the recommended number of cases or practice sessions for competency (Site et al., 2004; Stather et al., 2011), thus individual differences may result in inadequate levels of performance by some trainees, and the unnecessarily long training of others (Tolsgaard et al., 2015a). The initial step to overcome this obstacle as it was anticipated prior to conducting the study, was to create an un-assessed practice mode in the simulator which allowed those in the control group to practise with no structured instructions or feedback provided, in order to offer an equal opportunity of

training sessions when there were limited clinical sessions in the hospitals. However, none of the trainees in the control group had accessed the simulator. Trainees who attended frequent and structured clinical ultrasound training sessions averaged one session per a week with 4-5 supervised scans and performed better than other trainees who didn't. This number may correspond to EFSUMB recommendations. It is important to note that the majority of these trainees were enrolled on an ultrasound educational program with scheduled training.

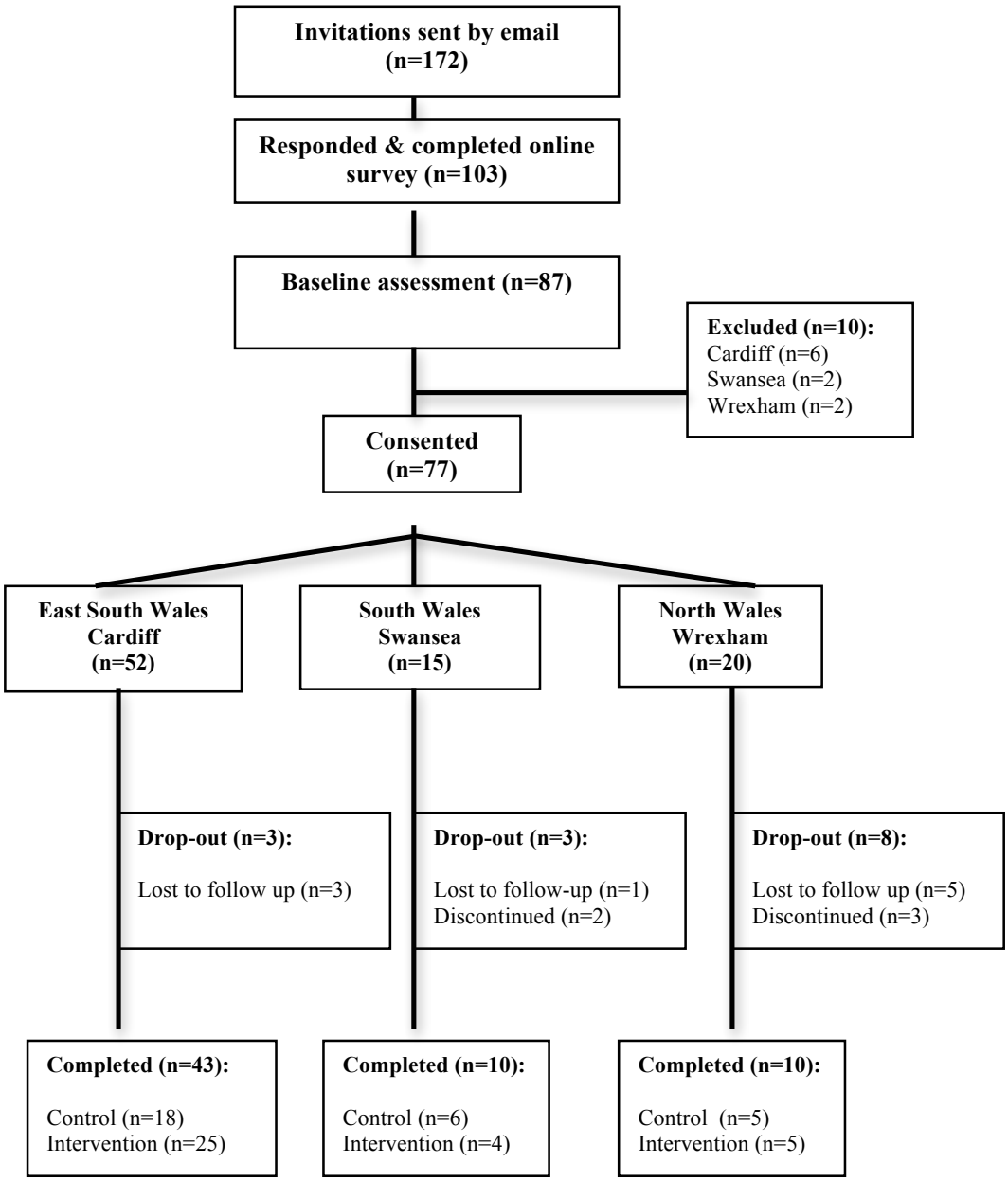
This RCT has a number of limitations and there were a number of factors that may have affected the significance of the findings. Limited access to simulation location, unprotected time given to trainees to practise on the simulator and lack of clinical sessions scheduled for trainees had an impact on the learning curves. Data extracted from the intervention group showed that a minority of trainees (six trainees) successfully completed the basic modules on the ultrasound simulator and it would have been far more beneficial if all participants had implemented the intervention positively. Recent studies suggest that for simulation training to be successful it must be integrated within the curriculum and made mandatory (Williams et al., 2013; Madsen et al., 2014). Consequently, with no specific guidelines available on how to implement simulation training within the ultrasound curriculum, simulation training remains a challenge. The findings of this study suggest that simulation training when additional to clinical ultrasound training would have a considerable effect on shortening the time required to learn basic TVUS and has beneficial and positive impact on the acquisition of skills hence benefitting the design and development of ultrasound training curricula.

Chapter 7: Tables and figures

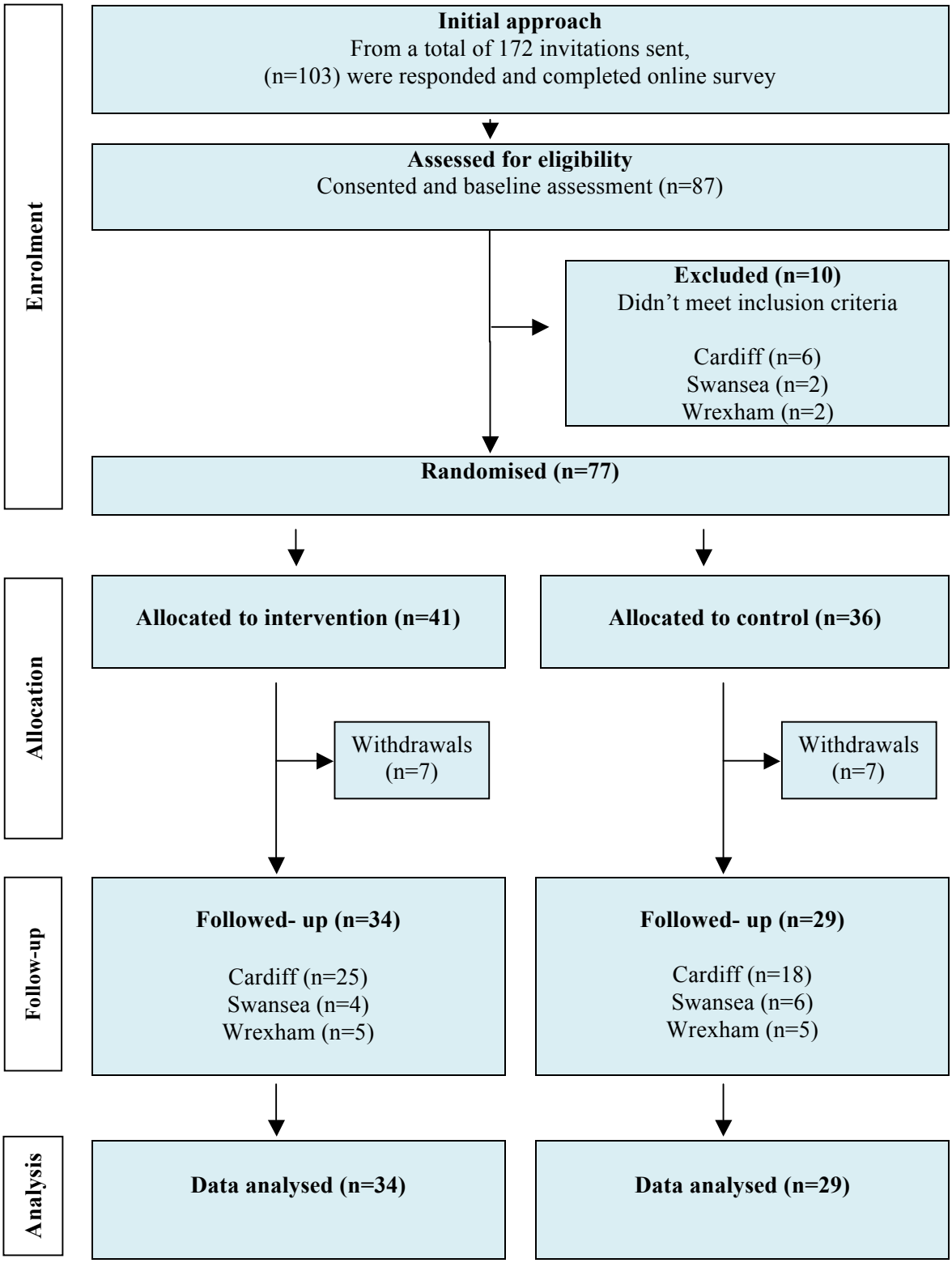
Flowchart (7.1): Diagram illustrates the three trial's phases.
Objective Structured Assessment of technical Skills (OSATS).



Flowchart (7.2): Flow chart of recruitment and participation in the three training areas within Wales.



Flowchart (7.3): illustrates the randomisation in the RCT



Flowchart (7.4): Sequence of primary and secondary data analysis of repeated measures and Kaplan Meier for estimating skill performances of intention-to-train and subgroups

Skill performance data analysis

Primary analysis

Intention-to-train

control (n=29) vs intervention (n=34)

Repeated Measures analysis	
Control vs intervention	by tests
Overall skill performance	NS
GYN1	NS
GYN2	NS
Early pregnancy (EP)	NS

Kaplan Meier analysis				
Control vs intervention	Primary end-point (Borderline pass)		Final end-point (Maximum score)	
	in days	by tests	in days	by tests
Overall skill performance	Sig.	NS	NS	NS
GYN1	Sig.	NS	NS	NS
GYN2	NS	NS	NS	NS
Early pregnancy (EP)	Sig.	NS	NS	NS
Pass (one) module	NS	NS	NS	NS
Pass (three) modules	Sig.	Sig.	NS	NS

Secondary analysis

Four subgroup analysis

G1:(C-CT,n=20), G2:(C+CT,n=9), G3:(intervention+T,n=21), G4:(intervention-T,n=13)

Four subgroups: G1, G2, G3, G4		by tests
Overall skill performance		Sig.
GYN1		Sig.
GYN2		Sig.
EP		Sig.
Individual groups		
GYN1	G1 vs G2	Sig.
	G1 vs G3	Sig.
	G1 vs G4	NS
	G2 vs G3	NS
	G2 vs G4	NS
	G3 vs G4	NS
GYN2	G1 vs G2	NS
	G1 vs G3	NS
	G1 vs G4	NS
	G2 vs G3	NS
	G2 vs G4	NS
	G3 vs G4	NS
EP	G1 vs G2	Sig.
	G1 vs G3	Sig.
	G1 vs G4	NS
	G2 vs G3	NS
	G2 vs G4	NS
	G3 vs G4	NS

Four subgroups: G1, G2, G3, G4	in days	by tests	in days	by tests
Overall skill performance	Sig.	Sig.	Sig.	Sig.
GYN1	Sig.	Sig.	Sig.	NS
GYN2	Sig.	Sig.	Sig.	Sig.
EP	Sig.	NS	Sig.	NS
Pass (one) module	Sig.	NS	Sig.	NS
Pass (three) modules	Sig.	Sig.	Sig.	Sig.
Individual groups				
G1 vs G2	Sig.	Sig.	Sig.	Sig.
G1 vs G3	Sig.	Sig.	Sig.	Sig.
G1 vs G4	NS	NS	NS	NS
G2 vs G3	NS	NS	NS	NS
G2 vs G4	NS	NS	NS	NS
G3 vs G4	Sig.	NS	Sig.	NS
G1 vs G2	Sig.	Sig.	Sig.	Sig.
G1 vs G3	Sig.	Sig.	Sig.	Sig.
G1 vs G4	NS	Sig.	NS	Sig.
G2 vs G3	NS	NS	NS	NS
G2 vs G4	NS	NS	NS	NS
G3 vs G4	Sig.	NS	NS	NS
G1 vs G2	NS	NS	NS	NS
G1 vs G3	Sig.	NS	Sig.	Sig.
G1 vs G4	NS	NS	NS	NS
G2 vs G3	NS	NS	NS	NS
G2 vs G4	NS	NS	NS	NS
G3 vs G4	Sig.	NS	NS	NS

Secondary analysis six subgroups

G1: (C-CT,n=20), G2: (C+CT,n=9) , G5: (+SIM-CT,n=10), G6: (+SIM+CT,n=11)

[groups G3 and G4 not included in the analysis]

Repeated Measures analysis

Six subgroup: G1, G2, G5, G6	by tests
Overall skill performance	Sig.
GYN1	Sig.
GYN2	Sig.
EP	Sig.

Subgroup vs Intention-to-train	
G1 vs intervention	Sig.
G2 vs intervention and control	NS
G5 vs control	Sig.
G6 vs intervention and control	NS

Individual groups		
GYN1	G1 vs G2	Sig.
	G1 vs G6	Sig.
	G3 vs G1, G4, G5, G6	NS
	G4 vs G5, G6	NS
	G5 vs G6	NS
GYN2	G1 vs G2, G3, G4, G5, G6	NS
	G2 vs G3, G4, G5, G6	NS
	G3 vs G4, G5, G6	NS
	G4 vs G5, G6	NS
	G5 vs G6	NS
EP	G1 vs G2	Sig.
	G1 vs G6	Sig.
	G3 vs G1, G2, G4, G5, G6	NS
	G4 vs G5, G6	NS
	G5 vs G6	NS

Kaplan Meier analysis

Six subgroup: G1, G2, G5, G6	Primary end-point (Borderline pass)		Final end-point (Maximum score)	
	in days	by tests	in days	by tests
Overall skill performance	Sig.	Sig.	Sig.	Sig.
GYN1	Sig.	Sig.	Sig.	Sig.
GYN2	Sig.	Sig.	Sig.	Sig.
EP	Sig.	NS	Sig.	NS
Pass (one) module	Sig.	NS	Sig.	NS
Pass (three) modules	Sig.	Sig.	Sig.	NS

Individual groups				
G1 vs G5	Sig.	NS	Sig.	NS
G1 vs G6	Sig.	Sig.	Sig.	Sig.
G2 vs G5	NS	NS	NS	NS
G2 vs G6	NS	NS	NS	NS
G5 vs G6	NS	NS	Sig.	Sig.
G1 vs G5	Sig.	Sig.	Sig.	Sig.
G1 vs G6	Sig.	NS	Sig.	Sig.
G2 vs G5	NS	NS	NS	NS
G2 vs G6	NS	Sig.	NS	Sig.
G5 vs G6	NS	NS	NS	NS
G1 vs G5	Sig.	NS	Sig.	NS
G1 vs G6	Sig.	NS	Sig.	Sig.
G2 vs G5	NS	NS	NS	NS
G2 vs G6	NS	NS	NS	NS
G5 vs G6	NS	NS	NS	NS

Sig.=significant, NS= not significant

Table (7.3): Global rating scale of experts' survey result

Survey statements	Median	%
5. How strongly do you agree/disagree that the conventional training (i.e. adhoc or locally organised training as per local practice),which is supported by the addition of an Ultrasound Simulator could improve the quality of training (better trainee's skills and expedited attainment of skills)?	Agree	57%
6.How strongly do you agree/disagree that conventional training alone would improve the trainee's skills over a long period of time?	Agree	100%
7. How strongly do you agree/disagree that beginners undertaking their training in the conventional way would acquire core ultrasound skills in a relatively short period	Neutral	43%
8. Could you indicate how many conventional training sessions (per day/per week/per month) are needed by a trainee to acquire the core skills in a relatively short period time without additional learning material, mannequin or simulator support?	Two weeks per a month	86%
9. Could you indicate how many conventional training sessions (per day/per week/per month) are needed by a trainee to acquire the core skills in a relatively short period time if supported by additional learning material, mannequin or simulator?	Two sessions per a week	57%
10. Have you ever heard about the effect of introducing the simulation to support the learning curve of ultrasound skills in short time?	Yes	86%
11. If you were a trainee, which of the training methods you prefer to have for your training purposes	Conventional training supplemented to simulation training	86%
12. Taking your own institution circumstances and as an expert in ultrasound, how long would a trainee in your institution take to achieve competency in core ultrasound skills utilising current conventional training resources?	9 months – one year	100%
13. In your opinion as an expert, how long should it take for the trainee to achieve competency in core ultrasound skills when conventional training is supplemented with simulation training?	Six months	57%

Table (7.4): Median interval (in days) between each test and the baseline for subjects in the control (n=29) and intervention (n=34) groups.

Median (Days)	Test 2	Test 3	Test 4	Test 5	Test 6
Control	40 (23-435)	126 (48-484)	185 (71-518)	225 (93-821)	248 (125-877)
Intervention	39 (21-212)	108 (40-280)	162 (82-326)	194 (120-371)	221 (154-453)

Table (7.5): Simulation learning tutorials and assignments as listed in the ultrasound simulator ScanTrainer®

	Module 1. Basic Gynaecological Skill: Full examination of the uterus		
Tutorial	**Description**	**Assignments**	**Attempts /pass**
T2	Orientation conventions	2.1 orientation in the sagittal plane	
		2.2 orientation in the coronal plane	
T3	Introduce the probe	3.1 direction and positioning	
		3.2 pressure	
T4	Examination of the uterus in sagittal plane	4.1 optimal demonstration of the uterus	
		4.2 optimal assessment of the uterine cavity	
		4.3 full procedure for assessment of the uterus in sagittal plane	
T5	Examination of the uterus in coronal plane	5.1 optimal demonstration of the uterus in the coronal plane	
		5.2 full procedure for the assessment of the uterus in coronal plane	
T6	Full examination of the uterus	6.1 full procedure for the assessment of the uterus in both planes	
T7	Examination of the ovaries and adnexa	7.1 examination o f the right ovary and adnexa	
		7.2 optimal assessment of the right ovary and adnexa	
		7.3 optimal assessment of the left ovary and adnexa	
T8	Final examination	8.1 full examination (anatomy revealed)	
		8.2 full examination (anatomy concealed)	
T9	Retroverted uterus	9.2 optimal assessment of the uterine cavity	
		9.3 full examination (anatomy revealed)	
		9.4 full examination (anatomy concealed)	
	Module 4. Basic Obstetrics Skills: Examination of the gestational sac		
Tutorial	**Description**	**Assignments**	**Attempts /pass**
T2	Orientation conventions	2.1 orientation in the sagittal plane	
		2.2 orientation in the coronal plane	
T3	Identification of the uterus	3.1 identification of the uterus in sagittal plane	
	Examination of the gestational sac	4.1 Imaging the gestational sac	
T4	Examination of fetus	5.1 identifying fetal heart beat (case2)	
		5.2 measuring CRL (case2)	
		5.3 assessing a sub-optimally positioned fetus	
T5	Examination of other structures	6.1 identify yolk sac	
		6.2 identify placenta	
T6	Final examination	7.1 full examination (anatomy revealed)	
		7.2 full examination (anatomy concealed)	
		7.3 case 2 (anatomy revealed)	
		7.4 case 2 (anatomy concealed)	

Table (7.6): Median of frequent training sessions attended by all participants (n=63) as recorded in trainees' logbooks.

Statement	Median
Frequency of clinical sessions attended	One per six weeks
Frequency of supervised ultrasound scans per a session	2-3 per a session
Frequency of accessing the simulation	One per four weeks
Time spent in practising TVUS with the simulator	27 minutes per session
Data extracted from control with training (n=9): Clinical sessions attended	1-2 sessions per two weeks
Number of supervised ultrasound scans per a session	4-5 cases per a session
Data extracted from six trainees successfully completed and passed basic modules in the simulator: Frequent access to simulation	Seven times
Time spent in practising TVUS with the simulator	1hr: 21 minutes per session
Average of total time spent from first to last session	6hrs: 15minutes

Chart (7.1): Plots of repeated measures result for two groups: intervention (n=34) and control (n=29) represent the overall skill performance of average score of combined three modules: GYN1/2 and early pregnancy.

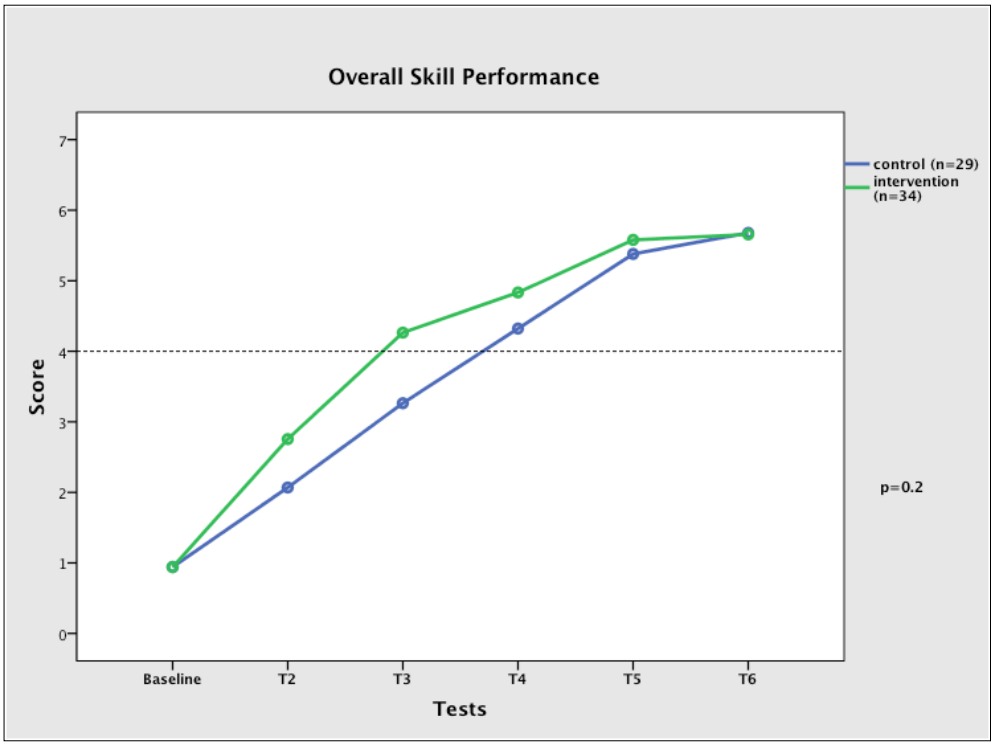


Table (7.1.A): Significance differences between intervention and control learning curves as tested by repeated measures ANOVA.

Control vs intervention	p-value
Baseline - test 2	0.053
Test 2- test 3	0.4
Test 3- test 4	0.1
Test 4- test 5	0.5
Test 5- test 6	0.2
Overall	0.2

Chart (7.1.1): Plots of repeated measures result presented the overall skill performance of average score of combined three modules: GYN1/2 and early pregnancy for four subgroup: G1: control did not receive clinical training (n=20), G2: control received clinical training (n=9), G3: intervention received any type of training: simulation and/or clinical training (n=21) and G4: intervention did not receive any training (n=13).

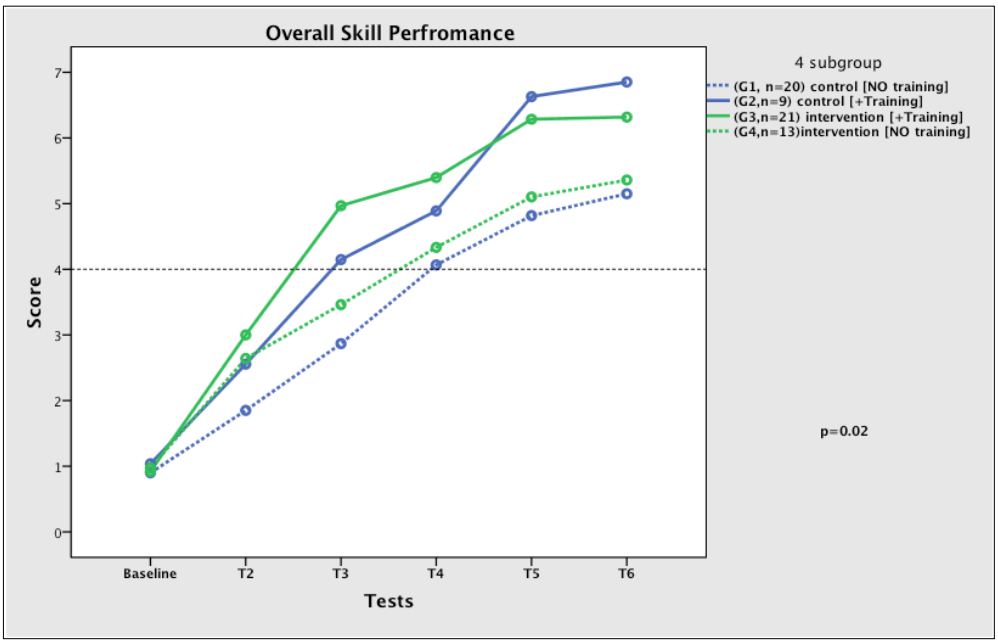


Table (7.1.B): Significance differences between four subgroup's learning curves as tested by repeated measures ANOVA

Between subjects in all four groups G1, G2, G3 and G4	**p-value**
Baseline - test 2	0.07
Test 2- test 3	0.1
Test 3- test 4	0.2
Test 4- test 5	0.6
Test 5- test 6	0.5
Overall	0.02

Chart (7.1.2): Plots of repeated measures result presented the overall skill performance of average score of combined three modules: GYN1/2 and early pregnancy for six subgroup: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G3 (-Sim-CT): intervention did not receive any type of training: no simulation, no clinical training (n=10) and G4 (–Sim+CT): intervention received clinical training alone (n=3), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).

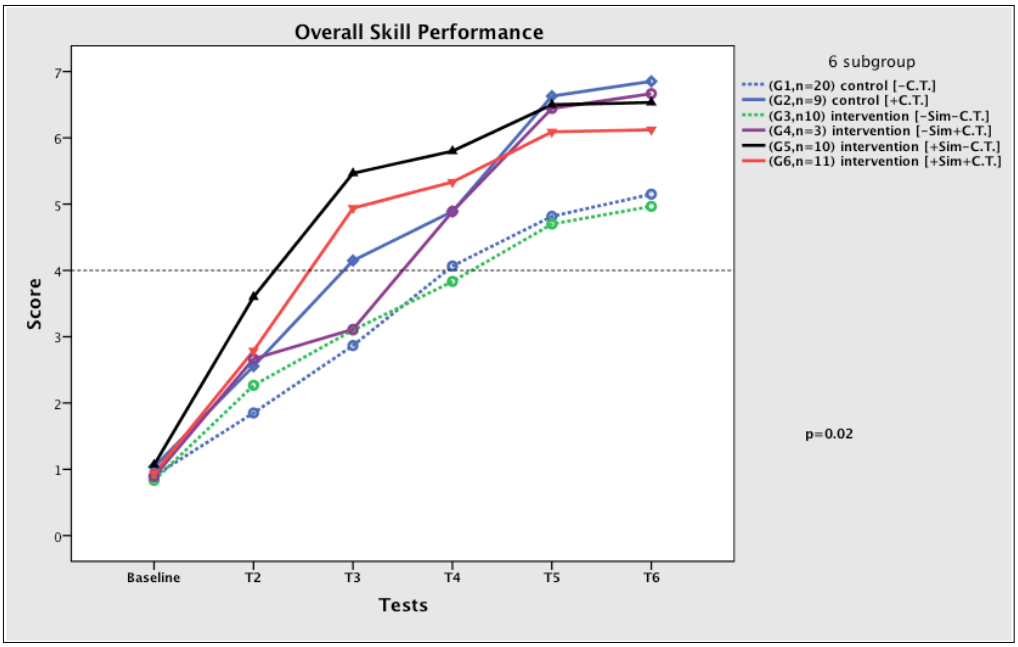


Table (7.1.C): Significance differences between six subgroups' learning curves as tested by repeated measures ANOVA

Six subgroup **G1, G2, G3, G4, G5, G6**	**p-value**
Baseline - test 2	0.07
Test 2- test 3	0.1
Test 3- test 4	0.2
Test 4- test 5	0.6
Test 5- test 6	0.5
Overall	0.02

Chart (7.2): Plots of repeated measures analysis result for intervention (n=34) and control (n=29) represent result of GYN1 assignment.

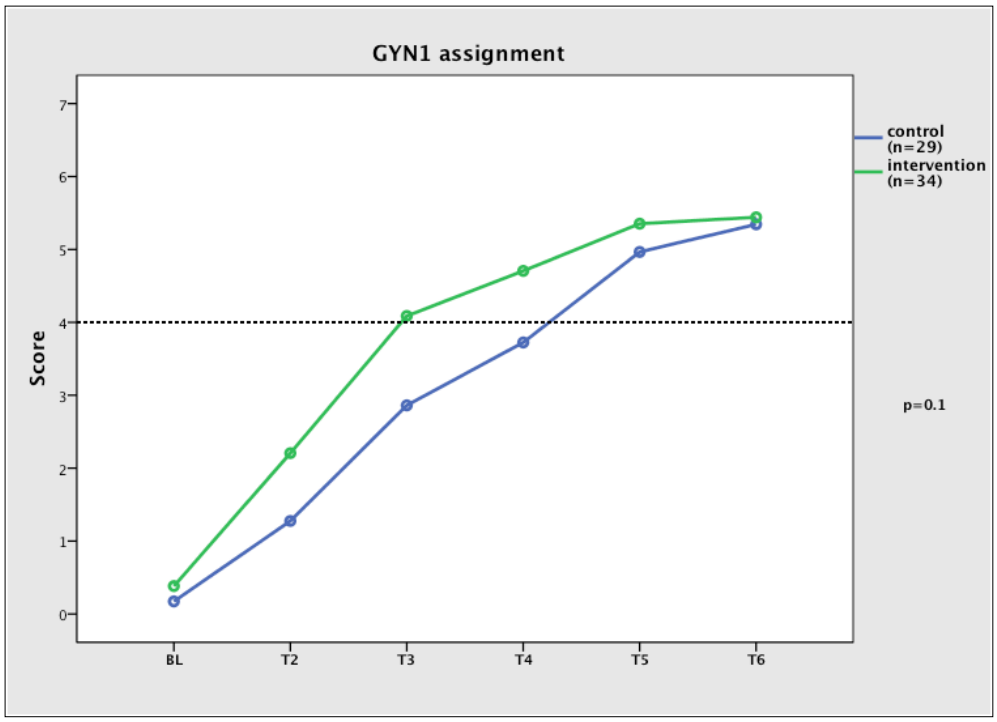


Table (7.2.A): Significance differences between intervention and control learning curves was tested by repeated measures ANOVA

	p-value		
	Within tests between control and intervention (TESTs)	within tests between subjects in (control)	within tests between subjects in (intervention)
At baseline	0.1		
Baseline - test 2	0.050	0.001	0.001
Test 2- test 3	0.6	0.006	0.001
Test 3- test 4	0.6	0.01	0.003
Test 4- test 5	0.3	0.01	0.1
Test 5- test 6	0.2	0.03	0.04
Overall	0.1		

Chart (7.2.1): Plots of repeated measures result presented GYN1 assignment for four subgroup: G1: control did not receive clinical training (n=20), G2: control received clinical training (n=9), G3: intervention received any type of training: simulation and/or clinical training (n=21) and G4: intervention did not receive any training (n=13).

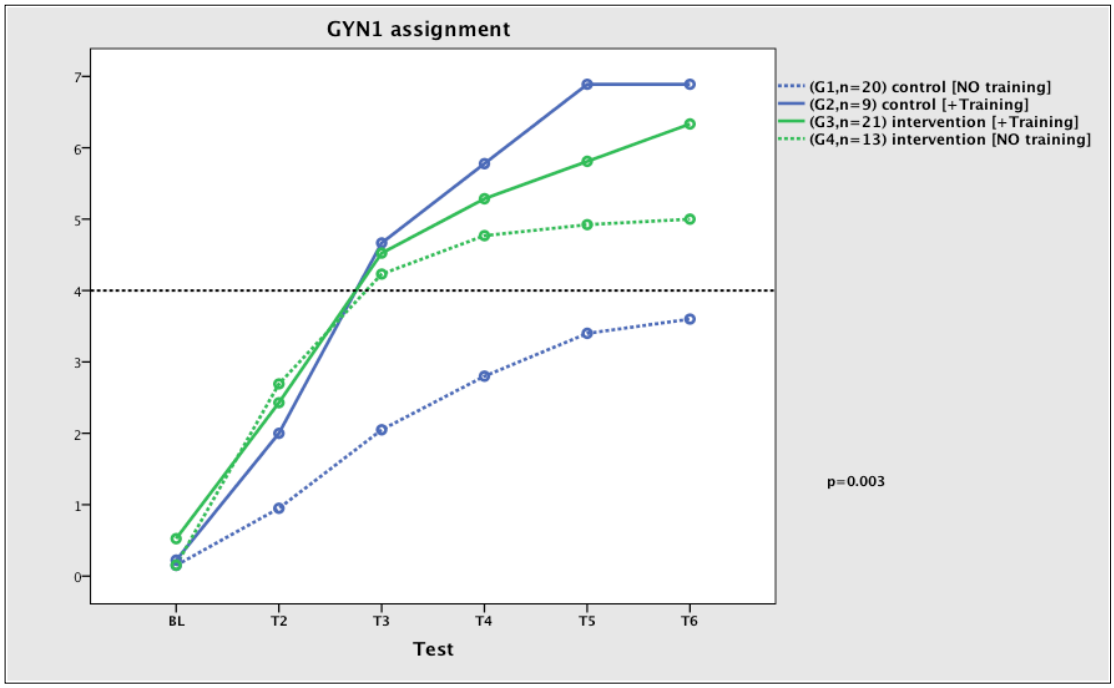


Table (7.2.B): Significance differences for the four groups was tested by repeated measures ANOVA.

		within tests, p-value			
	Between four groups (overall)	G1 Control-CT	G2 Control+CT	G3 Intervention +CT	G4 Intervention - CT
Baseline	0.2				
Baseline - test 2	0.1	0.01	0.01	0.001	0.005
Test 2- test 3	0.3	0.02	0.01	0.001	0.02
Test 3- test 4	0.8	0.06	0.08	0.004	0.2
Test 4- test 5	0.5	0.1	0.4	0.07	0.7
Test 5- test 6	0.2	0.1	0.2	0.04	0.6
Overall	0.003				

Non-parametric statistical significance difference between four subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	X	0.03	0.4	0.01	0.03
G1(control-CT) vs G3 (intervention+CT)	X	0.02	0.004	0.01	0.02	0.01
G1(control-CT) vs G4 (intervention-CT)	X	0.03	0.04	X	X	X
Kruskal-Wallis test (between four groups)	0.2	0.06	0.01	0.05	0.02	0.02

Chart (7.2.2): Plots of repeated measures result presented GYN1 assignment for six subgroup: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G3 (-Sim-CT): intervention did not receive any type of training: no simulation, no clinical training (n=10) and G4 (–Sim+CT): intervention received clinical training alone (n=3), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).

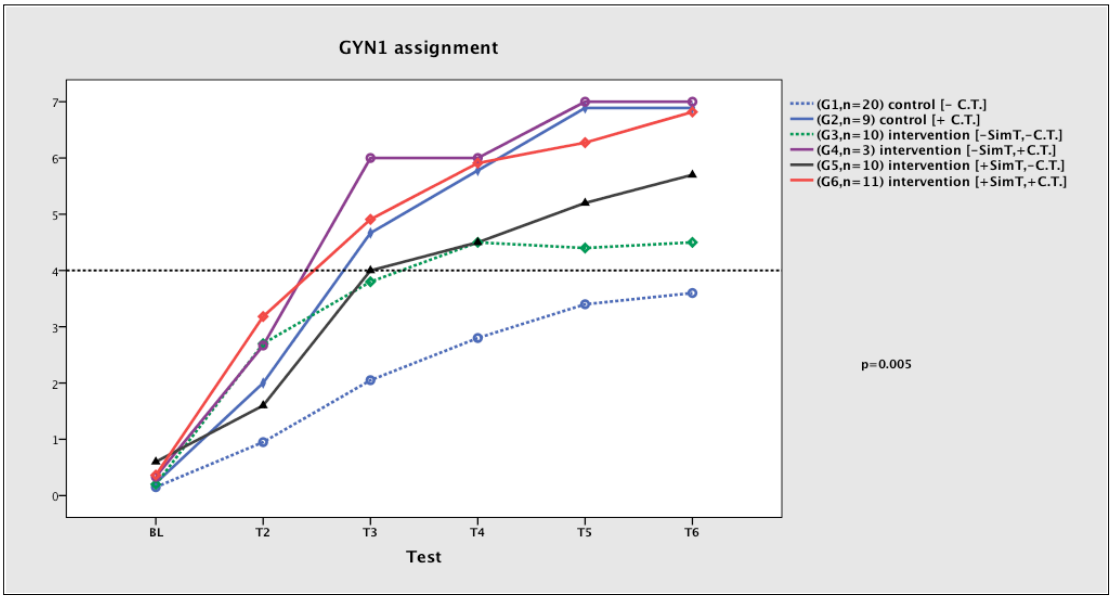


Table (7.2.C): Significance differences for the six subgroup was tested by repeated measures ANOVA

		within tests, p-value					
	Between six subgroup (overall)	G1 Con-CT	G2 Con+CT	G3 -S-CT	G4 -S+CT	G5 +S-CT	G6 +S+CT
Baseline	0.2						
BL-T2	0.8	0.01	0.01	0.002	0.1	0.1	0.003
T2- T3	0.3	0.02	0.01	0.1	0.1	0.01	0.03
T3- T4	0.8	0.06	0.08	0.2	0.1	0.052	0.03
T4- T5	0.6	0.1	0.04	0.8	0.4	0.1	0.3
T5- T6	0.5	0.1	0.2	0.6	0.1	0.1	0.1
Overall	0.005						

Non-parametric statistical significance difference between six subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	0.03	0.01	0.04	0.01	0.03
G1(control-CT) vs G6 (+Sim+CT)	X	X	0.008	0.01	0.04	0.02
G1(control-CT) vs G5 (+Sim-CT)	X	X	0.04	X	X	X
Kruskal-Wallis test (between six groups)	0.7	0.1	0.02	0.07	0.03	0.02

Chart (7.3): Plots of repeated measures analysis result for intervention (n=34) and control (n=29) represent result of GYN2 assignment.

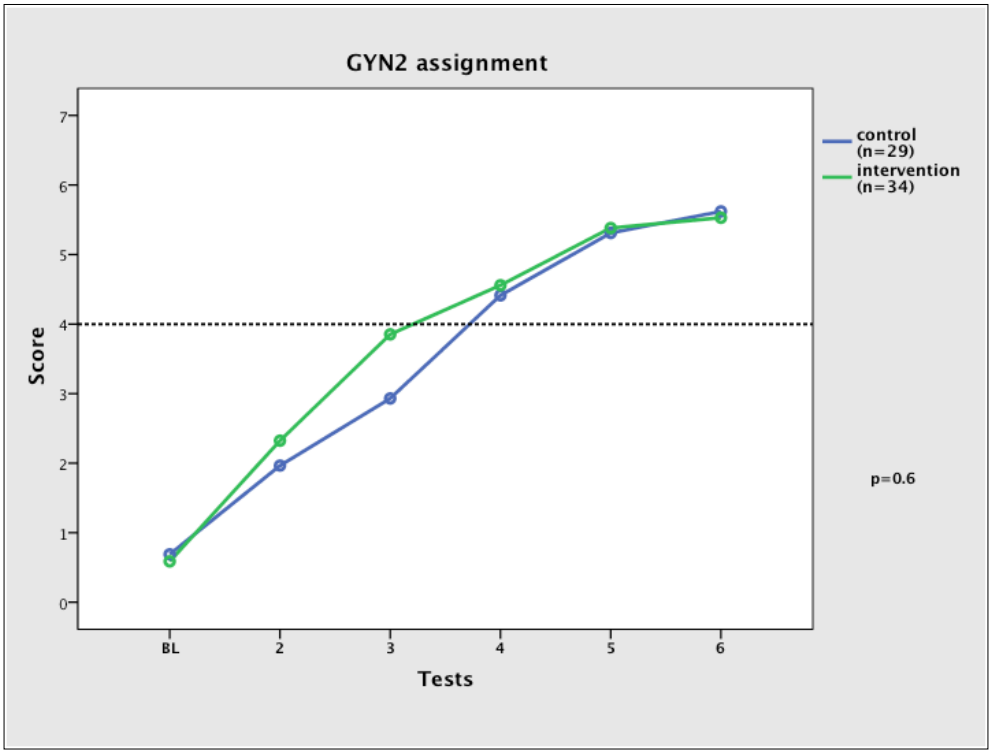


Table (7.3.A): Significance differences between intervention and control learning curves was tested by repeated measures ANOVA

		p-value	
	Within tests between control and intervention	within tests (control)	within tests (intervention)
Baseline	0.1		
Baseline - test 2	0.3	0.001	0.001
Test 2- test 3	0.3	0.02	0.001
Test 3- test 4	0.07	0.001	0.001
Test 4- test 5	0.9	0.04	0.09
Test 5- test 6	0.6	0.2	0.5
Overall	0.6		

Chart (7.3.1): Plots of repeated measures result presented GYN2 assignment for four subgroup: G1: control did not receive clinical training (n=20), G2: control received clinical training (n=9), G3: intervention received any type of training: simulation and/or clinical training (n=21) and G4: intervention did not receive any training (n=13).

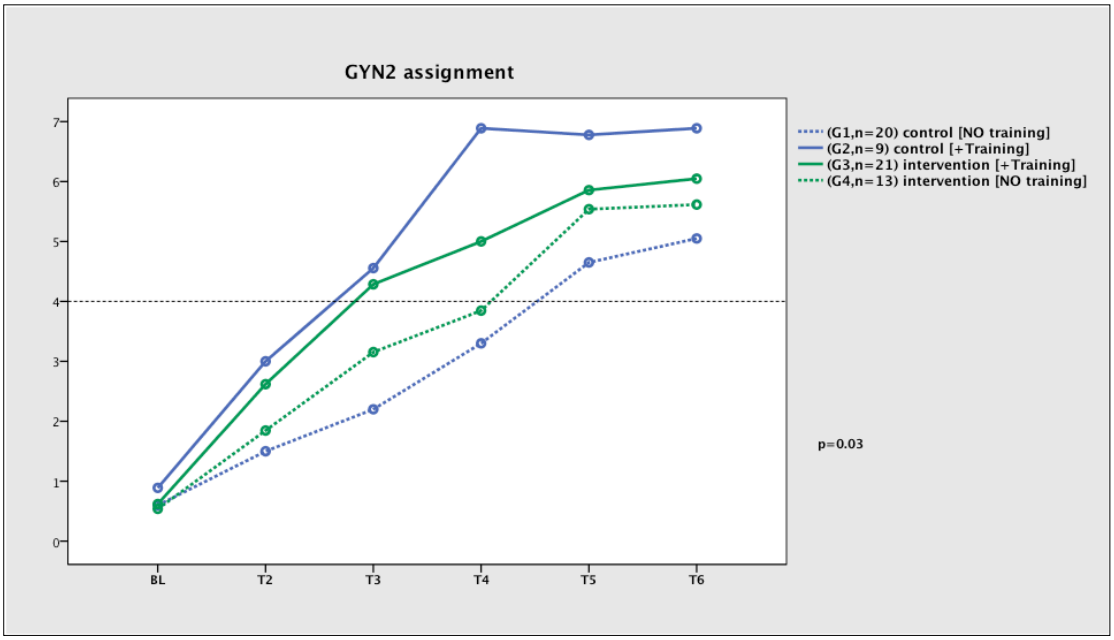


Table (7.3.B): Significance differences for the four subgroup was tested by repeated measures ANOVA

		within tests, p-value			
	Between four subgroup (overall)	G1 Control-CT	G2 Control+CT	G3 intervention +CT	G4 intervention - CT
Baseline	0.1				
Baseline - test 2	0.1	0.04	0.002	0.001	0.04
Test 2- test 3	0.5	0.1	0.054	0.002	0.1
Test 3- test 4	0.09	0.02	0.02	0.002	0.1
Test 4- test 5	0.3	0.03	0.6	0.1	0.04
Test 5- test 6	0.9	0.2	0.3	0.2	0.9
Overall	0.03				

Non-parametric statistical significance difference between four subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	X	0.04	0.009	X	X
G1(control-CT) vs G3 (intervention+T)	X	X	0.01	X	X	X
Kruskal-Wallis test (between four groups)	0.7	0.1	0.06	0.01	0.1	0.4

Chart (7.3.2): Plots of repeated measures result presented GYN1 assignment for six subgroup: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G3 (-Sim-CT): intervention did not receive any type of training: no simulation, no clinical training (n=10) and G4 (–Sim+CT): intervention received clinical training alone (n=3), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).

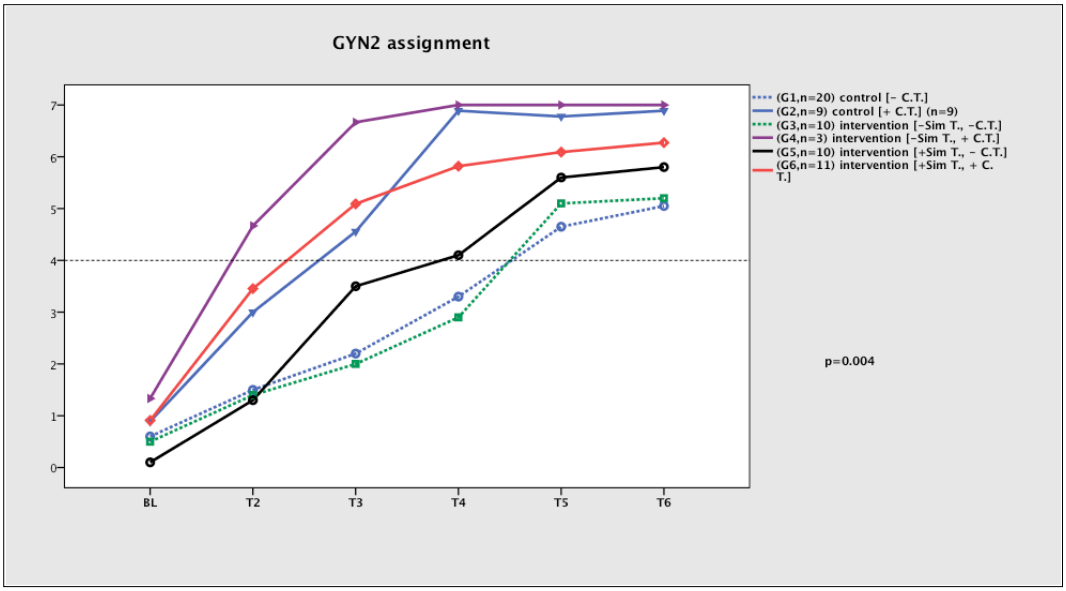


Table (7.3.C): Significance differences for the six groups was tested by repeated measures ANOVA

		within tests, p-value					
	Between six subgroup (overall)	G1 Con-CT	G2 Con+CT	G3 -S-CT	G4 -S+CT	G5 +S-CT	G6 +S+CT
Baseline	0.2						
BL-T2	0.03	0.04	0.002	0.1	0.1	0.01	0.001
T2- T3	0.4	0.1	0.054	0.4	0.4	0.02	0.02
T3- T4	0.2	0.02	0.02	0.1	0.4	0.051	0.03
T4- T5	0.2	0.03	0.6	0.04	0.4	0.7	0.6
T5- T6	0.9	0.2	0.3	0.9	0.4	0.5	0.3
Overall	0.004						

Non-parametric statistical significance difference between six subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	X	0.04	0.01	X	X
G1(control-CT) vs G6 (+Sim+CT)	X	0.01	0.01	0.04	X	X
Kruskal-Wallis test (between six groups)	0.7	0.1	0.02	0.07	0.03	0.02

Chart (7.4): Plots of repeated measures analysis result for intervention (n=34) and control (n=29) represent result of early pregnancy assignment.

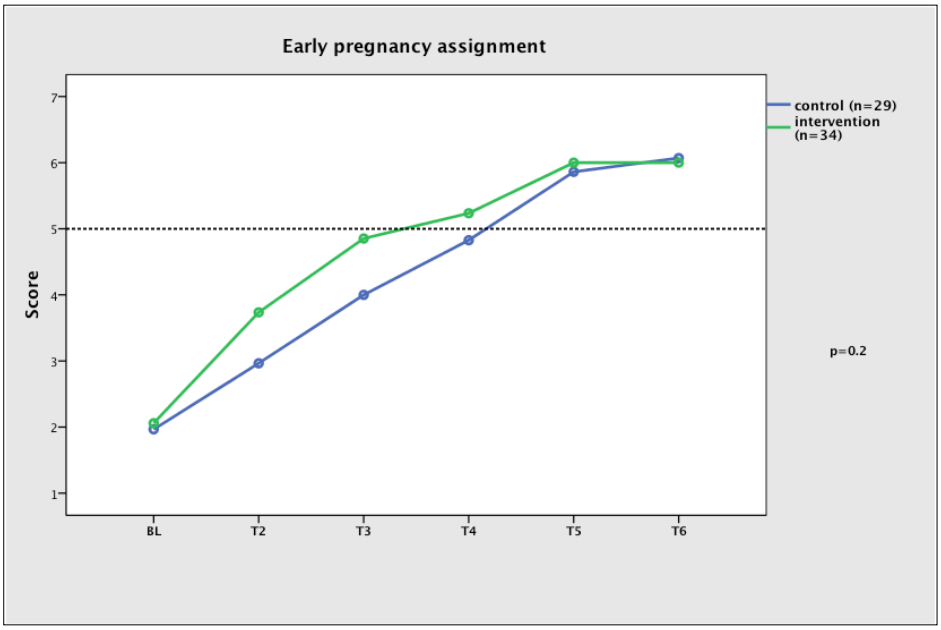


Table (7.4.A): Significance differences between intervention and control learning curves was tested by repeated measures ANOVA

		p-value	
	Within tests between control and intervention	within tests (control)	within tests (intervention)
Baseline	0.1		
Baseline - test 2	0.1	0.002	0.001
Test 2- test 3	0.8	0.006	0.001
Test 3- test 4	0.1	0.006	0.002
Test 4- test 5	0.5	0.01	0.001
Test 5- test 6	0.1	0.1	0.8
Overall	0.2		

Chart (7.4.1): Plots of repeated measures result presented early pregnancy assignment for four subgroup: G1: control did not receive clinical training (n=20), G2: control received clinical training (n=9), G3: intervention received any type of training: simulation and/or clinical training (n=21) and G4: intervention did not receive any training (n=13).

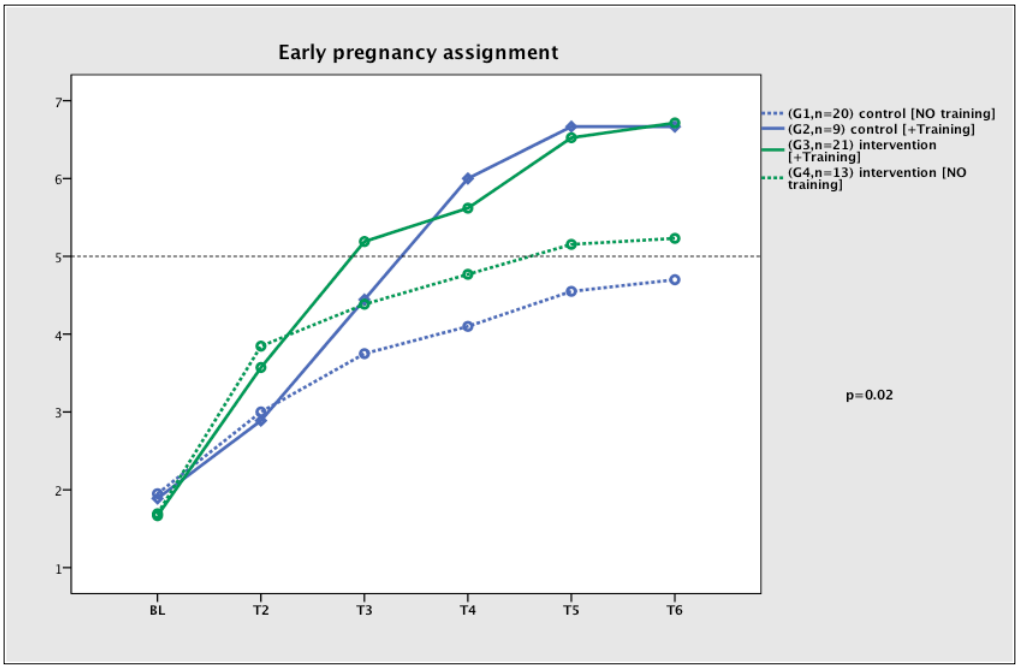


Table (7.4.B): Significance differences for the four subgroup was tested by repeated measures ANOVA

		within tests, p-value			
	Between four subgroup (overall)	G1 Control-CT	G2 Control+CT	G3 intervention +CT	G4 intervention - CT
Baseline	0.2				
Baseline - test 2	0.1	0.01	0.04	0.001	0.001
Test 2- test 3	0.1	0.08	0.07	0.001	0.2
Test 3- test 4	0.04	0.1	0.03	0.058	0.01
Test 4- test 5	0.6	0.2	0.1	0.002	0.04
Test 5- test 6	0.7	0.1	1.0	0.1	0.3
Overall	0.02				

Non-parametric statistical significance difference between four subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	X	X	0.02	0.01	0.03
G1(control-CT) vs G3 (intervention+CT)	X	X	0.01	0.02	0.001	0.001
Kruskal-Wallis test (between four groups)	0.6	0.3	0.1	0.06	0.005	0.003

Chart (7.4.2): Plots of repeated measures result presented early pregnancy assignment for six subgroup: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G3 (-Sim-CT): intervention did not receive any type of training: no simulation, no clinical training (n=10), G4 (–Sim+CT): intervention received clinical training alone (n=3), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).

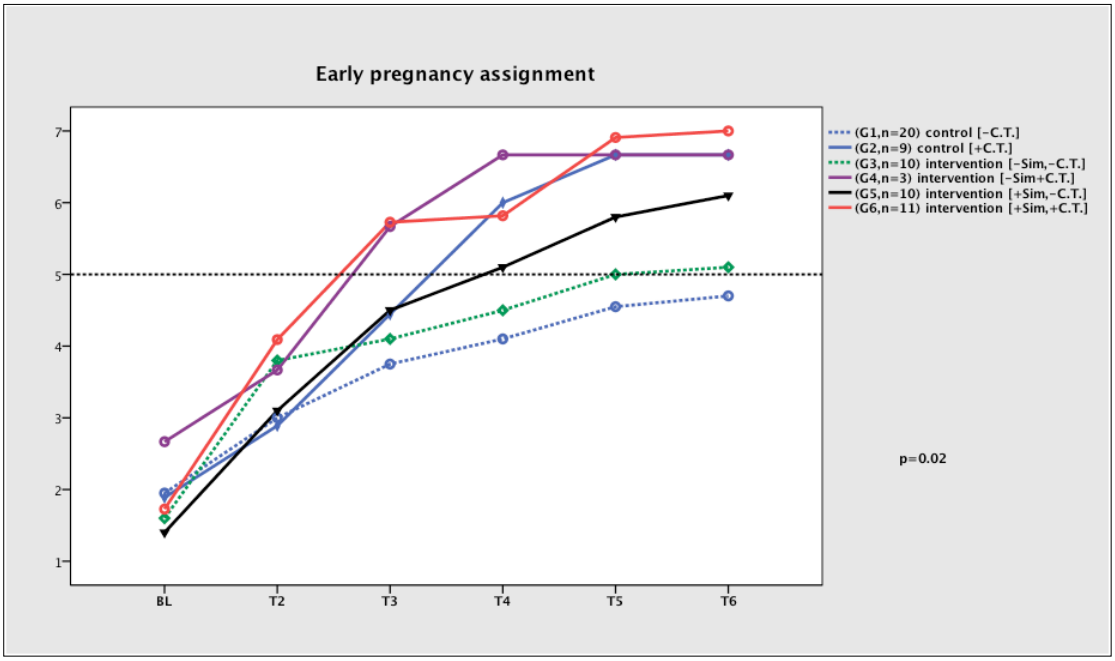


Table (7.4.C): Significance differences for the six groups was tested by repeated measures ANOVA

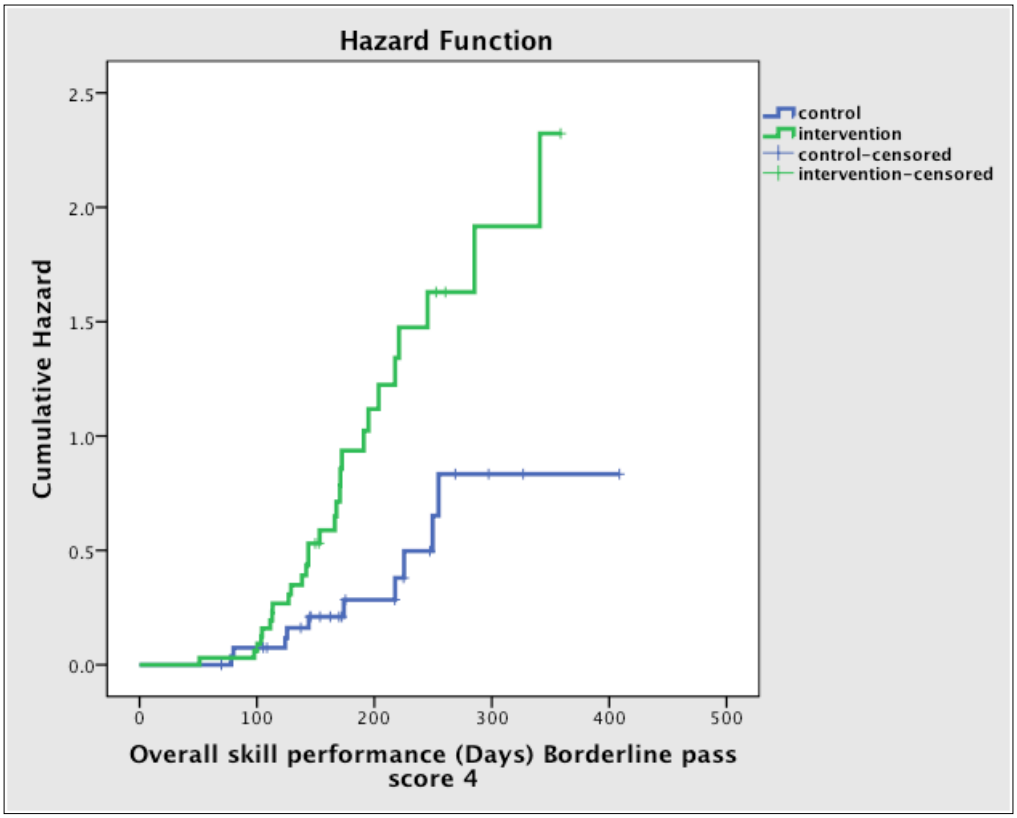
		within tests, p-value					
	Between six subgroup (overall)	G1 Con-CT	G2 Con+CT	G3 -S-CT	G4 -S+CT	G5 +S-CT	G6 +S+CT
Baseline	0.2						
BL-T2	0.1	0.01	0.04	0.002	0.2	0.003	0.002
T2- T3	0.3	0.08	0.07	0.5	0.07	0.02	0.002
T3- T4	0.08	0.1	0.03	0.03	0.2	0.1	0.6
T4- T5	0.7	0.2	0.1	0.09	0.3	0.6	0.01
T5- T6	0.7	0.1	1.0	0.3	0.3	0.1	0.3
Overall	0.02						

Non- parametric statistical significance difference between six subgroup

	Mann-Whitney U tests (between two groups)					
	BL	**T2**	**T3**	**T4**	**T5**	**T6**
G1(control-CT) vs G2 (control+CT)	X	X	X	0.03	0.01	0.04
G1(control-CT) vs G6 (+Sim+CT)	X	X	0.01	0.04	0.002	0.005
Kruskal-Wallis test (between six groups)	0.3	0.4	0.1	0.07	0.008	0.01

Chart (7.5): Plots of Kaplan Meier cumulative hazard illustrated the overall skill performance in attainment of borderline score of 4 (chart 7.5a) and a maximum score (7) (chart 7.5b) in (days) as scored by the control (n=29) and intervention (n=34) groups.

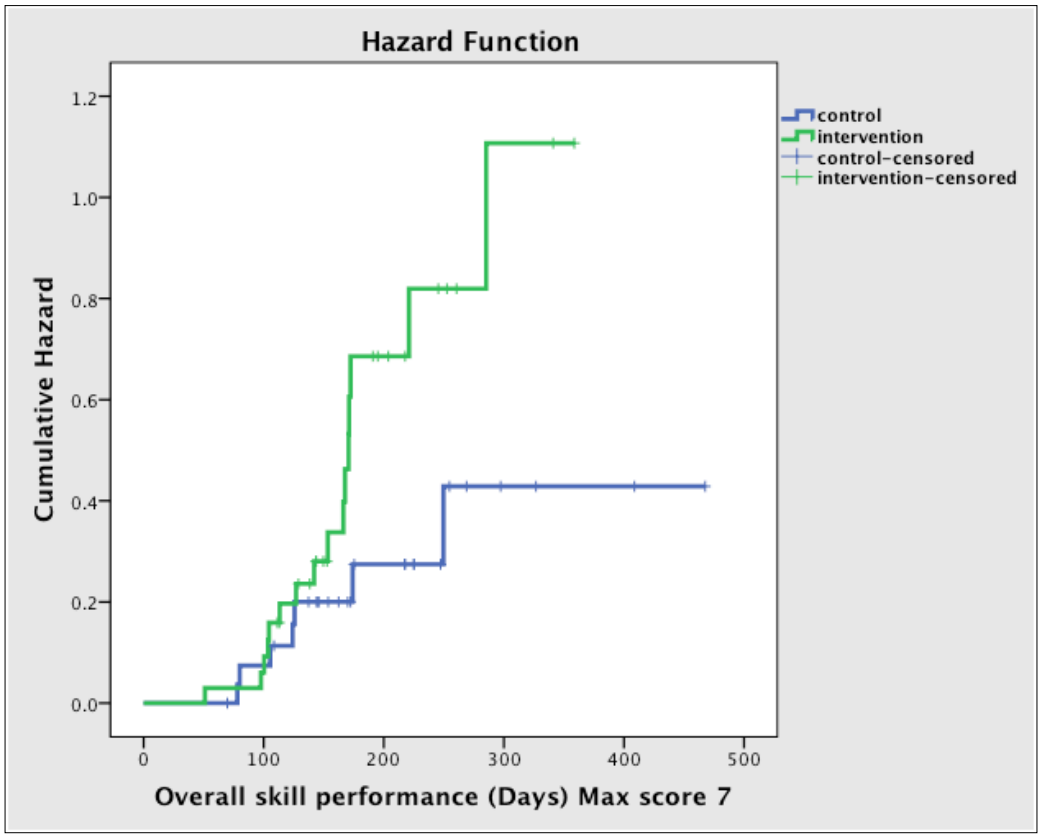
(Chart 7.5.a)



Overall skill performance (combined three modules)
Attainment of borderline score of 4 in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	254	7.052	1	0.008
Intervention (n=34)	167			

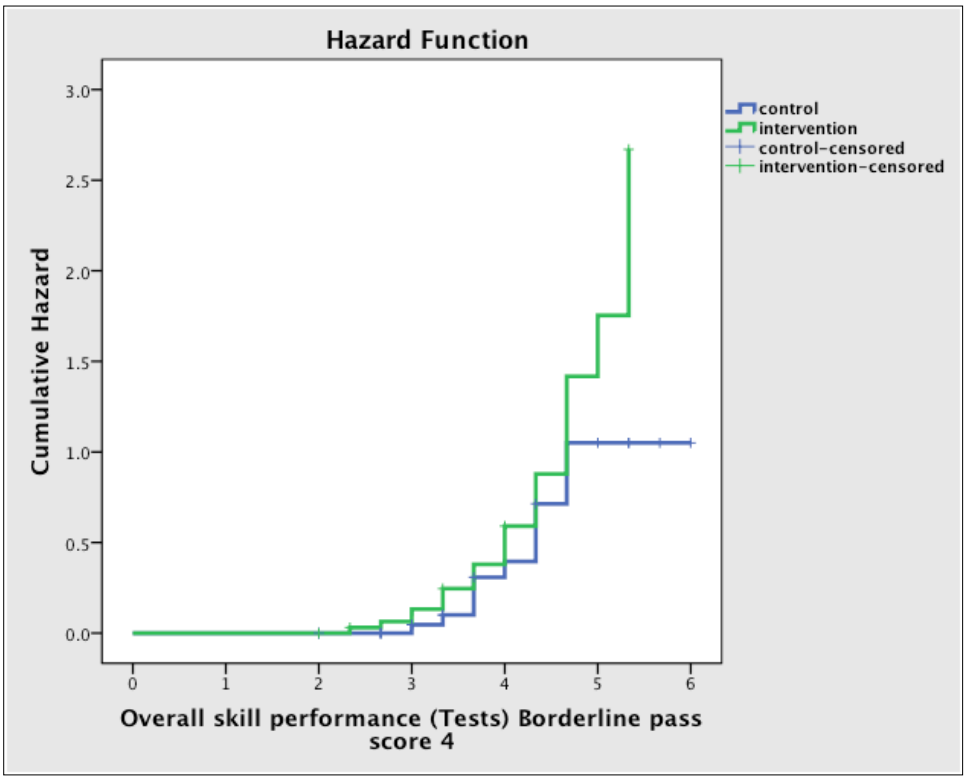
(**Chart 7.5.b)**



Overall skill performance (combined three modules)
Attainment of maximum score (7) in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	248	3.05	1	0.08
Intervention (n=34)	221			

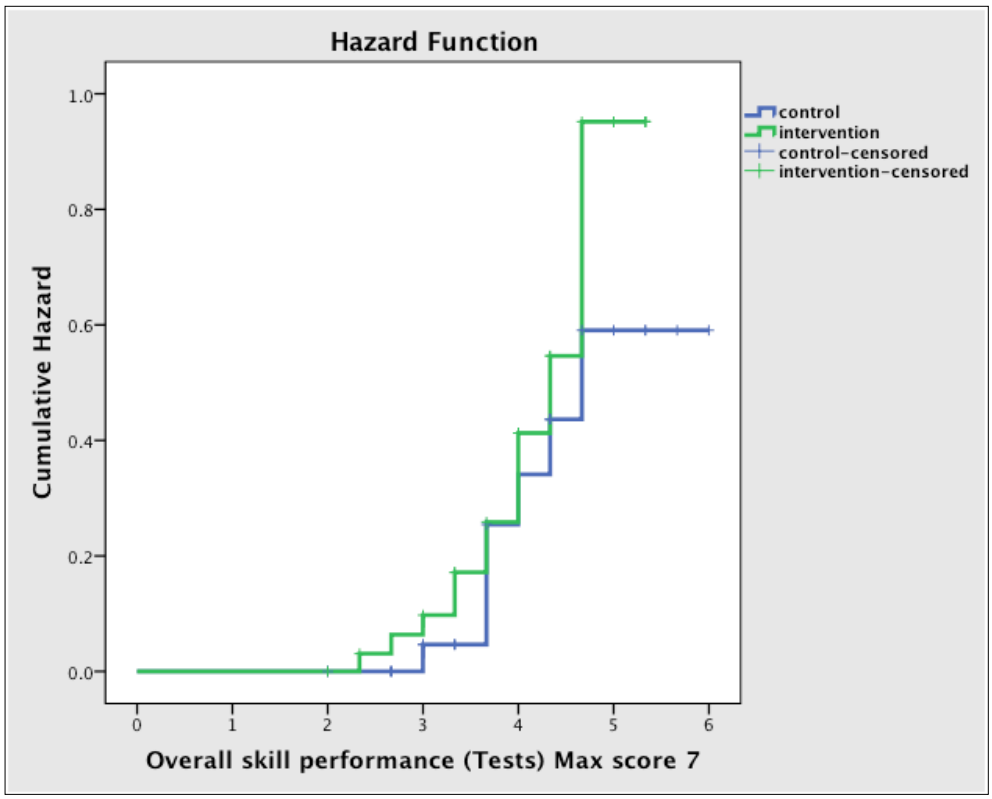
(**Chart 7.5.c)**



Overall skill performance (combined three modules)
Attainment of borderline score of 4 by Tests

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (Tests)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	4	3.478	1	0.06
Intervention (n=34)	4			

(**Chart 7.5.d)**

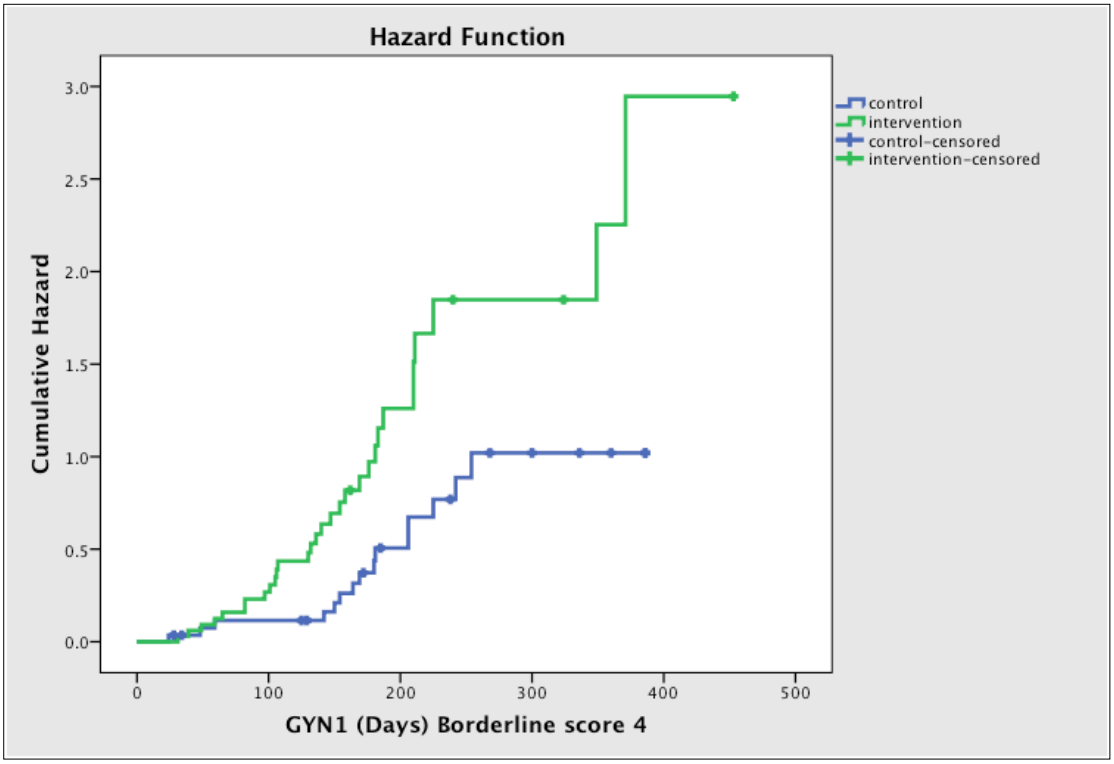


Overall skill performance (combined three modules)
Attainment of maximum score (7) by Tests

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (Tests)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	5	1.453	1	0.2
Intervention (n=34)	4			

Chart (7.6): Plots of Kaplan Meier cumulative hazard illustrated the performance in the attainment of borderline score of 4 (chart 7.6.a) and maximum score of (7) (chart 7.6.b) in GYN1 as scored by the control (n=29) and intervention (n=34).

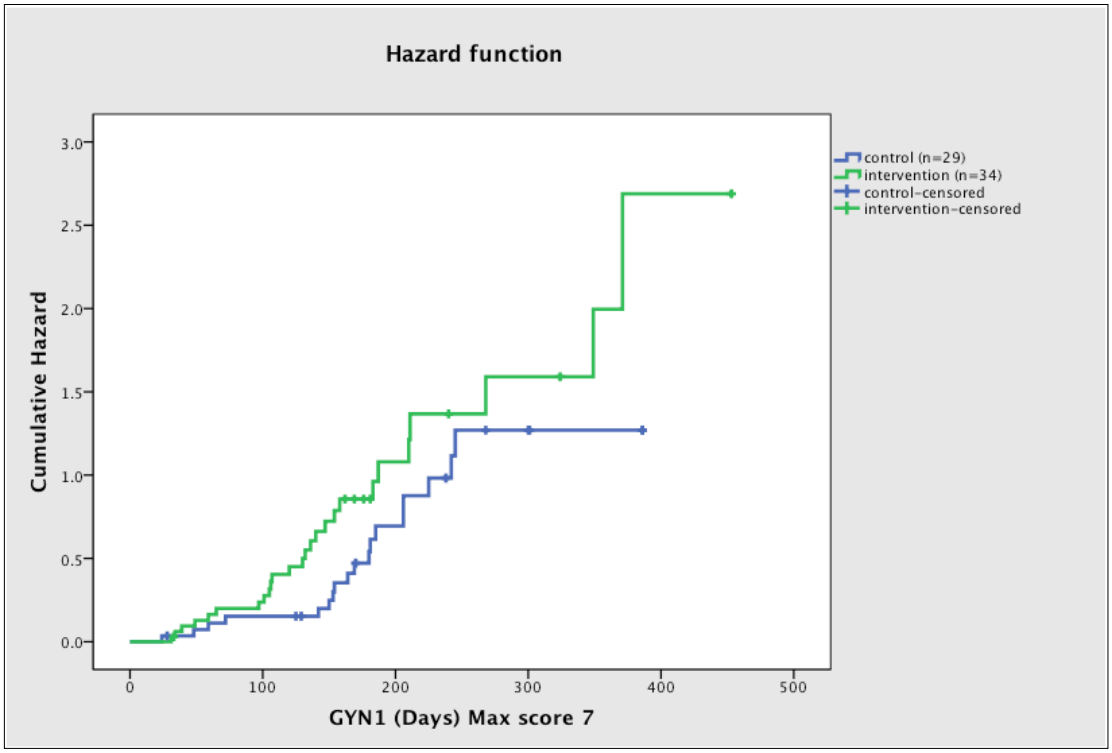
(Chart 7.6.a)



GYN1 assignment
Attainment of borderline score of 4 in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	225	7.517	1	0.006
Intervention (n=34)	147			

(**Chart 7.6.b)**

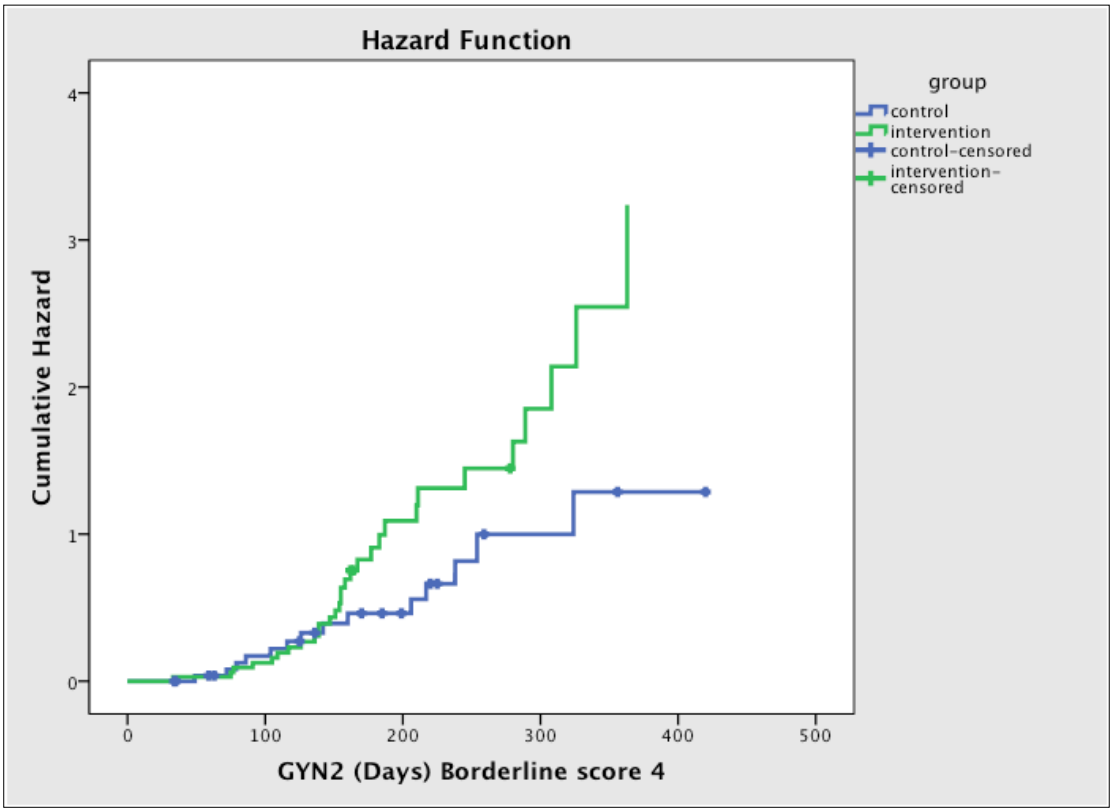


GYN1 assignment
Attainment of maximum score (7) in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	185	2.754	1	0.09
Intervention (n=34)	147			

Chart (7.7): Plots of Kaplan Meier cumulative hazard illustrated the performance in the attainment of borderline score of 4 (chart 7.7.a) and maximum score of (7) (chart 7.7.b) in GYN2 as scored by the control (n=29) and intervention (n=34).

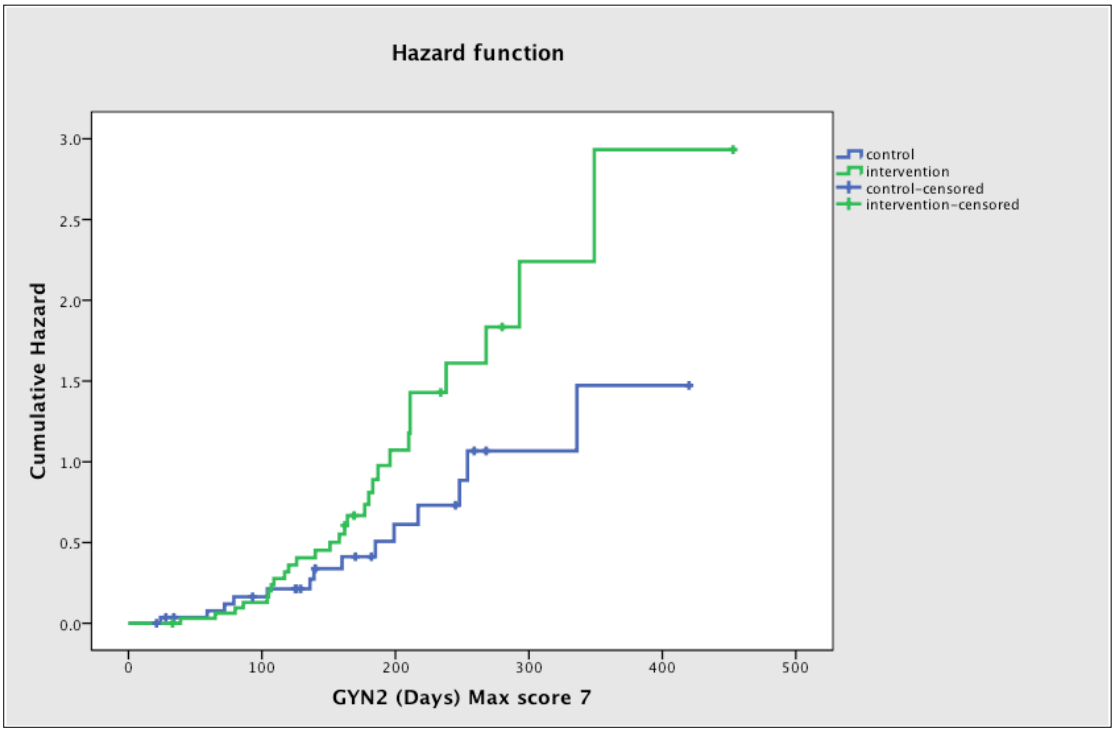
(Chart 7.7.a)



GYN2 assignment
Attainment of borderline score of 4 in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	238	3.925	1	0.04
Intervention (n=34)	158			

(Chart 7.7.b)

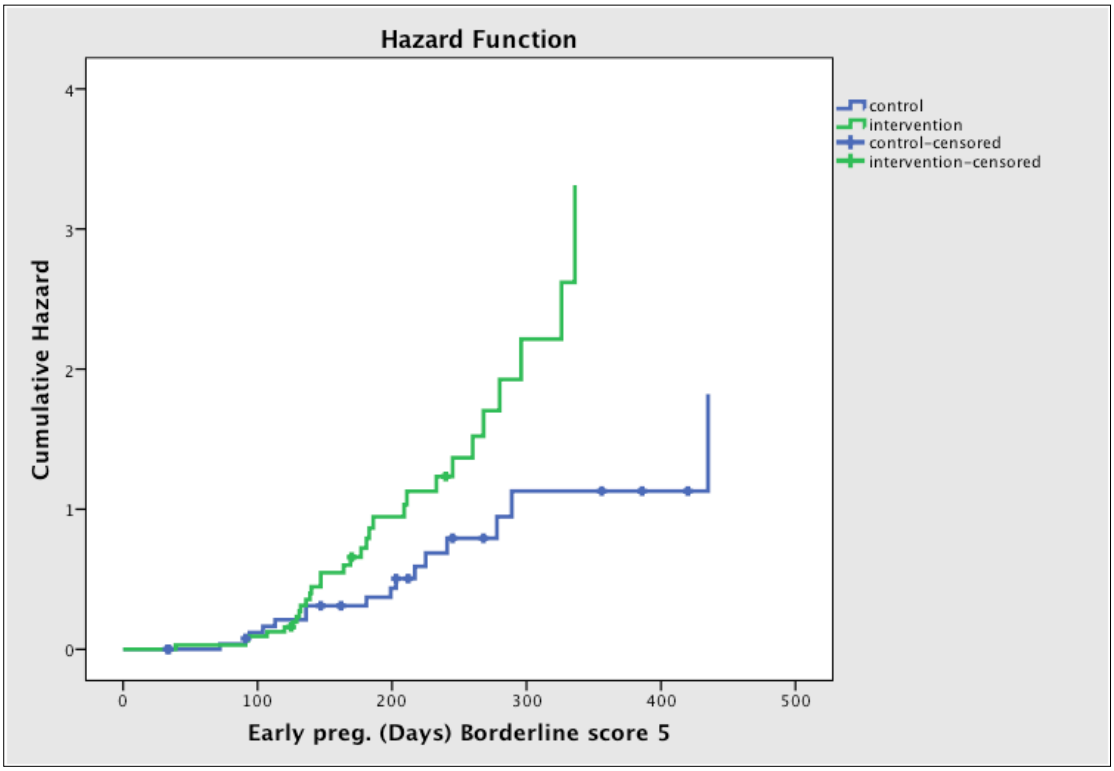


GYN2 assignment
Attainment of maximum score (7) in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	217	2.741	1	0.09
Intervention (n=34)	177			

Chart (7.8): Plots of Kaplan Meier cumulative hazard illustrated the performance in attainment of borderline score of 4 (chart 7.8.a) and maximum score of (7) (chart 7.8.b) in early pregnancy as scored by the control (n=29) and intervention (n=34).

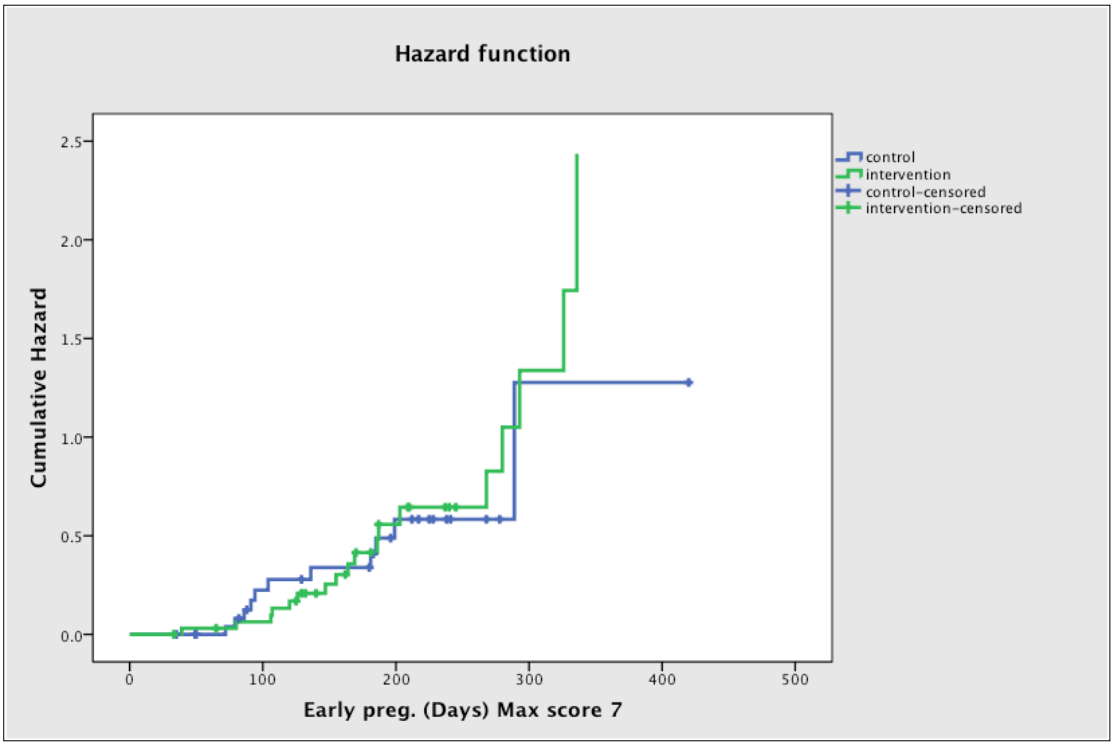
(Chart 7.8.a)



Early pregnancy assignment
Attainment of borderline score of 5 in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	241	3.925	1	0.04
Intervention (n=34)	177			

(Chart 7.8.b)

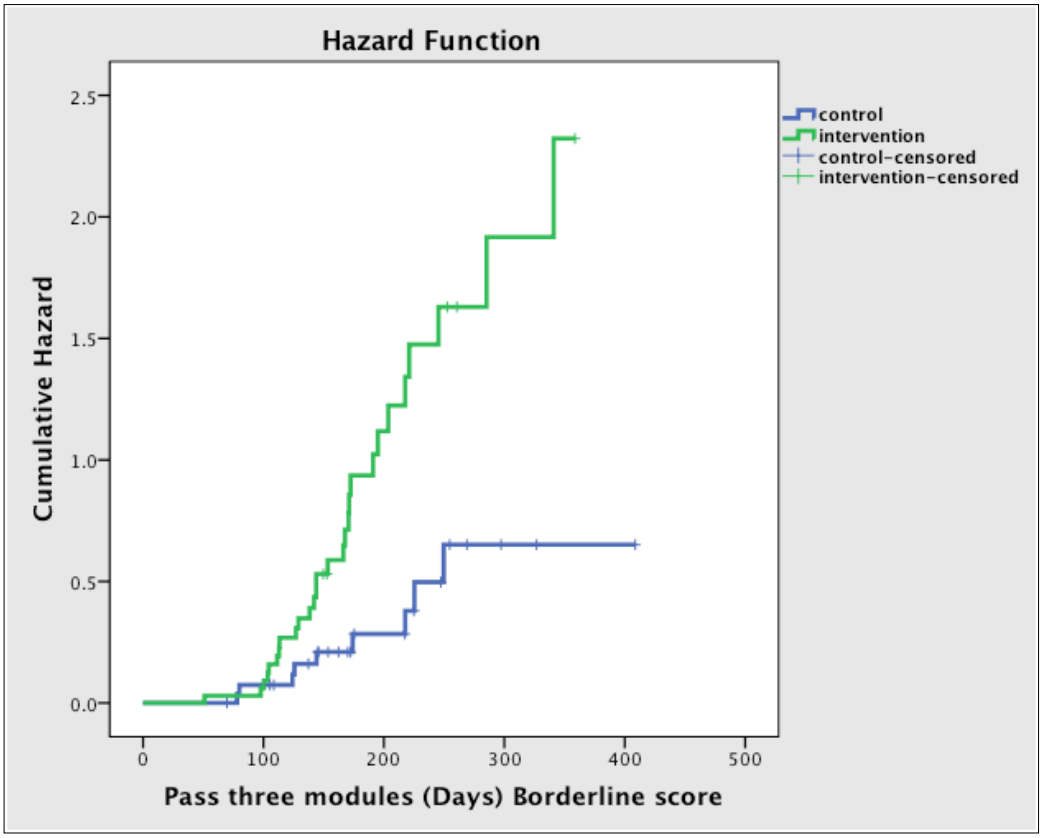


Early pregnancy assignment
Attainment of maximum score (7) in DAYS

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	289	0.90	1	0.7
Intervention (n=34)	268			

Chart (7.9): Plots of Kaplan Meier cumulative hazard illustrated the overall skill performance (pass three modules). The attainment in the borderline score of 4 (chart 7.9.a) and a maximum score (7) (chart 7.9.b) in (days) as achieved by the control (n=29) and intervention (n=34) groups.

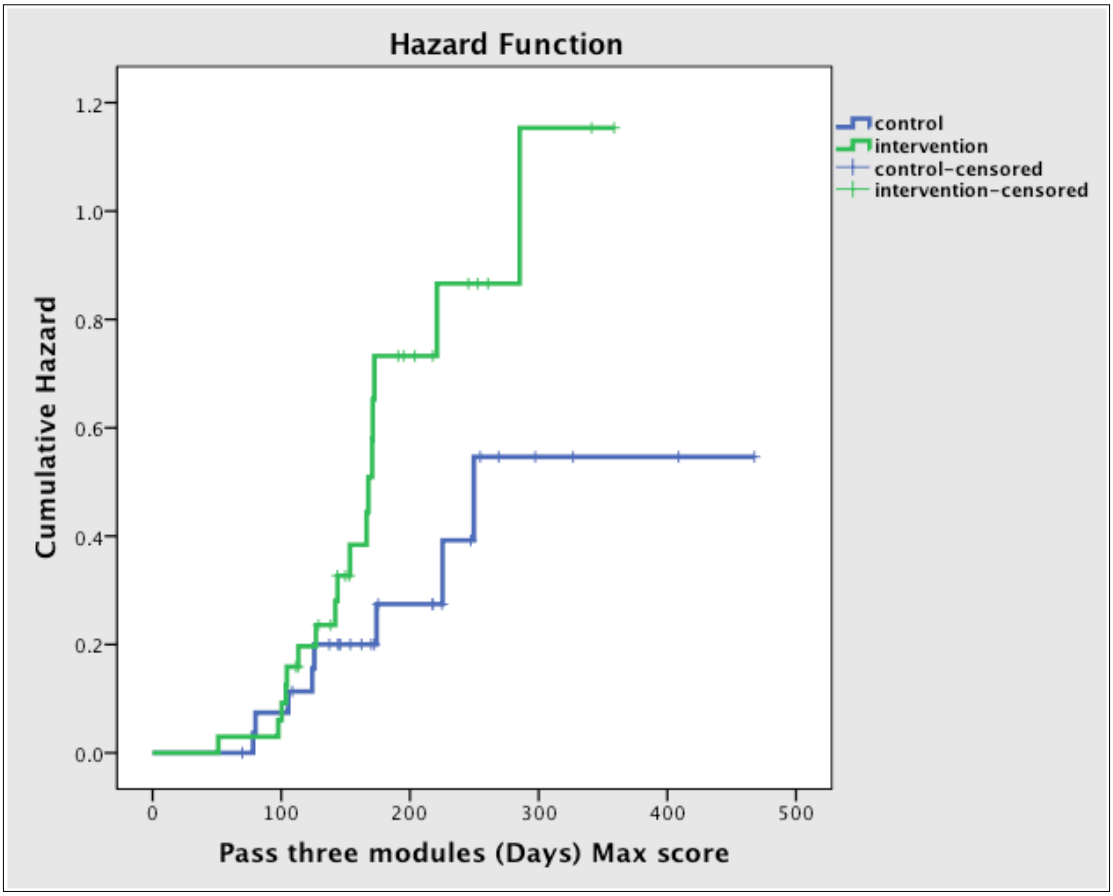
(Chart 7.9.a)



Overall skill performance
Attainment of borderline score of 4 in DAYS
Pass three modules

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	210	8.058	1	0.005
Intervention (n=34)	167			

(Chart 7.9.b)

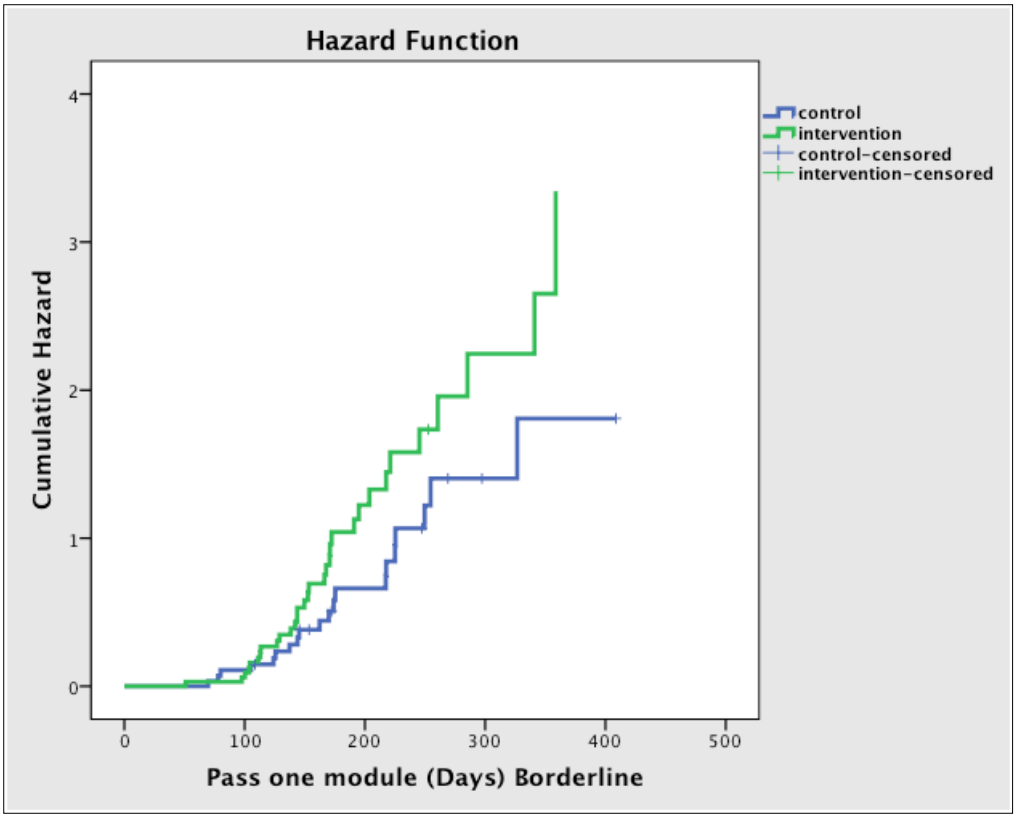


Overall skill performance
Attainment of maximum score (7) in DAYS
Pass three modules

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	210	2.875	1	0.09
Intervention (n=34)	172			

Chart (7.10): Plots of Kaplan Meier cumulative hazard illustrated the overall skill performance (pass one module). The attainment in the borderline score of 4 (chart 7.10.a) and a maximum score (7) (chart7.10.b) in (days) as achieved by the control (n=29) and intervention (n=34) groups.

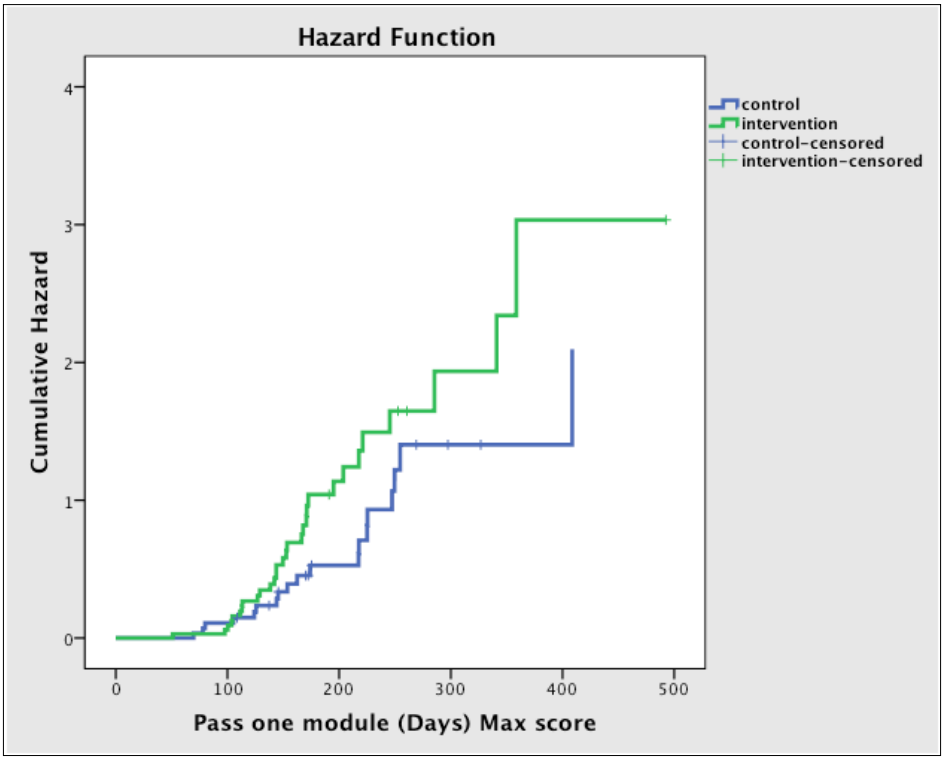
(Chart 7.10.a)



Overall skill performance
Attainment of borderline score of 4 in DAYS
Pass one module

Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	217	2.159	1	0.1
Intervention (n=34)	153			

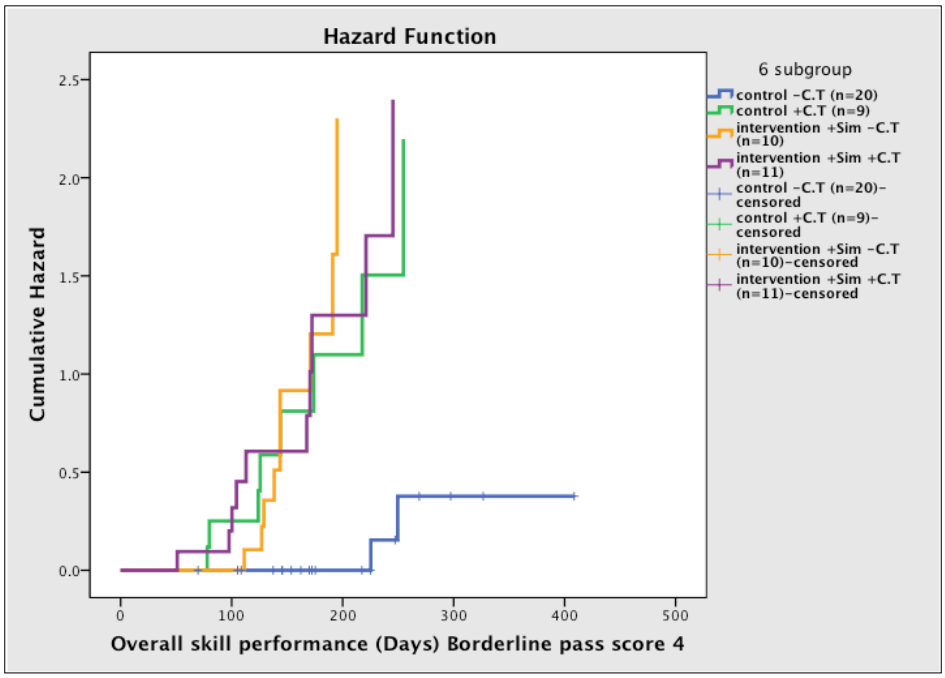
(Chart 7.10.b)



Overall skill performance
Attainment of maximum score (7) in DAYS
Pass one module

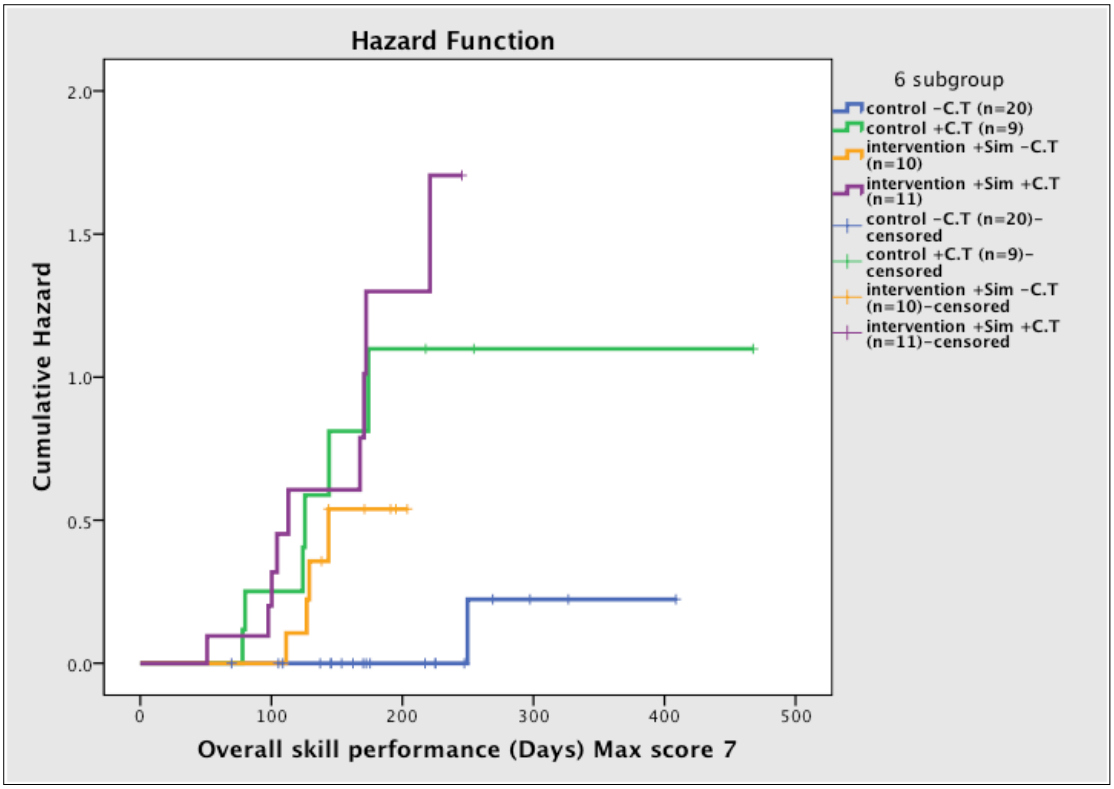
Control vs. Intervention		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
Control (n=29)	217	2.075	1	0.1
Intervention (n=34)	153			

Chart (7.11): Plots of Kaplan Meier cumulative hazard illustrated the performance in the attainment of borderline score of 4 for overall skill performance (combined three modules). The result is estimated in (days) for six subgroups: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).



Overall skill performance
Attainment of borderline score of 4 in DAYS

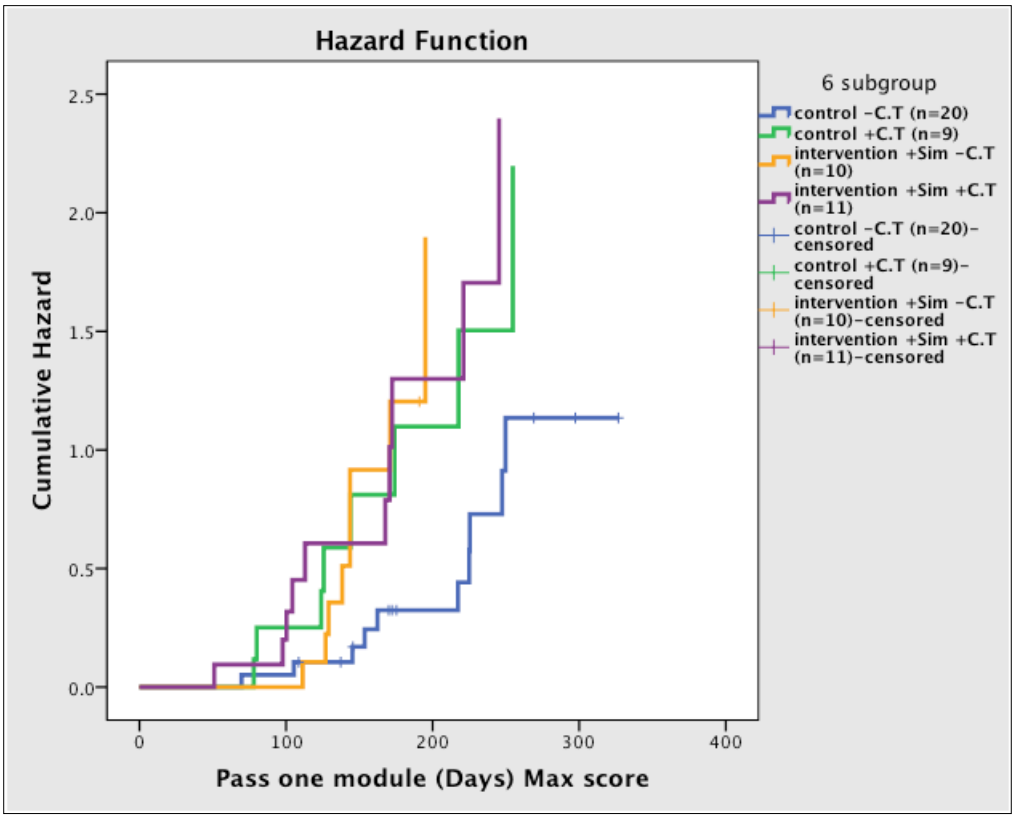
G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	225	25.129	3	0.0001
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	143			
G6:Intervention (+Sim+CT), (n=11)	167			



Overall skill performance
Attainment of maximum score (7) in DAYS

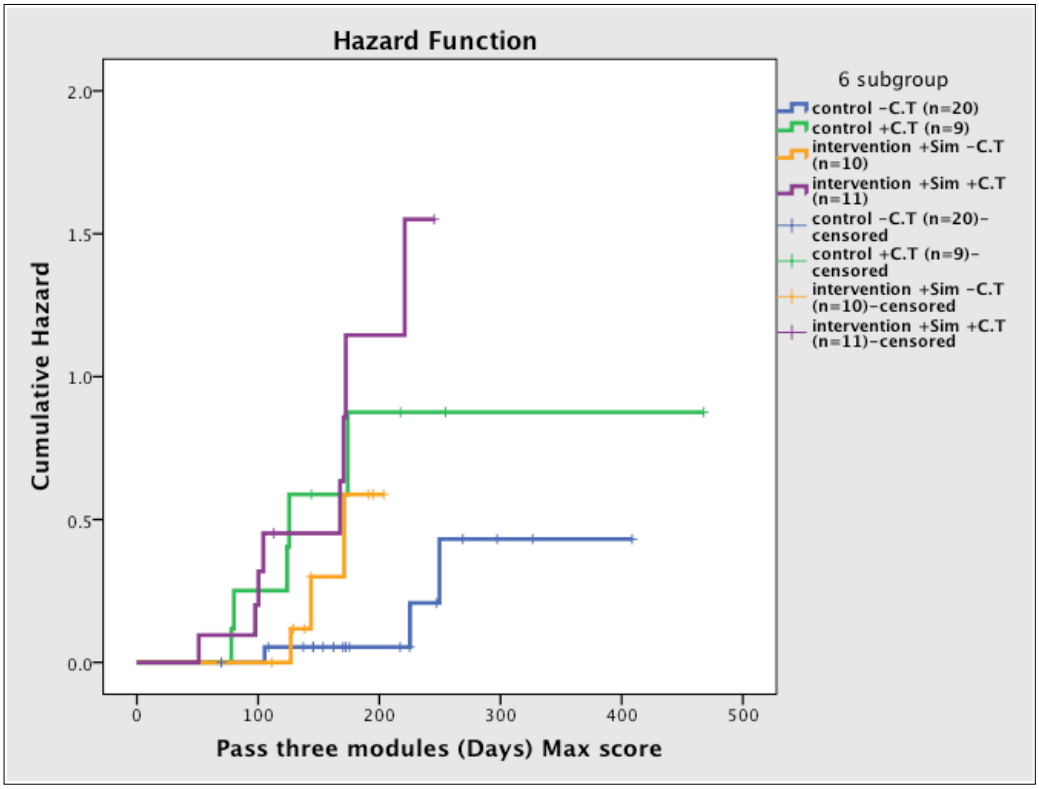
G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	278	19.178	3	0.0001
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	130			
G6:Intervention (+Sim+CT), (n=11)	167			

Chart (7.12): Plots of Kaplan Meier cumulative hazard ilistrated the performance in the attainment of maximum score (7) to pass one and three modules. The result is estimated in (days) for six subgroups: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11)



Overall skill performance
Attainment of maximum score (7) in DAYS (pass one modules)

G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	225	9.301	3	0.02
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	143			
G6:Intervention (+Sim+CT), (n=11)	167			

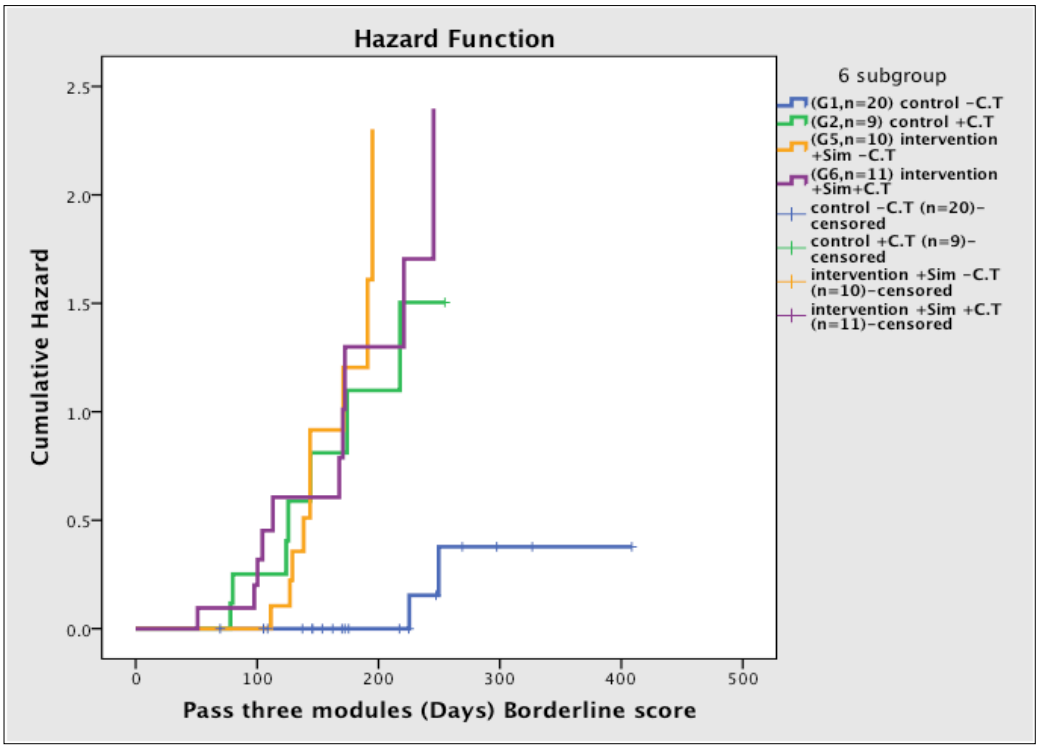


Overall skill performance
Attainment of maximum score (7) in DAYS (pass three modules)

G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	225	12.727	3	0.005
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	143			
G6:Intervention (+Sim+CT), (n=11)	167			

Chart (7.13): Plots of Kaplan Meier cumulative hazard illustrated the performance in the attainment borderline score of 4 for overall skill performance (combined three modules) to pass one and three modules. The result is estimated in (days) for six subgroups: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).

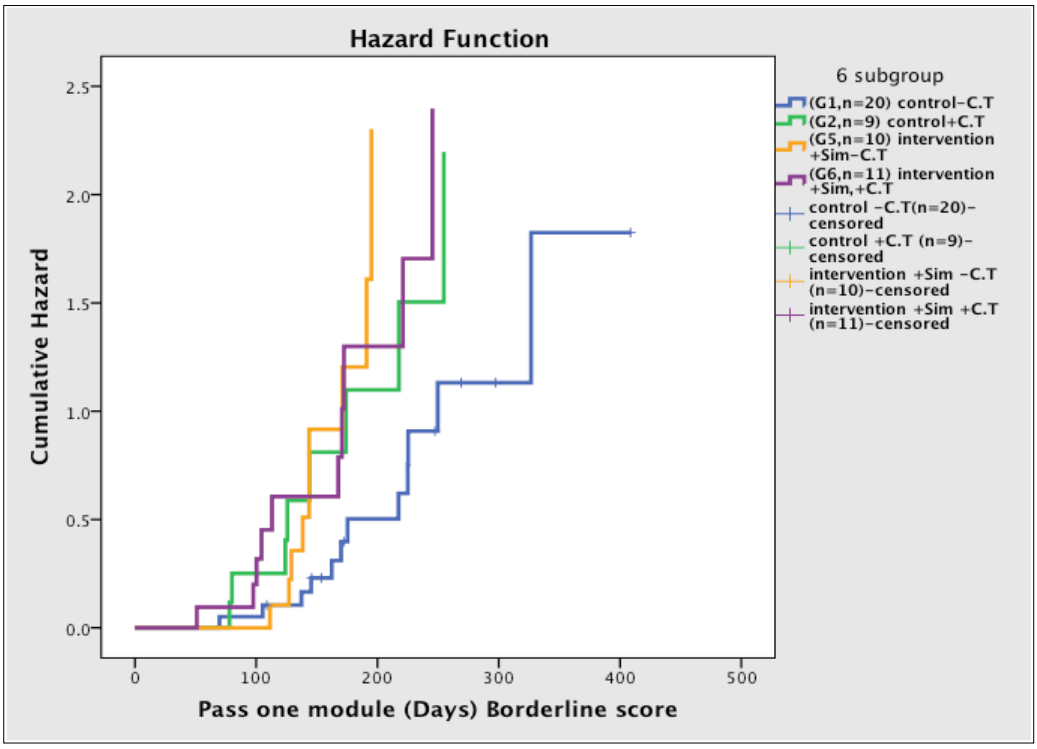
(Chart 7.13.a)



Overall skill performance
Attainment of borderline score of 4 in DAYS (pass three modules)

G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	225	24.3777	3	0.0001
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	143			
G6:Intervention (+Sim+CT), (n=11)	167			

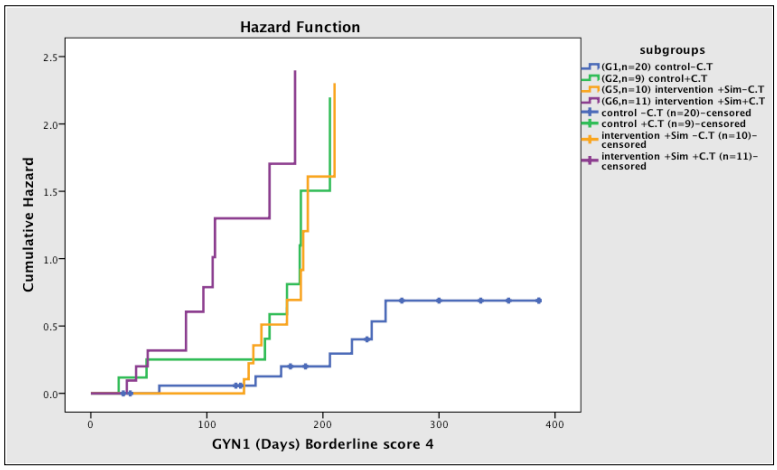
(Chart 7.13.b)



Overall skill performance
Attainment of borderline score of 4 in DAYS (pass one module)

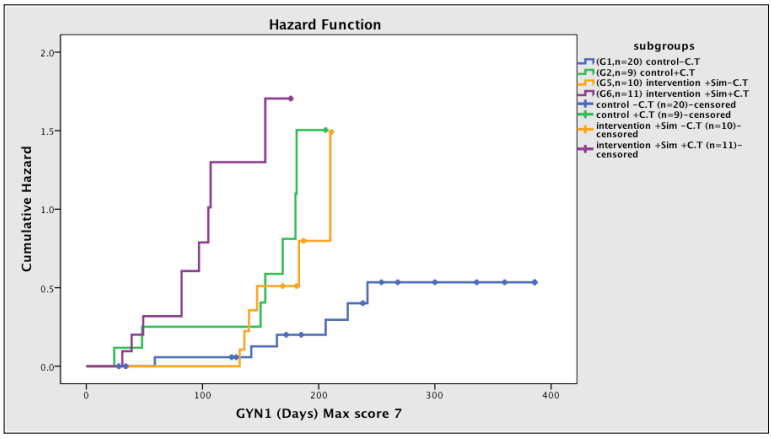
G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	225	8.964	3	0.03
G2:Control (+training), (n=9)	144			
G5:Intervention (+Sim-CT), (n=10)	143			
G6:Intervention (+Sim+CT), (n=11)	167			

Chart 7.14: Plots of Kaplan Meier cumulative hazard illustrated the performance in the attainment maximum score of 7 and borderline score of GYN1 assignment. The result is estimated in (days). The six subgroups: G1 (control-CT): control did not receive clinical training (n=20), G2: (control+CT): received clinical training (n=9), G5 (+Sim-CT): intervention received simulation training alone, G6 (+Sim+CT): intervention received simulation and clinical training (n=11).



GYN1 assignment
Attainment of borderline score of 4 in DAYS

G1, G2,G5,G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	267	26.484	3	0.0001
G2:Control (+training), (n=9)	169			
G5:Intervention (+Sim-CT), (n=10)	169			
G6:Intervention (+Sim+CT), (n=11)	97			



GYN1 assignment
Attainment of maximum score of 7 in DAYS

G1, G2, G5, G6		Log Rank (Mantel-Cox)		
Group	**Median (days)**	**Chi-Square**	**df**	**Significance**
G1:Control (no training), (n=20)	267	20.667	3	0.0001
G2:Control (+training), (n=9)	169			
G5:Intervention (+Sim-CT), (n=10)	183			
G6:Intervention (+Sim+CT), (n=11)	97			

Chapter eight

End of Trial survey

CHAPTER 8

Participants' perceptions of the effectiveness of simulation practice: End of Trial survey

---

Ultrasound is an integral part of obstetrics and gynaecology training. The Royal College of Obstetrics and Gynaecology's (RCOG) ultrasound training programme is competency-based and designed to ensure trainees develop the skills they need to use in clinical obstetrics and gynaecology ultrasound practice. The challenge of acquiring sufficient skills in a reduced training time in order to function safely at skilled practitioners' level is a problem not only confined to obstetrics and gynaecology but which also applies to all specialties where trainees need to acquire practical skills (Rosenblatt and Abrams, 2002; Ahmed et al, 2011; Moss et al., 2011, Madsen et al., 2014; Konge et al., 2015). In ultrasound training, trainees find 'experience in scanning' difficult to access and struggle to reach RCOG required competencies (Burden et al., 2011). The specific causes of limited training opportunities are presented as unrealistically long scanning lists; decreased doctors' hours and increased patient expectations. Therefore, ultrasound simulation training generates an additional opportunity to deliver flexibility and a systematic learning approach for trainees, away from engaging in busy clinics. One of the greatest benefits of the simulator is its ability to provide interactive learning opportunities with a large number of cases in a controlled and safe environment (Chung et al., 2011; Blum et al., 2013; Chalouhi et al., 2015a). As a final step in this randomised controlled trial, the participants were surveyed about their perceptions with simulation practice as supplemental to clinical training, and were evaluated to determine the benefits and limitations of training models.

8.1 Aim and objectives

The aim of this survey was to explore trainees' perceptions of simulation training as supplemental to their clinical training. The objectives of this end-of-trial survey were to investigate current ultrasound training delivered to obstetrics and gynaecology trainees in order to determine: (1) the benefits and limitations of ScanTrainer® ultrasound simulator compared to a physical model i.e. mannequin Blue Phantom™ (Figure 8.1); (2) the barriers and obstacles that have contributed to the gap in learning transvaginal ultrasound and (3) to clarify the potential solutions/suggestions that might help in enhancing current ultrasound training.

8.2 Subjects and method

The sample included subjects who were primarily obstetrics and gynaecology speciality trainees (ST) in the Welsh Deanery and were enrolled in this randomised controlled trial (RCT). These participants were mostly trainees at ST1–ST3 level, other NHS staff and students of the MSc programme at Cardiff University. The sample included subjects who completed the RCT as well as those who withdrew. The study had been reviewed and ethically approved by the South East Wales Research Ethics Committee SEWREC (NHS REC Reference 10/WSE02/75).

8.2.1 Internet-based questionnaire

The 'Google docs survey', an internet-based questionnaire tool was used here to investigate current obstetrics and gynaecology ultrasound training as well as trainees' self confidence in performing transvaginal ultrasound (TVUS) after simulation and clinical training. The cover letter was emailed along with the survey, which began with thanking participants and expressing appreciation of their enrolment. Also the overall objectives of the survey and length of time needed to complete it were stated. The survey's statements are helped depicted in exploring trainees' opinion about current barriers and potential ways of overcoming those barriers while learning TVUS in obstetrics and gynaecology (Appendix 8.1). The survey statements were previously piloted at an early stage of the trial as a part of student selected component (SSC) project undertaken by two undergraduate medical students (Langan et al., 2012; Mullins et al., 2012). The link to end of trial survey was accessible for a period of six weeks starting from the date on which the survey questionnaires were distributed, with three follow-up reminders sent as regular interval to subjects who had not

completed the survey. In order to collect multidimensional information regarding trainees' perceptions of simulation and clinical training, the survey was designed to include: (1) quantitative data which was represented in general statements, multiple choice options and a global rating scale, and (2) open comments field.

8.2.2. The pilot survey

The pilot survey was distributed in August 2012 to twenty-nine participants who were enrolled on this RCT at that time and included those who withdrew or were lost to follow-up but willing to participate in completing the survey. The survey was used to assess the benefits and limitations as well as barriers to and reasons for engagement with the ultrasound simulator for learning TVUS training. The survey was available in paper and electronic formats and by phone interviews to maximise convenience for the participants. If a paper or electronic copy was chosen, participants were contacted to ensure they had received the forms. The participants were asked to add any further comments not included in the survey that they thought were important to add and share with other participants. Hence, some of the questions in the pilot survey were modified and re-written to fulfil and cover further important aspects from the trainees' perspective (Appendices 8.2 and 8.3). The survey consisted of generic information e.g. where the trainee accessed the simulator, frequency and ease of access, and perceived benefits of and barriers to accessing the simulator. The questions in the survey were based on multiple choice, global rating scale and open question formats.

8.3 Outcome measures

The desired outcomes were to explore the potential factors that influenced the learning of TVUS skills. In addition, it was also hoped that the survey would address the obstacles, differences, benefits and limitations of two learning approaches: simulation and clinical training. Trainees were asked to complete six sections in the survey, which were (1) general information regarding respondent's name, hospital, age, gender and clinical position; (2) general statements of ultrasound training, as this section outlines reasons for engaging in ultrasound training and taking part; (3) benefits and limitations of ScanTrainer® ultrasound simulator; (4) accessibility, obstacles and length of time spent in practising with the simulator; (5) thirteen statements about simulation training in terms of assessing the quality of service given by the ultrasound simulator in delivering cognitive and practical knowledge of TVUS

in gynaecology and early pregnancy practice, and (6) participation status, whether the respondents completed the trial or not potential methods of enhancing TVUS training as well as whether trainees had future plans to carry on using the simulator after the study ended.

8.4 Statistical data analysis

IBM SPSS Statistics version 20.0 was used to collect data and to conduct analysis. The Shapiro-Wilk test revealed that the data distribution were not normal, $p<0.05$. The questions in the survey consist of categorical data that was analysed as descriptive data. Median scores were obtained by the respondents in rating thirteen statements related to quality service delivered by ScanTrainer®. The significant different between the control and intervention groups in rating these statements was tested by Mann-Whitney U, where the significant considers at $p<0.05$. Result of pilot survey is found in appendices 8.2 and 8.3.

8.5 Result

A total of seventy-seven invitations were sent but only forty-four responses were received giving a response rate of 57%. In total, 17 (39%) respondents were enrolled in the control group while 27 (61%) were in the intervention group (Table 8.1).

Trainees' prior ultrasound experience and reasons of engagement to randomised controlled trial are reported in Table 8.2. Participants' perception of accessing the simulator and the benefits of using simulation training in learning TVUS is found in Table 8.3. Frequency of access to the simulation training centres during the trial and the obstacles faced by trainees when practising with the simulator are found in Table 8.4. Regular assessment received by trainees during the trial and the accompanying feedback provided about the TVUS performance is described in Table 8.5. In terms of assessing the quality of service provided by the ScanTrainer® ultrasound simulator, thirteen statements about the simulator were assessed and median scores were given to each statement, as shown in Table 8.6. The differences between the simulator and other TVUS learning models is listed in Table 8.7 while as well as the factors that made the simulator good at teaching TVUS skills are detailed in Tables 8.8. The benefits and limitations in learning TVUS in gynaecology and early pregnancy are

outlined in Table 8.9. Trainees' perceptions on possible solutions/suggestions that would enhance their experience with simulation training are found in Table (8.10).

The trainees offered a number of free comments e.g.,

"No simulator can accurately recreate a patient experience, verbal, no verbal communications etc., but is an excellent tool to grasp basic skills technique and etiquette with transvaginal scanning"

"Safe environment, but I do feel it should be an adjunct to live training rather than a sole means of training; two years later I had no live experience"

"I was very lucky with its location, and proximity to work area"

"I moved from my previous hospital to another one as a part of my training. It was great to train on the simulator and I'd have continued if I had the possibility"

8.6 Discussion

The findings in this study draw attention to important issues related to current ultrasound training in obstetrics and gynaecology. The response rate was high and the majority of respondents gave positive feedback about transvaginal ultrasound (TVUS) simulation training. Eighty percent of responses highly agreed that the ultrasound simulator was a helpful tool in teaching and assessing basic TVUS skills. As discussed in recent studies, simulation training may not replace clinical training in learning core ultrasound skills, however it is considered to be a useful tool in preparing trainees prior to them entering clinical settings (Williams et al., 2013; Madsen et al., 2014). The importance of simulation prior to clinical training may also refer to the intimate nature of TVUS scans, which makes practice opportunities severely limited. Hence the simulator offers a wide platform for trainees to learn from mistakes with unlimited repetitions, unlike the opportunities for learning provided by examining real patients (Sidhu et al., 2012; Blum et al., 2013; Chen et al., 2015b). In contrast, transabdominal (TAS) scanning in obstetrics is easier than TVUS, because that type of examination is not embarrassing to patients, thus recruiting volunteers for training purposes is more likely to be achievable (Blum et al., 2013).

Furthermore, a high number of responses in this survey highlighted that a major benefit of the ultrasound simulator was the self-directed learning element tool and that it provided feedback for individual performances without the need for a physical instructor to be present. In addition, the regular feedback allows broader conclusions to be drawn on trainee's progress (Fletcher et al., 2003). The simulator's limitations, included, (1) simulation training was neither realistic enough, nor had similar experience as on patients, (2) the difficulty in completing a number of tutorials because of inadequate explanation, and (3) the lack of patient interaction and/or the simulator not being realistic enough to feel a frozen pelvis or tenderness.

Subjectively, trainees had high expectations that simulation would provide real scanning. However, the virtual reality environment is an alternative option that would assist in learning a systematic approach towards ultrasonography and offers a scenario that is very close to real ones (Blum et al., 2013; Chalouhi et al., 2014; Tolsgaard et al., 2015a). In order to overcome some of these limitations, the Ph.D researcher (A.A.) was available to assist, guide and assess trainees on a regular basis and to discuss obstacles they encountered during simulation practice. This enhanced their understanding of weaknesses in their performance. In addition, A.A. was keen to report any technical issues to manufacturer, which might have affected on its quality service. The respondents rated A.A. assistance with a median score of 8 out 10. This may lead to the conclusion that continuous monitoring of trainees' performance on a regular basis, either through monitoring of progress on the simulator or through frequent formative assessment of trainees' skills, would enhance individual performance and acquisition of skills (Tolsgaard et al., 2015a).

One of the obstacles highlighted was absence of network/connection between simulators, which meant that the trainees had to use the same simulator to ensure that all data was saved on a single system. Indeed, this is an important learning point where a simulator on one site should link with others via 'icloud' to enable trainees as well as trainers to assess and monitor progress wherever they are. Moreover, motivating trainers and supervisors to support trainees and review their feedback is equally important as it encourages trainees to use simulation for learning.

There have been substantial arguments about what constitutes a suitable and proper location for the simulator in order to make it accessible for trainees and staff (Issenberg et al., 2005; Awtrey et al., 2010; Lateef F, 2010). Some respondents agreed that the simulator should preferably be at simulation training centres. However, others claimed that clinical and work locations are more suitable for simulation practice. The highest score was given to simulation training centres, as this location would facilitate easy access as well as enabling trainees to seek and ask for technical assistance if needed. In addition, training in simulation centres provides un-interrupted practice, unlike hospital clinic or ward-based simulators. One of the obstacles mentioned in the study was the difficulty trainees had in securing protected training time to practise on the simulator and thus simulation training centres were better suited locations for training. Some have argued that the best place for the simulator is anywhere that is surrounded by convenient equipment, space and learning materials which correspond to the nature of the training service offered (Huang H, 2002; Kaufman D, 2003; Issenberg et al., 2005; Awtrey et al., 2010; Lateef F, 2010; Walker et al., 2013; Burckett-St Laurent et al., 2016).

According to report published by the Department of Health "A Framework for Technology Enhanced Learning", 'technology has an important role to play in the continuum of managed learning processes' (Department of Health, 2011). The approach of combining a variety of different educational methods is commonly referred to as "blended learning" (Seymour et al., 2002; Pereira et al., 2007), and includes simulation, face-to-face teaching and e-learning in addition to traditional training. Previous studies pointed to the effectiveness of blended ultrasound training in preparing trainees to perform transvaginal (TVUS) and transabdominal (TAS) examinations by practising with mannequins and on e-learning modules (Alsalamah et al., 2009; Murugan et al., 2009). These studies also suggested that Blue Phantom™ mannequin-training model had a positive effect on improving basic ultrasound skills in obstetrics and gynaecology also with peer training approach. Consequently, as outlined in the chapter seven, six months was suggested by experts to be an adequate period for trainees to become competent. However, the majority of respondents stated that six months was too short for beginners to reach competency in performing basic TVUS when continuous and uninterrupted practice was not available. Thirty-two percent of respondents believed that engaging in simulation training was a training

requirement of the RCOG, with the aim of improving healthcare in women. However, NHS workload and the current shortage of sonographers weren't taken into account as the training programme is implemented at regional and local levels (RCOG, 2014). As trainees were expected to be independent and competent in performing basic obstetrics and gynaecology scanning by year four of speciality training, this would indicate that curriculum should be restructured to include blended learning material in order to fill the gap which exists in conventional training. Similar to the findings of Kodz (2003), trainees' views reflected the lack of training and suggested the need for a flexible, organised, well-supervised and supportive educational environment to help them gain core skills.

To summarise, simulation training supplemental to clinical training had positive effects on trainees' learning experience, performance and skills. However, the limitations of the simulator as a training tool, should not be ignored and the developers should work towards further enhancement to achieve better learning outcomes.

Chapter 8: Tables and Figures

Figure (8.1): Blue Phantom™ transvaginal mannequin

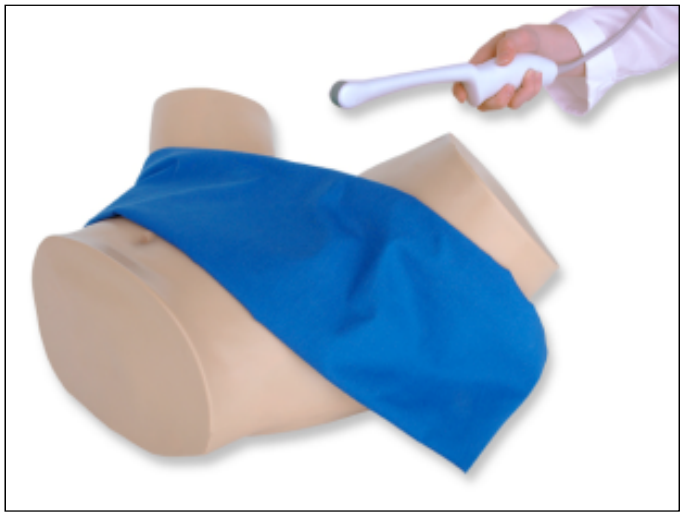


Table (8.1): Demographics of respondents (n=44)

	No. of respondents (%)	
Groups		
Control	17	(39%)
Intervention	27	(61%)
Total	**44**	
Location		
South East Wales (Cardiff)	33	(75%)
South Wales (Swansea)	4	(9%)
North Wales (Wrexham)	7	(16%)
Total	**44**	
Age		
24-30	13	(30%)
31-40	20	(45%)
41-50	8	(18%)
+50	3	(20%)
Gender		
Female	37	(87%)
Male	7	(16%)
Position		
Senior clinician	1	(2%)
Specialist trainee (ST)	32	(72%)
Postgraduate student	6	(14%)
Midwife	1	(2%)
Nurse	1	(2%)
Academic	1	(2%)
Clinical fellow	1	(2%)
Associate specialist	1	(2%)
Distribution of speciality trainees per years of training		
ST1	6	(14%)
ST2	8	(18%)
ST3	6	(14%)
ST4	6	(14%)
ST5	3	(7%)
ST6	2	(5%)
ST7	3	(7%)
Participation status		
Withdrawn	2	(5%)
Completed	42	(95%)

Table (8.2): Trainees' previous ultrasound experience before participation

	No. of respondents (%)	
Type of ultrasound training received in last two years		
Transvaginal	4	(9%)
Trans abdominal	10	(23%)
Both	26	(59%)
No training received	4	(9%)
Location of ultrasound training received		
Gynaecology clinics	3	(7%)
Antenatal clinics	8	(18%)
Early pregnancy assessment unit	2	(5%)
All of the above	22	(50%)
Radiology	3	(7%)
Fetal medicine training sessions	2	(5%)
Previous ultrasound training with mannequin		
Yes	15	(34%)
No	29	(66%)
List of reasons influenced the participant engagement to ultrasound training		
RCOG modules requirement	14	(32%)
Tutor/trainer's request	9	(20%)
Attending ultrasound workshop	7	(16%)
Personal interest	5	(11%)
Clinical practice requirement	7	(16%)
All of the above	2	(5%)

Table (8.3): Accessibility, benefits and limitations of simulation training

	No. of respondents (%)	
Location of accessing the simulator		
Training room in labour ward, UHW, Cardiff	11	(25%)
Cardiff Medicentre, Cardiff	12	(27%)
Singleton Hospital, Swansea	4	(9%)
Wrexham Maelor Hospital, Wrexham	6	(14%)
Total	**33**	
Simulator compared to other training services in its effectiveness to familiarize trainees with core ultrasound skills		
Better	20	(45%)
Equivalent	9	(20%)
Can't judge	15	(34%)
Benefits of simulator in addition to other training method		
Yes	30	(68%)
No	1	(2%)
Don't know	13	(30%)
Benefits of simulator in enhancing trainees competence in undertaking TVUS		
Yes	34	(77%)
No	2	(5%)
Don't know	8	(18%)
Simulator has limitations		
Yes	21	(48%)
No	3	(7%)
Don't know	20	(45%)
Simulator is good at teaching and assessment		
Yes	33	(75%)
No	5	(11%)
Don't know	6	(14%)
Simulator is useful and helpful tool		
Yes	35	(80%)
No	3	(7%)
Don't know	6	(14%)

Table (8.4): Access frequency to the simulator and potential obstacles

	No. of respondents (%)	
Access frequency to the simulator for training		
1-2 times per week	1	(2%)
1-2 times per a month	8	(18%)
Once / 3 months	5	(11%)
Accessed only once	8	(18%)
Never had accessed the simulator at all	22	(50%)
Obstacles in accessing the simulator		
Being in control group	6	(14%)
Work/duties and other commitments	12	(27%)
Wasn't given protected training time	11	(25%)
Kept failing tasks and lost interest	4	(9%)
Location of simulator not suitable	4	(9%)
Travel distance	4	(9%)
Haptic overheating makes a short sessions	3	(7%)
All of the above	2	(5%)
Overall 'un-interrupted practice' in one simulation training session		
Less than half an hour (<30 minutes)	8	(18%)
Less than an hour (31-60 minutes)	6	(14%)
More than an hour (>60 minutes)	5	(11%)
Didn't have 'un-interrupted' session	8	(18%)

Table (8.5): Assessments and feedback received during the trial

	No. of respondents (%)	
Frequency of assessment by the PhD researcher during trial		
Every 4 weeks	16	(36%)
Every 4-6 weeks	16	(36%)
Every 2-3 months	5	(11%)
Every >3 months	7	(16%)
Usefulness of feedback provided by the PhD researcher after each assessment session *(scale of 10-point- very useful)*	Median	
0----------10	8.0	
Duration of six month to gain competence in core skills is		
Too long	5	(11%)
Too short	21	(48%)
Adequate	18	(41%)
Presence of the instructor during the simulation session is		
Very helpful	31	(70%)
Somewhat helpful	13	(30%)
Need to approach trainer/instructor to help during session		
Needed help every session	10	(23%)
Needed only once	10	(23%)
Didn't need help at all	4	(9%)

Table (8.6): Median score of thirteen statements rated by the respondents (n=44) related to quality of service obtained by the ScanTrainer® ultrasound simulator. The significant difference between two groups tested by Mann-Whitney U.

Statement	Median			Mann-Whitney U test
	Control	Intervention	Overall	p-value
Statement 1: The simulator provides easy access to practice endovaginal ultrasound scanning **0……10**	8	7	7.5	0.8
Statement 2: The simulator is excellent for training beginners in ultrasound scanning skills **0……10**	8	8	8	0.5
Statement 3: It is a good process for teaching and learning a systematic approach in scanning **0……10**	8	8	8	0.7
Statement 4: Training with the simulator helps in familiarising the trainee with core ultrasound skills (endovaginal scanning skills) **0……10**	8	8	8	0.8
Statement 5: The ultrasound image appeared realistic in the simulator **0……10**	7	7	7	0.6
Statement 6: The instructions and the 3D depiction of anatomy have a very useful role in understanding the ultrasound image and orientation **0……10**	8	8	8	0.8
Statement 7: All buttons on the simulator were handy and well explained **0……10**	7	7	7	0.4
Statement 8: Providing force feedback on the operator's hand helps in simulating real scanning **0……10**	7	6	7	0.3
Statement 9: Training with the simulator reflects a similar or very close experience to a patient/volunteer scan **0……10**	6	5	6	0.2
Statement 10: My confidence in scanning is enhanced after training with the simulator **0……10**	7	7	7	0.4
Statement 11: Training is better at simulation training centres **0……10**	7	6	6	0.2
Statement 12: Training is better at clinical stations e.g. ANC, labour ward **0……10**	8	8	8	0.3
Statement 13: The simulator helps overcome the current shortage of learning capacity in hospitals and training centres **0……10**	8	7	7.5	0.9

Table (8.7): Trainees' quotes of potential differences between the ScanTrainer® ultrasound simulator and the mannequin: Blue Phantom™ in terms of learning core transvaginal ultrasound in gynaecology and early pregnancy.

Differences
"Overriding advantages is accessibility of the simulator"
"Familiar with basic and learning technique"
"Presenting cases of real patients"
"Driving machine"
"Picture much clearer and adjustable compared to pictures presented with mannequin"
"Anatomical diagram next to the scan picture assists in early skill development"
"Mannequin images are poor with lots of artifacts"

Table (8.8): Trainees' quotes of potential factors that make the simulator good at teaching TVUS skills

Factors
"Can be done anywhere at own time"
"Difficult practice the exact measurement of some findings in real life in clinic because of the limitation in time"
"Structured approach to examination and learn basic gynaecology anatomy and pathologies"
"Good at understanding of transvaginal scanning technique"
"Standardise was to gain image required"
"Learning further features such as labelling, placing callipers, dual screen etc."
"Very helpful with orientation"
"Haptic feedback and pressure gauge to suggest appropriate force to apply"
"Good overall"

Table (8.9): Trainees' quotes of benefits and limitations of the simulator.

Benefits
"Learning platform for beginners"
"Accelerate both skills and confidence build-up for beginners"
"All time practice outside clinical environment"
"Portability allows flexibility and repeatability of training"
"Doesn't require patient for trainee to master basic ultrasound anatomy and pathologies"
"Structured methods, provided better orientation especially for beginner who start training from the core modules"
"Provide clear instructions about examination and described anatomy in details"
"Variety of assignments provided in the ScanTrainer are beneficial for trainee to be familiar with different cases"
"Number of simulated cases much higher than in live patients"
"Consistent supervision and structured method"
"Should mandatory prior to patient transvaginal scan"
"Learn basic skills and technique before contact patient"
"Helped to locate structures to look for when "let loose" on patient"
"Gives extra familiarity with holding probe and use features to gain the images wanted"
Limitations
"Lack of recorded voice which explains anatomy and orientation before starting modules"
"Additive at present in view of the fact that scan training is limited otherwise absent"
"With some practices, item wouldn't pass even if the performance seem to be ok and that tolerances are fine clinically"
"Not the same as scanning patient, especially postmenopausal patients when ovaries can be quite difficult to locate"
"Anatomy isn't as obvious in reality, patient has high BMI's etc"
"Lack of patient interaction"
"Patients have varying anatomy that require special scanning technique due to achieve good image, i.e. using left hand to press on pelvic to bring uterus to front, which not feasible with the simulator"
"I did find sometimes I was working visually to what I knew the machine wanted"
"Can't get real feel of frozen pelvis or tenderness in assessing"

Table (8.10): Trainees' quotes about possible solutions that would enhance their experience with the ScanTrainer® ultrasound simulator

Solutions/suggestions
"Accessibility at all times"
"Good instruction by team"
"Good instruction by the simulator"
"Meeting Amal (PhD researcher) on regular basis for the assessment"
"Given protected time for training"
"Easily accessible in the department"
"Make all simulators connected where enable to log in and find your results on any simulator used"
"If I had been in clinical practice at the time"
"If it situated on or around labour ward or on-call room"
"Feedback on progress"
"More cases in the assignment are useful"
"Not realistic enough and the software is limited and rigid"
"It can't recognised the coronal plane so I kept failing the tasks"
"Having allotted time within rota to use it, followed by session in gynaecology clinics and EPAU to consolidate learning"
"Relocating it to antenatal clinics as that were most of the scan is done"
"Availability in every hospital"
"New training approach rather than the conventional one"
"More supervised time while using the simulator"
"Simulator has some exercise that are near to impossible to pass, which is very frustrating"
"Frequent practice will keep the skills updated"
"Setting feedback from other participants"

Chapter nine

General Discussion

CHAPTER 9

General discussion

This chapter is a general discussion of the overall findings in this thesis in comparison with the available evidence in the literature. To the best of the researcher's knowledge, the projects conducted in this thesis, including what is undoubtedly the largest ultrasound simulation educationally-driven randomised controlled trial to date. Therefore, some of these projects, such as face and content validity, intra- and inter-observer reliability of scoring systems developed for the assessment of obstetrics and gynaecology ultrasound skills, were conducted recently and no comparable studies using virtual reality simulators for TVUS have yet been published at the time of submission. The current findings were compared to available relevant findings in the literature including published studies in obstetrics and gynaecology ultrasound and other different medical specialities. As a result, review of the literature revealed a strong relationship between simulation training, enhanced performance and the acquisition of skills. When comparing findings in this thesis with other relevant work on ultrasound simulator ScanTrainer®, the results confirmed the validity and reliability of the simulation in offering a reliable practice and systematic learning approach in TVUS. Although some of our findings were consistent with other studies, the RCT was different in that it was conducted in an uncontrolled learning environment.

In reviewing the literature, no data was found on the assessment of time required for novice trainees to gain basic TVUS skills in a non-controlled conventional ultrasound learning environment. Very little was found in the literature on time, in hours, needed for the acquisition of skills by practising on the ultrasound simulator ScanTrainer® within a controlled learning environment. This study set out with the aim of assessing the time required for trainees to receive the simulation supplemental to clinical training in order to gain basic TVUS skills, with those who received clinical training

alone. Although the results of this study showed no statistically significant difference between the simulation and control groups, simulation training proved to be beneficial in shortening the time of learning TVUS. This result may be explained by the fact of the heterogeneity of subjects in the simulation and control groups. When subgroup analysis was considered, significant results were demonstrated. Subjects who received simulation training alone were able to reach the primary endpoint and pass all simulation modules in a shorter time compared to other groups. The findings of this study are consistent with those of Morgan et al. (2010) and Williams et al. (2013) who found no significant difference between simulator and control groups in post-test (after 10 hours of TVUS training), however the simulator group scored higher than the control at final plateau. That similarity leads to a conclusion that simulation training would add beneficial effects in terms of speeding up skills' acquisition and shortening the required learning time compared to conventional training.

From a medical education perspective, most challenges with simulation-supplemented skills training were fundamentally in determining the level of competency for trainees to gain basic skills (Tetzlaff J, 2007; Scalese et al., 2008; Ward et al., 2014). In the assessment of ultrasound skills, a number of published studies demonstrated that the change in trainees' performance and learning curves depended on attaining experts level (Madsen et al., 2014), acquisition of skills (Tolsgaard et al., 2014a) or use of specific parameters for assessing final skill level by repetition or with a comparison to other methods of training (Morgan et al., 2010; Williams et al. 2013, Madsen et al. 2014). The learning curve endpoint of trainee performance was clearly defined in our study by attaining the maximum score following successfully completing specific modules in the simulator. Similar to Madsen et al (2014) findings, a contrasting-group method was used to determine the borderline pass score for trainees to minimal competence in performing TVUS. Notwithstanding the scores of borderline (pass/fail level) in our study were different due to different in measurement parameters and simulation modules selection. In accordance with Dreyfus and Dreyfus (1980), the model of skills acquisition to master skills, novices required free training to gain basic skills and the threshold for practice as being minimally competent, may potentially be the borderline pass score. Our findings also demonstrated that ScanTrainer® was able to distinguish between subjects with different levels of TVUS experience indicating its construct validity. Tolsgaard et al.'s (2015a) construct validity study in the use of

ultrasound simulators addressed recent calls for valid and reliable assessment instruments to ensure high quality of scans provided by trainees in obstetrics and gynaecology. The authors also demonstrated that their Objective Structured Assessment of Ultrasound Skills (OSAUS) scale discriminated between trainees with different levels of competence and established credible pass/fail scores for two types of ultrasound examination. Agreeably, ultrasound simulation training has positive impacts on trainees' knowledge and performance, another subsequent study by Tolsgaard et al. (2015a) had shown that simulation-based ultrasound training leads to substantial improvement in clinical performance and that is sustained after 2 months of clinical training.

Our face and content validity study also demonstrated that the ScanTrainer® has a feel and look (face validity) that are realistic and tutorial structure (content validity) that is relevant for actual TVUS scanning. No comparable face and content validity studies addressing virtual reality simulators for TVUS in obstetrics and gynaecology have been published in the literature at the time of submission of this thesis. This finding correspond with the approach adopted in other studies, which confirmed face validity by seeking experts' and novice opinion and content validity as determined by experts' rating (Kenney et al., 2009; Gavazzi et al., 2011; Schreuder et al., 2014; de Vries et al., 2016).

Validation of simulation metrics with automated feedback is extremely important for any simulation system. In this thesis, the findings showed no significant difference between simulation metrics feedback and the observer ratings, which indicates that simulation metrics are robust and consistent with human judgement. Additionally, simulation metrics-based performance was able to provide accurate evaluation of trainees' level of practical skill in gynaecology and early pregnancy ultrasound, and coupled with feedback, provided an effective means of assisting trainees to improve their performance.

Studies related to the reliability of the ScanTrainer® simulator were consistent with studies in other medical specialities (Newble D, 2004; Norcini J, 2005; van der Vleuten and Schuwirth, 2005), such as in laparoscopic procedures (Stefanidis et al., 2009) and Cardiopulmonary resuscitation (Al-Rasheed et al., 2013). We were unable

to find comparable studies utilising ultrasound simulators. The findings showed high reliability between simulation metrics and the observer, similar to the findings of Scalese et al. (2008) and unbiased outcome as concluded in other published studies (Tarara et al., 2014; Yanes et al., 2016). The findings were also consistent with those studies in that ultrasound performance varies considerably. This leads to the conclusion that assessment of recorded simulated performance is an effective method to rehearse the assessment technique among observers as well as improve the quality of assessment. The use of checklists or a global rating scale (GRS) in the assessment of ultrasound skills has been reported previously (Alsalamah et al., 2009; RCOG, 2016; Tolsgaard et al., 2015a), however they vary considerably in the way they were developed or evaluated and hence their applicability to the wider ultrasound community is fundamental. Studies have also shown that an observer's pedagogical beliefs can influence that person's ability to use a rating system as intended and also might conflict with the underlying theoretical foundation of the evaluation system (Henry et al., 2010; Tarara et al., 2014). In this thesis, checklists produced a higher ratings agreement than the use of GRS. In contrast, Larsen et al. (2008) reported that the GRS was more effective than yes/no-based checklists when using video recordings of laparoscopic gynaecological procedures to improve quality assurance. Several authors had argued that despite the fact that the checklist is more objective and rules out partiality, the range of scores available in a GRS could improve reliability by allowing more variation in ratings but reducing the exact agreement on a particular score (Penny et al., 2000; Scott et al., 2000; Graham et al., 2012; Barry et al., 2013; Tolsgaard et al., 2013). The subjectivity of a GRS scoring system enables it to give more feedback on performance, whereas the checklist scoring system is limited to deciding whether the performance is a pass or a fail. Given all the above factors, the checklist was considered to be the more appropriate method with simulation assessment.

The results of the end-of-trial survey draw attention to important issues related to current ultrasound training in obstetrics and gynaecology. A high number of responses highlighted that a major benefit of the ultrasound simulator was the self-directed learning element tool and that it provided feedback for an individual's performance without the need for a physical instructor to be present. In addition, the regular feedback allows broader conclusions to be drawn on a trainee's progress. This

leads to the conclusion that simulation training supplemental to clinical training had positive effects on trainees' learning experience, performance and skills. This is consistent with Williams et al (2013) in that training on the ScanTrainer® was found to affect a trainee's confidence to progress to clinical scanning. However, some limitations were reported about the ScanTrainer®, which were similar to the results of many studies with simulators (Chalouhi et al., 2015a). Limitations include the absence of blood flow (colour Doppler) or fetal movements during the examination, as well as the absence of interaction between the candidate and the patient. Some of these limitations could be resolved by improving the quality of the simulation system while other may be prohibited for the time being due to technological challenges.

To summarise, the findings of this research were encouraging and confirmed several important aspects about validation and evaluation of the simulator. They also demonstrated the initial steps to reach competency in learning TVUS scans and that trainees should not be considered fully competent after training on a virtual reality simulator. Konge et al. (2015) proposed a three-step approach consisting of learning the necessary anatomy and theory (step one), simulation-based training (step two), and supervised practice on patients (step three), before performing independent procedures. Testing can ensure that basic (early phase) competency is achieved, accelerate learning and improve performance. The authors proposed that all three steps should end with the ability to identify the competence level of the trainee before proceeding to the next step (Konge et al., 2015; Lucereau et al., 2016). Further investigations such as predictive validity to determine transferability of skills to clinical practice are crucial. As eloquently stated by Issenberg et al. (2001), "evidence-based outcomes must show these systems to be effective instruments for teaching and assessment, and medical educators must be willing to effect change in medical education to ensure the appropriate use of these systems in the next millennium".

Chapter ten

Study limitations, future work and Conclusion

CHAPTER 10

Limitations, future work and conclusion

10.1 Possible limitations

Possible limitations of the current study were identified for each experiment conducted in the thesis. In face and content validity study, higher ratings given by non-experts than experts with regard to the relevance of the simulator to actual TVUS and its realism to simulate the movements required to perform an examination of the female pelvis, highlighted the fact that such realism is crucial for non-experts for several reasons. This may be because experts need to develop greater understanding of the strengths and limitations of the simulator compared to trainees (Shanmugan et al., 2014). Alternatively, beginners in the early stages of learning ultrasound skills were able to address their learning needs through simulated learning compared to the experts who expect variety and advanced or more complex performance rather than basic tutorials (Hung et al., 2011). Virtual ultrasound and haptics were used instead of a mannequin to improve realism of the simulation movement, allowing measurement of the force applied to the probe and providing somewhat realistic force-feedback during scanning. However, this had the limitation of allowing a lower range of movements to the probe while lacking a simulated environment exemplified by the absence of a physical mannequin (Chalouhi et al., 2015a). Another potential limitation of face and content validity study was the small sample size. However the sample size was larger than those in Hung et al.'s (2011) study, where a sample size of six participants in each of the expert and pure novice groups was deemed adequate to achieve significance at 80% statistical power on the basis of available literature data.

A limitation of the intra– and inter observer reliability scoring systems developed for the assessment of obstetrics and gynaecology ultrasound skills, was the use of video recordings that either revealed or did not reveal the anatomy of the female pelvis.

Those that did so enabled the observer to monitor the movement of the probe during the examination easily. However, this difficulty is unavoidable because video recordings of trainee performances with revealed and concealed anatomy were recorded for the randomised controlled study. The use of checklists and global rating scoring (GRS) systems had its limitations. Despite the fact that the checklist was more objective and ruled out partiality, the range of scores available in a GRS could improve reliability by allowing more variation in ratings but reducing the exact agreement on a particular score. The subjectivity of a GRS scoring system enables it to give more feedback on performance, whereas the checklist scoring system is limited to deciding whether the performance is a pass or a fail. Discrepancy between observers' ratings may have been related to the observer's degree of familiarity with the trainee/person who was being evaluated, thus familiarity may encourage bias. Moreover, it has been reported that the use of recorded videos for the evaluation of trainee performance leads to less accurate evaluation than the use of direct observation of the trainees' performance. Direct contact with the trainees enabled them to explain, clarify and identify some issues during the examination and this was offered for one observer but not to another one who was blindly evaluating the performance without awareness of the trainees' lack of knowledge. Possible solutions to improve reliability may include clear instructions given to the trainee at the onset and to add audio to the video recording, so the trainee explains her/his actions.

With validation of simulation metrics, limitation sometimes leads to flawed conclusions of correct or optimal practice, as the metric has inflexible and limited ability in comparison to human assessors. In other words, sub-optimal performance of ultrasound skills performed in the simulator may be considered as a failure while the same performance would be accepted if it occurred in a real scan. Again, the limitations of computerised settings in the simulator should be considered and addressed. In addition, the small sample size and the self-selection method to invite participants may limit the conclusions although self-selection method was widely used and accepted. This may reflect inherent bias as participation inevitably required participants' interest in taking part in the study. In the construct validity study, the lack of significant difference between intermediate and expert scores on all TVUS skills included in the assessment was one of the study's limitations. It suggested that if the intermediates and experts were allowed to practice prior to collecting data, their

scores could have been significantly different in more tasks. In addition, the small sample size may have contributed to the lack of significant difference.

In the randomised controlled trial limitations, there were a number of factors that may have affected the significance of the findings. The heterogeneity of the subjects randomly assigned to intervention and control impacted on the slope of the curves and significance of outcomes. In addition, small sample size in the subgroups limited the degree of significance. Other possible factors were limited access to the simulation location, unprotected time given to the trainees to practise on the simulator and lack of clinical ultrasound training sessions scheduled for trainees, which altogether impacted on the learning curves and significance of outcomes.

Despite the fact that the virtual reality environment is an alternative option that would assist in learning a systematic approach towards ultrasonography and offers a scenario that is very close to the real one, trainees' perceptions about the limitations of the ScanTrainer® varied and they stated that (1) simulation training was neither realistic enough, nor had similar experience as on patients, (2) the difficulty in completing a number of tutorials because of inadequate explanation, and (3) the lack of patient interaction and/or the simulator not being realistic enough to feel a frozen pelvis or tenderness. In order to overcome some of these limitations, the PhD researcher (A.A.) was available to assist, guide and assess trainees on a regular basis and to discuss obstacles they encountered during simulation practice. This enhanced trainees' understanding of the inherent weaknesses in their performance.

10.2 Strength of research

The main strength element in this research was that it was being conducted within a non-controlled learning environment unlike most of the published studies which were conducted within a controlled learning environment. Moreover, similar opportunities were offered to all participants under identical clinical training conditions. This led to concluding that the amount of training received, whether simulation or clinical training, reflected the actual current training in obstetrics and gynaecology ultrasound. This also enhanced our understanding of potential barriers and limitations of learning TVUS skills that mimicked trainees' motivation during the learning process. On the other hands, the virtual reality ultrasound simulator ScanTrainer® is a computerised

system that allows for developing, updating and improving its system in order to enhance the quality of training and assessment.

10.3 Future work

Future studies should focus on improving the realism of ScanTrainer® and assessing the validity of new simulation metrics such as a construct validity study to differentiate between high levels of ultrasound experience; experts and intermediate practitioners, with large sample size and preferably with more advance simulation modules. This would help enhance our understanding of performance standard and determine trainees' level of competence in attaining expert level. In addition, a predictive validity study using ScanTrainer® to determine the transferability of skills to clinical practice is essential.

10.4 Conclusion

This multidimensional research set out with the overall aim to determine the length of time required for trainees to acquire the skills necessary to perform TVUS, with the simulation training supplemental to clinical training. In addition, it identified the validity and reliability of the virtual reality ultrasound simulator ScanTrainer®, and showed high face, content and constructs' validity that supports the research hypotheses. It also has a potential role in the assessment of clinical skills. Nevertheless, the impact of simulation on the learning curves requires further investigation.

References

References

[1.] Abrahamson, S. et al. 1969. Effectiveness of a simulator in training anesthesiology residents, *Journal Medical Education* 44, pp. 515-519.

[2.] Ackerman, J. 1998. The Visible Human Project. *Proc IEEE* 86, pp. 504-511.

[3.] Adamchak, S. et al. 2000. *A guide to monitoring and evaluating adolescent reproductive health programs*. Tool Series No. 5, Pathfinder International, Washington, D.C.

[4.] Adler, M. et al. 2007. Development and evaluation of high fidelity simulation case scenarios for paediatrics resident education. *Ambul Pediatr.* 7(2), pp. 182-186.

[5.] Aggarwal, R. et al. 2007. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg.* 246(5), pp. 771-779.

[6.] Ahmed, M. et al. 1996. Virtual reality in medicine. *British Journal of Urology.* 80(3), pp. 46-52.

[7.] Ahmed, K. et al. 2011. Educational research in urology: current status and future challenges. *BJU International.* 107(12), pp. 1872-1873.

[8.] Aiello, P. et al. 2012. A Constructivist Approach to Virtual Reality for Experiential Learning. *E-learning and digital media.* 9(3), pp. 317-324.

[9.] Al-Elq, A. 2010. Simulation-based medical teaching and learning. *J Family Community Med.* 17(1), pp. 35–40.

[10.] Al-Rasheed, R. et al. 2013. Simulation intervention with manikin-based objective metrics improves CPR instructor chest compression performance skills without improvement in chest compression assessment skills. *Simul Healthcare.* 8(4), pp. 242-252.

[11.] Alsalamah, A. et al. 2009. *The effectiveness of 'Blended Learning Environment' on ultrasound skills and competencies in Obstetrics and Gynaecology*. MSc thesis, Cardiff University

[12.] Alzahrani, T. et al. 2013. Validation of the da Vinci Surgical Skill Simulator across three surgical disciplines. *Canadian Urological Association Journal.* 7, pp.7-8.

[13.] Amso, N. and Griffiths, A. 2005. The role and applications of ultrasound in ambulatory gynaecology. *Best Pract Res Clin Obstet Gynaecol.* 19(5), pp. 693-711.

[14.] Awtrey, C. et al. 2010. The Simulation and Skills Center at Beth Israel Deaconess Medical Center. *J Surg Education.* 67(4), pp. 255-257.

[15.] Aydin, A. et al. 2014. Face and content validation of the prostatic hyperplasia model and holmium laser surgery simulator. *J Surg Educ.* 71(3), pp. 339-344.

[16.] Bahner, D. et al. 2013. Integrated medical school ultrasound: development of an ultrasound vertical curriculum. *Crit Ultrasound J.* 5(1), p. 6

[17.] Bajura, M. et al. 1992. Merging virtual objects with real world: seeing ultrasound imagery within the patient. *Associated for Computing Machinery.* 26(2), pp. 203-210.

[18.] Bakker, N. 2008. *Design and Evaluation of Video Portfolios: Reliability, generalizability, and validity of an authentic performance assessment for teachers*. Holland Netherlands Organisation for Scientific Research.

[19.] Bandura, A. 2001. Social cognitive theory: an agentic perspective. *Annual Review of Psychology.* 52, pp. 1-26

[20.] Barry, M. et al. 2013. Improving the content and face validity of OSCE assessment marking criteria on an undergraduate midwifery programme: a quality initiative. *Nurse Educ Pract.* 13(5), pp. 477-480.

[21.] Barsuk, J. et al. 2009. Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Crit Care Med.* 37(10), pp. 2697-2701.

[22.] Bayona, S. et al. 2009. A new assessment methodology for virtual reality surgical simulators. *Computer Animation and Virtual Worlds.* 20(1), pp. 39–52.

[23.] Beard, J. et al. 2005. Assessing the technical skills of surgical trainees. *British Journal Surgery.* 92(6), pp. 778-782.

[24.] Beyer, L. et al. 2011. Impact of laparoscopy simulator training on the technical skills of future surgeons in the operating room: a prospective study. *Am J Surg* 202(3), pp. 265-272.

[25.] Biese, K. et al. 2009. Using screen-based simulation to improve performance during pediatric resuscitation. *Acad Emerg Med.* 16(Suppl 2), pp. S71-75.

[26.] Billings, C. and Reynard, W. 1984. Human factors in aircraft accidents: Results of a 7-year study. *Aviation, Space, and Environmental Medicine.* 55(10), pp. 960-965.

[27.] Blättler, C. et al. 2011. Representational Momentum in Aviation. *Journal of experimental Psychology: HUman perception and performance* 37(5), pp. 1569-1577.

[28.] Blum, T. et al. 2013. A review of computer-based simulators for ultrasound training. *Simul Healthcare.* 8(2), pp. 98-108.

[29.] Bø, L. et al. 2010. Efficiency of ultrasound training simulators: method for assessing image realism. *Minimally Invasive Therapy and Allied Technologies* 19(2), pp. 69-74.

[30.] Bose, R. et al. 2009. Transesophageal echocardiography simulator: a new learning tool. *J Cardiothorac Vasc Anesth.* 23(4), pp. 544-548.

[31.] Boulet, J. et al. 2003. Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology*. 99(6), pp. 1270-1280

[32.] Bowling, C. et al. 2010. Testing and validation of low-cost cystoscopy teaching model: Randomised controlled trial. *Obstet Gynecol.* 116(1), pp. 85-91.

[33.] Bradly, P. 2006. History of simulation in medical education and possible future directions. *Medical Education.* 4(3), pp. 254-262.

[34.] Brehmer, M. and Swartz, R. 2005. Training on bench models improves dexterity in ureteroscopy. *Eur Urol*. 48(3), pp. 458-463.

[35.] Bretz, R. et al. 1992. The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management.* 18, pp. 321–352.

[36.] Bright, E. et al. 2012. Face validity, construct validity and training benefits of a virtual reality turp simulator. *International Journal of Surgery* 10(3), pp. 163-166.

[37.] Bruner, J. 1966. *Toward a Theory of Instruction*. Harvard University Press Cambridge, MA

[38.] Bruppacher, H. et al. 2010. Simulation-based training improves physicians' performance in patient care in high-stakes clinical setting of cardiac surgery. *Anesthesiology.*112, pp. 985–992

[39.] Burckett-St Laurent, D. et al. 2016. Teaching ultrasound-guided regional anesthesia remotely: a feasibility study. *Acta Anaesthesiol Scand.* 0(0).

[40.] Burden, C. et al. 2011. Integration of laparoscopic virtual-reality simulation into gynaecology training. *BJOG.* 118(3), pp. 5-10.

[41.] Byrne, A. and Greaves, J. 2001. Assessment instruments used during anaesthetic simulation: review of published studies. *Br. J. Anaesth.* 86(3), pp. 445-450.

[42.] Candy, P. 1991. *Self-direction for Lifelong Learning. A comprehensive guide to theory and practice*. San Francisco: Jossey-Bass.

[43.] Carney, P. et al. 2016. Tools to Assess Behavioral and Social Science Competencies in Medical Education: A Systematic Review. *Acad Med.* 0(0).

[44.] Carolan-Rees, G. and Ray, A. 2015. The ScanTrainer obstetrics and gynaecology ultrasound virtual reality training simulator: A cost model to determine the cost viability of replacing clinical training with simulation training. *Ultrasound.* pp. 1-6

[45.] Carter, F. et al. 2005. Consensus guidelines for validation of virtual reality surgical simulators. *Surg Endosc.* 19(12), pp. 1523-1532.

[46.] Cash, A. et al. 2012. Rater calibration when observational assessment occurs at large scale: Degree of calibration and characteristics of raters associated with calibration. *Early Childhood Research Quarterly.* 27, pp. 529-542.

[47.] Chalasani, V. et al. 2011. Development and validation of a virtual reality transrectal ultrasound guided prostatic biopsy simulator. *Can Urol Assoc J.* 5(1).

[48.] Chalouhi, G. et al. 2015a. Evaluation of trainees' ability to perform obstetrical ultrasound using simulation: challenges and opportunities. *Am J Obstet Gynecol.* S0002-9378(15), pp. 02237-02231.

[49.] Chalouhi, G. et al. 2015b. Ultrasound simulators in obstetrics and gynecology: state of the art. *Ultrasound in Obstetrics & Gynecology.* 46(3), pp. 255-263.

[50.] Chen, H. et al. 2015. Improving Trainee Competency and Comfort Level with Needle Driving Using Simulation Training. *Pain Med.* 0(0), pp. 56.

[51.] Chenkin, J. et al. 2015. Defining the learning curve of point-of-care ultrasound for confirming endotracheal tube placement by emergency physicians. *Crit Ultrasound J.* pp. 7-14.

[52.] Chopra, C. et al. 1994. The Leiden anaesthesia simulator. *British Journal of Anaesthesia.* 73, pp. 287-292.

[53.] Choudhry, N. et al. 2005. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med.* 142(4), pp. 260-273.

[54.] Chung, G. et al. 2011. *The Effects of Practicing With a Virtual Ultrasound Trainer on FAST Window Identification, Acquisition, and Diagnosis. National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*. University of California, Los Angeles.

[55.] Cline, B. et al. 2008. Human performance metrics for a virtual reality simulator to train chest tube insertion. *IEEE Systems and Information Engineering Design Symposium*, pp. 168 - 173.

[56.] Conaghan, P. and Lockey, A. 2009. Feedback to feedforward. A positive approach to improving candidate success. *Notfall Rettungsmed.* Suppl 2 (12), pp. 45–48.

[57.] Controlled-trials. 2013. Online avaiable at http://www.controlled-trials.com/ISRCTN53915329?q=&filters=conditionCategory:Not%20Applicable&sort=&offset=196&totalResults=636&page=2&pageSize=100&searchType=basic-search [Accessed 18/02/2013].

[58.] Cook, D. et al. 2011. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. *JAMA.* 306(9), pp. 978-988.

[59.] Cooper, J. and Taqueti, V. 2004. A brief history of the development of mannequin simulators for clinical education and training. *Qual Saf Health Care.* 13(Suppl 1), pp. 11-18

[60.] General Medical Council (GMC) 2016. Borderline group scoring method. [online] available at http://www.gmc-uk.org/doctors/plab/borderline_group_scoring_faqs.asp [accessed 22/02/2016].

[61.] Cox, S. et al. 1994. Assessment of the in-training examination in obstetrics and gynaecology. *Obstet Gynecol*. 84(6), pp. 51-54.

[62.] Cremers, S. et al. 2005. Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). *Ophthalmology.* 112(10), pp. 1655–1660.

[63.] Crofts, J. et al. 2006. Training for shoulder dystocia: a trial of simulation using low-fidelity and high-fidelity mannequins. *Obstet Gynecol.* 108(6), pp. 1477-1485.

[64.] de Vries, A. et al. 2016. The Simbla TURBT Simulator in Urological Residency Training: From Needs Analysis to Validation. *J Endourol.*

[65.] de Wit-Zuurendonk, L. and Oei, S. 2011. Serious gaming in women's health care. *BJOG.* 118 (3), pp. 17-21.

[66.] Deering, S. et al. 2009. Use of a postpartum haemorrhage simulator for instruction and evaluation of residents. *J Graduate Med Education.* 1(2), pp. 260-263.

[67.] Denson, J. and Abrahamson, S. 1969. A computer-controlled patient simulator.*JAMA.* 208, pp. 504-508. Cited in: van Merurs WL (2006) Evolution of acute care simulation: a European perspective. Medisch Journaal. 2035:2116.

[68.] Derossis, A. et al. 1998. Development of a model for training and evaluation of laparoscopic skills. *American Journal of Surgery.* 175, pp. 482-487.

[69.] Dillard, A. 2002. Validation of advanced flight simulators for human-factors operational evaluation and training programs. *Foundation '02 V7V Workshop.* John Hopkins University Applied Physics Laboratory.

[70.] Distlehorst, L. et al. 2000. Teaching and learning in medical and surgical education; lessons learned for the 21st Century. *Lawrence Erlbaum Associate Inc.*, p. Mahwah.

[71.] Donovan, J. and Radosevich, D. 1999. A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology.* 84(5), pp. 795-805.

[72.] Downing, M. et al. 2006. RESEARCH METHODOLOGY: Procedures for Establishing Defensible Absolute Passing Scores on Performance Examinations in Health Professions Education, *Teaching and Learning in Medicine*, 18(1), pp. 50-57

[73.] Draycott, T. et al. 2008. Improving Neonatal Outcome Through Practical Shoulder Dystocia Training. *Obstet & Gynecol.* 112(1), pp. 14-20.

[74.] Dreyfus, S. and Dreyfus, H. 1980. *A five stage model of the mental activities involved in directed skill acquisition*. Unpublished report supported by the Air Force Office of Scientific Research (AFSC) USAF. Berkeley: University of California.

[75.] Dupon, W. and Plummer, W. 1990. *Power and Sample Size Calculations: A Review and Computer Program, Controlled Clinical Trials 1990*, 11:116–128. The program is used for power and sample size calculation is PS. [Online] available at http://biostat.mc.vanderbilt.edu/PowerSampleSize

[76.] Dulan, G. et al. 2012. Content and face validity of a comprehensive robotic skills training program for general surgery, urology, and gynecology. *American Journal of Surgery.* 203(4), pp. 535-539.

[77.] Edwards, S. and Rajaratnam, V. 2009. Digital video documentation as evidence of clinical skill acquisition. *J Vis Commun Med.* 32((3-4)), pp. 78-85.

[78.] Ehricke, H. 1998. SONOSim3D: a multimedia system for sonography simulation and education with an extensible case database. *Eur J Ultrasound.* 7(3), pp. 225-300.

[79.] Ennen, C. and Satin, A. 2010. Training and assessment in obstetrics: the role of simulation. *Best Pract Res Clin Obstet Gynaecol.* 24(6), pp. 747-758.

[80.] Erdogan, A. et al. 2016. Development and validation of clinical performance assessment in simulated medical emergencies: an observational study. *BMC Emerg Med.* 16(1), p. 4.

[81.] Ericsson, K. et al. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev.* 100, pp. 363– 406.

[82.] Ericsson K 2007. Deliberate practice and the modifiability of body and mind: toward a science of the structure and acquisition of expert and elite performance. *International J Sport Psychology.* 38(4), pp. 4-34.

[83.] Evans, L. et al. 2010. Simulation training in central venous catheter insertion: improved performance in clinical practice. *Acad Med.* 85(9), pp. 1462-1469.

[84.] Eversbusch, A. and Grantcharov, T. 2004. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surg Endosc.* 18(10), pp. 18(10):1514-1518.

[85.] Federal Aviation Authority, FAA. 1990. Line-oriented flight training programs. *Advisory Circular*, pp. 120-135.

[86.] Feinstein, A. and Cannon, H. 2001. *Fidelity, verifiability, and validity of simulation: Constructs for evaluation*. Developments in Business Simulation and Experiential Learning.

[87.] Feldman, L. et al. 2004. Using simulators to assess laparoscopic competence: ready for widespread use?. *Surgery.* 135(1), pp. 28-42.

[88.] Ferro, T. 1993. *The influence of affective processing in education and training. In: Flannery DD. Applying Cognitive Learning Theory to Adult Learning*. San Francisco: Jossey-Bass, pp. 25–33.

[89.] Fisher, D. et al. 2010. *Handbook of Driving Simulation for Engineering, Medicine, and Psychology.* CRC Press, Taylor & Francis Group, Boca Raton, Florida.

[90.] Fisher, J. et al. 2006. Development and face and content validity of an eye surgical skills assessment test for ophthalmology residents. *Ophthalmology.* 113(12), pp. 2364-2370.

[91.] Fletcher, G. et al. 2003. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *British Journal of Anaesthesia.* 90(5), pp. 580-588.

[92.] Flight simulation. 2014. [online] available at http://pong-story.com/intro.htm [Accessed: 16/8/14].

[93.] Forrest, K. and McKimm, J. 2010. Using simulation in clinical education. *Br J Hosp Med (Lond).* 71(6), pp. 345-349.

[94.] Fried, G. et al. 2004. Proving the Value of Simulation in Laparoscopic Surgery. *Annals of Surgery* 240(3), pp. 518-528.

[95.] Gaba, D. and DeAnda, A. 1988. A comprehensive anaesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology*. 69, pp. 387-394.

[96.] Gaba, D. 2004. The future vision of simulation in health care. *Qual Saf Health Care.* 13(Suppl 1), pp. ii2–i10.

[97.] Gallagher, A. et al. 2012. Prospective, randomized assessment of the acquisition, maintenance, and loss of laparoscopic skills. *Annals of Surgery.* 256(2), pp. 387-393.

[98.] Gallagher, C. and Tan, J. 2010. The current status of simulation in the maintenance of certification in Anesthesia. *International Anesthesiology Clinics*. 48(3), pp. 83-99.

[99.] Garuda, S. et al. 2002. *Efficacy of a computer-assisted Endoscopic simulator in training Residents in Flexible Sigmoidoscopy*. American College of Gastroenterology Seattle ,WA.

[100.] Gavazzi, A. et al. 2011. Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). *Ann R Coll Surg Engl.* 93(2), pp. 152-156.

[101.] Gibbs, V. 2014. A proposed new clinical assessment framework for diagnostic medical ultrasound students. *Ultrasound.* 22(2), pp. 113-117.

[102.] Gilliam, A. and Acton, S. 2007. Echocardiographic Simulation for Validation of Automated Segmentation Methods. *Image Processing, 2007. ICIP 2007. IEEE International Conference.* 5, pp. 529 - 532.

[103.] Girzadas, D. et al. 2009. Hybrid simulation combining a high fidelity scenario with a pelvic ultrasound task trainer enhances the training and evaluation of endovaginal ultrasound skills. *Acad Emerg Med.* 16(5), pp. 429-435

[104.] Good, M. and Gravenstein, J. 1989. Anesthesia simulator and training devices. *Int Anesth Clinical.* 27, pp. 161-166

[105.] Good, M. 2003. Patient simulation for training basic and advanced clinical skills. *Medical Education*. 37(1), pp. 14-21.

[106.] Goodyear, P. 1999. *Pedagogical frameworks and action research in open and distance learning*. Centre for Studies in Advanced Learning Technology (CSALT), Dept. of Educational Research, Lancaster University, Lancaster, United Kingdom.

[107.] Gordon, M. et al. 2012. Non-technical skills training to enhance patient safety: a systematic review. *Med Educ*. 46, pp.1042–1054.

[108.] Gould, D. 2010. Using simulation for interventional radiology training. *The British Journal of Radiology.* 83, pp. 546–553.

[109.] Graham, M. et al. 2012. *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. The Center for Educator Compensation Reform (CECR). pp.33.

[110.] Grober, E. et al. 2010. Intraoperative assessment of technical skills on live patients using economy of hand motion: establishing learning curves of surgical competence. *Am J Surg.* 199(1), pp. 81-85.

[111.] Grossman, E. 1919. Dry Shooting for Airplane Gunners. *Popular Science Monthly,* pp. 13-14.

[112.] Gulgin, H. and Hoogenboom, B. 2014. The functional movement screening (FMS)™: an inter-rater reliability study between raters of varied experience. *The International Journal of Sports Physical Therapy.* 9(1), pp. 14-20.

[113.] Gupta, V. and Devarajan, V. 2004. *Realistic anatomical texture for laparoscopic surgery simulation:. medicine meets virtual reality*, Newport Press, pp. 124-126.

[114.] Gurusamy, K. et al. 2008. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *British Journal of Surgery.* 95(9), pp. 008-1097.

[115.] Haerizadeh, H. and Frappell, J. 2014. Simulation training in minimal access gynaecology. *Obstetrics, Gynaecology and Reproductive Medicine.* 24(4), pp. 122-124.

[116.] Hallgren, A. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psycol.* 8(1), pp. 23-34.

[117.] Hatala, R. et al. 2008. Assessing cardiac physical examination skills using simulation technology and real patients: a comparison study. *Med Educ.* 42(6), pp. 628-636.

[118.] Hay, C. et al. 2015. Can an Educational Simulation Among Obstetrics and Gynecology Residents Improve Competency in Obtaining a Pap Test?. *Obstetrics & Gynecology.* 126(Suppl 4), pp. 4:47

[119.] Haywood, K. et al. 2004. Reviewing measures of outcome: reliability of data extraction. *J Eval Clin Pract.* 10(2), pp. 329-337.

[120.] Health and Social Care Information Centre (HSCIC) 2012. Health and Social Care Information Centre, NHS Maternity statistics England April 2010 to March 2011, London. [online] Available at http://www.hscic.gov.uk/article/2021/Website-Search?productid=10061&q=miscarriage&sort=Relevance&size=10&page=3&area=both - top [Accessed 10/02/2016].

[121.] Department of Healthcare. 2011. A framework of technology enhanced learning. [online] Available at https://www.gov.uk/government/publications/a-framework-for-technology-enhanced-learning [Accessed 16.08.2014]

[122.] Hellaby, M. 2013. *Healthcare simulation in practice: Simulation in a virtual area*. First edition. M&K publishing, Keswich. Cumbria.

[123.] Henry, A. et al. 2010. *Rater calibration when observational assessment occurs at large-scale: Degree of calibration and characteristics of raters associated with calibration*. Doctoral dissertation, University of Virginia.

[124.] Higdon, D. 2008. Flight Training Simulators, Simulator review. *World Aircraft Sales Magazine*, pp. 120-124.

[125.] Hogle, N. et al. 2007. Documenting a learning curve and test-retest reliability of two tasks on a virtual reality training simulator in laparoscopic surgery. *J Surg Educ.* 64(6), pp. 424-430.

[126.] Honey, M. and Hilton, M. 2011. *Learning Science: Computer Games, Simulations, and Education*. Washington DC: National Academies Press.

[127.] Hsi, R. et al. 2014. Content and face validation of a curriculum for ultrasonic propulsion of calculi in a human renal model. *J Endourol.* 28(4), pp. 459-463.

[128.] Hsu, J. et al. 2016. Force feedback vessel ligation simulator in knot-tying proficiency training. *Am J Surg.* 221(2), pp. 411-415.

[129.] Huang, H. 2002. Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology.* 33(1), pp. 27–37.

[130.] Hume, W. 1983. Aviation Medicine? *British Medical Journal.* 287, pp. 24-31.

[131.] Hung, A. J. et al. 2011. Face, content and construct validity of a novel robotic surgery simulator. *J Urol*. 186(3), pp. 1019-1024.

[132.] Hunt, E. et al. 2008. Simulation of in-hospital pediatric medical emergencies and cardiopulmonary arrests: highlighting the importance of the first five minutes. *Pediatrics.* 121(1), pp. 34-43.

[133.] International Society of Ultrasound in Obstetrics and Gynecology ISUOG. 2014. ISUOG Education Committee recommendations for basic training in obstetric and gynecological ultrasound, *Ultrasound Obstet Gynecol.* 43, pp. 113-116

[134.] Issenberg, S. et al. 2001. Simulation and new learning technologies. *Medical Teacher.* 23(1), pp. 16-23.

[135.] Issenberg, S. et al. 2005. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach.* 27(1), pp. 10-28.

[136.] Jaffer, U. et al. 2014. Validation of a novel duplex ultrasound objective structured assessment of technical skills (DUOSATS) for arterial stenosis detection. *Heart Lung Vessel.* 6(2), pp. 92–104.

[137.] Jalink, M. et al. 2014. Construct and concurrent validity of a Nintendo Wii video game made for training basic laparoscopic skills. *Surg Endosc.* 28(2), pp. 537-542.

[138.] Janse, J. et al. 2013. Hysteroscopic sterilization using a virtual reality simulator: assessment of learning curve. *J Minim Invasive Gynecol.* 20(6), pp. 775-782.

[139.] Jensen, M. et al. 2003. Interpretation of visual analog scale ratings and change scores: a reanalysis of two clinical trials of postoperative pain. *J Pain.* 4(7), pp. 407-414.

[140.] Jones, J. et al. 1997. Assessing bedside cardiologic examination skills using ‘‘Harvey’’ a cardiology patient simulator. *Acad Emerg Med.* 4(10), pp. 980-985.

[141.] Kaufman, D. 2003. Applying educational theory in practice. *BMJ.* 25, pp. 213-216.

[142.] Kazemi, H. et al. 2010. Assessing suturing techniques using a virtual reality surgical simulator. *Microsurgery.* 30(6), pp. 479-486.

[143.] Kearsley, G. 2010. The theory Into practice database. [Online] available at http://tip.psychology.org [Accessed 22.02.2016].

[144.] Kelly, D. et al. 2012. Face, content, and construct validation of the da Vinci Skills Simulator. *Urology.* 79(5), pp. 1068-1072.

[145.] Kenney, P. et al. 2009. Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. *Urology*. 73(6), pp. 1288-1292.

[146.] Kim, J. et al. 2009. A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). *Simul Healthc* 4(1), pp. 6-16.

[147.] Kim, E. et al. 2016. The effect of electromagnetic guidance system on early learning curve of ultrasound for novices. *Korean J Anesthesiol.* 69(1), pp. 15-20.

[148.] King, N. et al. 2016. A Review of Endoscopic Simulation: Current Evidence on Simulators and Curricula. *J Surg Educ.* 73(1), pp. 12-23.

[149.] Kirkpatrick, D. 1994. *Evaluating training programs; the four levels*. Berrett-Koehler Publishers, San Francisco, CA.

[150.] Kneebone, R. 2003. Simulation in surgical training: educational issues and practical implications. *Medical Education.* 37(1), pp. 267-277.

[151.] Knowles, M. 1975. *Self-directed Learning: a Guide for Learners and Teachers*. Chicago, IL: Follett.

[152.] Knudson, M. and Sisley, A. 2000. Training Residents Using Simulation Technology: Experience with Ultrasound for Trauma. *Journal of Trauma.* 48(4), pp. 659-666.

[153.] Kodz, J. 2003. *Working long hours, a review of the evidence*. Department of Trade and Industry.London.UK

[154.] Kohn, L. et al. 2000. *To Err is human: Building a better health care system*. National Academy of Sciences, National Academies Press, Washington, DC.

[155.] Kolozsvari, N. et al. 2011. Sim one, do one, teach one: considerations in designing training curricula for surgical simulation. *J Surg Educ.* 68(5), pp. 421-427.

[156.] Konge, L. et al. 2013. Using virtual-reality simulation to assess performance in endobronchial ultrasound. *Respiration.* 86(1), pp. 59-65.

[157.] Konge, L. et al. 2015. Simulator training for endobronchial ultrasound: a randomised controlled trial. *Eur Respir J.* 46(4), pp. 1140-1149.

[158.] Kovacs, G. et al. 2000. A randomized controlled trial on the effect of educational interventions in promoting airway management skill maintenance. *Ann Emerg Med.* 36(4), pp. 301-309.

[159.] Kumar, A. et al. 2015. Current status of robotic simulators in acquisition of robotic surgical skills. *Curr Opin Urol.* 25(2), pp. 168-174

[160.] Kundhal, P. and Grantcharov, T. 2009. Psychomotor performance measured in a virtual environment correlates with technical skills in the operating room. *Surg Endosc.* 23(3), pp. 645-649.

[161.] Kunkler, K. 2006. The role of medical simulation: an overview. *Int J Med Robot.* 2(3), pp. 203-210.

[162.] Kutz, M. et al. 2006. *Engineering Applications of Virtual Reality. Mechanical Engineers' Handbook: Materials and Mechanical Design*, Volume 1, Third Edition.

[163.] Lang, T. and Secic, M. 2006. *How to report statistics in Medicine: Annotated guidelines for authors, editors and reviewers*. American College of Physicians, Philadelphia.

[164.] Langan, G. et al. 2012. *Transvaginal ultrasound simulation training and its impact on clinical skills development*. SSC Project. Cardiff University.

[165.] Lanier, J. 2011. Virtual Reality Computer Simulation. [online] Available at http://www.referenceforbusiness.com/encyclopedia/Val-Z/Virtual-Reality-Computer-Simulation.html [Accessed 11.12.2011].

[166.] Larsen, C. et al. 2008. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG.* 115(7), pp. 908-916.

[167.] Lateef, F. 2010. Simulation-based learning: Just like the real thing. *J Emerg Trauma Shock.* 3(4), pp. 348-352.

[168.] Letterie, G. 2002. How virtual reality may enhance training in obstetrics and gynecology. *American Journal of Obstetrics and Gynecology.* 187(3), pp. 37-40.

[169.] Lin, Y. et al. 2014. Development and validation of a surgical training simulator with haptic feedback for learning bone-sawing skill. *J Biomed Inform.* 48, pp. 122-129.

[170.] Link, E. 2000. *The Link Flight Trainer*, A historic Mechanical Engineering Landmark. *ASME International* Roberson Museum and Science Centre, Binghamton, New York.

[171.] Loftin, R. 2006. The future of Simulation. . *In Virtual Media for Military Application.* Neuillysur Seine, France. Online available at http://www.rto.nato.int/abstracts.asp [Accessed: 25.10.2011].

[172.] Loveless, M. et al. 2011. A Simulation Program for Teaching Obstetrics and Gynecology Residents the Pediatric Gynecology Examination and Procedures. *Journal of Pediatric and Adolescent Gynecology.* 24(3), pp. 127-136.

[173.] Lucas, L. et al. 2006. Topics in ultrasound education: use of patient models to teach endovaginal ultrasound skills. *Ultrasound Quarterly* 22, pp. 119-120.

[174.] Lucereau, B. et al. 2016. Learning curve of robotic assisted anastomosis: Shorter than the laparoscopic technique? An educational study. *Ann Vasc Surg.* 0(0).

[175.] Lynch, J. et al. 2010. Video games and surgical ability: a literature review. *J Surg Educ.* 67(3), pp. 184-189.

[176.] Macedonia, C. et al. 2003. Simulation Laboratories for training in Obstetrics and Gynaecology. *American College of Obstetricians and Gynaecologists* 102(2), pp. 388-392.

[177.] Macmillan, A. and Cushieri, A. 1999. Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: Predictive and concurrent validity. *Am J Surg.* 177, pp. 274-277.

[178.] Madsen, M. et al. 2014. Assessment of performance and learning curves on a virtual reality ultrasound simulator. *Ultrasound Obstet Gynecol*. 44(6),pp. 693-9

[179.] Magee, D. et al. 2007. An augmented reality simulator for ultrasound guided needle placement training. *Med Biol Eng Comput* 45(10), pp. 957-967.

[180.] Maithel, S. et al. 2006. Construct and face validity of MIST-VR, Endotower, and CELTS, Are we ready for skills assessment using simulators?. *Surg EndosC.* 20, pp. 104–112.

[181.] Mandavia, D. et al. 2000. Ultrasound training for emergency physicians--a prospective study. *Acad Emerg Med.* 7(9), pp. 1008-1014.

[182.] Mann, K. 1999. Motivation in medical education: how theory can inform our practice. *Acad Med*, pp. 237–239.

[183.] Manting, M. 2014. interprofessional Education and obstetric ultrasound. *Donald School J Ultrasound Obstet Gynecol* 8(1), pp. 72-76.

[184.] Markowitz, J. et al. 2011. Development and validation of a web-based assessment tool for the extended focused assessment with sonography in trauma examination. *J Ultrasound Med.* 30(3), pp. 371-375.

[185.] Maul, H. et al. 2004. Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. *Ultrasound Obstet Gynecol.* 24(5), pp. 581-585.

[186.] Mayrose, J. and Myers, J. W. 2007. Endotracheal intubation: application of virtual reality to emergency medical services education. *Simul Healthcare.* 2(4), pp. 231-234.

[187.] McDougall, E. et al. 2006. Construct validity testing of a laparoscopic surgical simulator. *J Am Coll Surg.* 202(5), pp. 779-787.

[188.] Medaphor plc, ScanTrainer®, Cardiff, UK. 2016. [online] Available at http://www.medaphor.com/scantrainer/ [Accessed 30.02.2016].

[189.] Mema, B. and Harris, I. 2016. The Barriers and Facilitators to Transfer of Ultrasound-Guided Central Venous Line Skills from Simulation to Practice: Exploring Perceptions of Learners and Supervisors. *Teach Learn Med.* pp. 5-10.

[190.] Merién, A. et al. 2010. Multi-disciplinary Team Training in a Simulation setting for acute Obstetrics Emergency: A systematic Review. *Obstetrics and Gynaecology.* 115(5), pp. 1021-1031.

[191.] Merz, E. 2006. Ultrasound Simulator - an Ideal Supplemental Tool for Mastering the Diagnostics of Fetal Malformations or an Illusion?. *Ultraschall in Med.* 27(4), pp. 321-323.

[192.] Miller, D. et al. 2012. Improving teamwork and communication in trauma care through in situ simulations. *Acad Emerg Med*. 19, pp. 608–612

[193.] Minardi, H. and Ritter, S. 1999. Recording skills practice on videotape can enhance learning: a comparative study between nurse lecturers and nursing students. *Journal of Advanced Nursing.* 29(6), pp. 1318-1325.

[194.] Mishra, S. et al. 2010. Validation of virtual reality simulation for percutaneous renal access training. *Journal Endourology.* 24(4), pp. 635-640.

[195.] Moak, J. et al. 2014. Training in transvaginal sonography using pelvic ultrasound simulators versus live models: a randomized controlled trial. *Acad Med.* 89(7), pp. 1063-1068.

[196.] Modell, J. et al. 2006. Using the human patient simulator to educate students of veterinary medicine. *Journal of Veterinary Medical Education.* 29(2), pp. 111-116.

[197.] Moglia, A. et al. 2015. A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery. *Eur Urol.* S0302-2838(15), pp. 92-9

[198.] Monsky, W. et al. 2002. Using a sonographic simulator to assess residents before overnight call. *AJR Am J Roentgenol.* 178, pp. 35-39.

[199.] Moore, M. and Bennett, C. 1995. The learning curve for Laparoscopic cholecystectomy. The southern Surgeons Club. *American Journal Surgery,* 170(1), pp. 55-59

[200.] Morgan, H. et al. 2010. *Comparison of Two Different Methods of Teaching Ultrasound Scanning to Medical Students*. SSC Project, Cardiff University.

[201.] Moss, E. et al. 2011. Is gynaecological surgical training a cause for concern?: a questionnaire survey of trainees and trainers. *BMC Med Educ.* 11, pp. 32.

[202.] Muller, M. et al. 2005. Teaching antiarrhythmic therapy and ECG in simulator-based interdisciplinary undergraduate medical education. *Br J Anaesth.* 95(3), pp. 300-304.

[203.] Mullins, S. et al. 2012. *Ultrasound Simulation training and its impact on clinical skills development:barriers to access*. SSC Project. Cardiff University.

[204.] Murugan, M. et al. 2009. *Peer Training of Transvaginal Ultrasound Scanning*. SSC Project. Cardiff University.

[205.] Nara, N. et al. 2009. The introduction and effectiveness of simulation-based learning in medical education. *Intern Med.* 48(17), pp. 1515-1519.

[206.] Newble, D. 2004. Techniques for measuring clinical competence objective structured clinical examinations. *Medical Education.* 38(2), pp. 199-203.

[207.] Nicholson, W. et al. 2006. Face and Content Validation of Virtual Reality simulation for carotid Angiography. *Simulation in Healthcare.* 1(3), pp. 147-150.

[208.] Nitsche, J. et al. 2009. In-utero stenting: development of a low-cost high-fidelity task trainer. *Ultrasound Obstet Gynecol.* 34(6), pp. 720-723.

[209.] Nitsche, J. and Brost, B. 2013. Obstetric ultrasound simulation. *Semin Perinatol.* 37(3), pp. 199-204.

[210.] Norcini, J. 2005. Current perspectives in assessment: the assessment of performance at work. *Med Educ.* 39(9), pp. 880-889.

[211.] O'Leary, S. et al. 2008. Validation of a networked virtual reality simulation of temporal bone surgery. *Laryngoscope.* 118(6), pp. 1040-1046.

[212.] O'Brien, K. et al. 2015. Does training on a Virtual Reality (VR) transvaginal ultrasound simulator, increases the number of ultrasound scans an obstetrics and gynaecology trainee performs?. [online available at] http://www.medaphor.com/scantrainer/publications/.

[213.] Okuda, Y. et al. 2009. The Utility of Simualtion in Medical Education: What Is the Evidence? *Mount Sinai Journal of Medicine.* 76(4), pp. 330-343.

[214.] Oremus, M. et al. 2012. Inter-rater and test-retest reliability of quality assessments by novice student raters using the Jadad and Newcastle-Ottawa Scales. *BMJ Open.* 2(4).

[215.] Overly, F. et al. 2007. High-fidelity medical simulation as an assessment tool for paediatric residents' airway management skills. *Pediatr Emerg Care* 23(1), pp. 11-15.

[216.] Oxford English Dictionary. 2016. Online available at http://www.oed.com [Accessed 22.02.2016].

[217.] Paisly, A. et al. 2001. Validity of surgical simulation for the assessment of operative skill. *Br J Surg.* 88, pp. 152-132.

[218.] Parush, A. et al (2002), learning histories in simulation- based teaching: the effects on self-learning and transfer. *Computer and Education.* 39(4), pp. 319-332

[219.] Pedersen, P. et al. 2014. Virtual-reality simulation to assess performance in hip fracture surgery. *Acta Orthop.* 85(4), pp. 403-407.

[220.] Penny, J. et al. 2000. Using rating augmentation to expand the scale of an analytic rubric. *The Journal of Experimental Education.* 68(3), pp. 269-287.

[221.] Pereira, J. et al. 2007. Effectiveness of using blended learning strategies for teaching and learning human anatomy. *Med Educ* 41(2), pp. 189-195.

[222.] Persoon, M. et al. 2010. A simulator for teaching transrectal ultrasound procedures: how useful and realistic is it? Simulation in healthcare. *The Journal of the Society for Simulation in Healthcare* 5(5), pp. 311-314.

[223.] Pettineo, C. et al. 2009. Simulated emergency department procedures with minimal monetary investment: cricothyrotomy simulator. s*imul Healthcare.* 4(1), pp. 60-64.

[224.] Piaget, J. 1973. *To understand is to invent: the future of education*. Grossman, New York.

[225.] Ping, Y. and Sconing, J. 2008. Estimating standard errors of cut scores for item rating and mapmark procedures: a generalizability theory approach. *Educational and Psychological Measurement* 68(1), pp. 25–41.

[226.] Pittini, R. et al. 2002. Teaching invasive perinatal procedures: assessment of a high fidelity simulator-based curriculum. *Ultrasound Obstet Gynecol.* 19(5), pp. 478-483.

[227.] Platts, D. et al. 2012. The use of computerised simulators for training of transthoracic and transoesophageal echocardiography: the future of echocardiographic training? *Heart Lung and Circulation.* 21(5), pp. 267-274.

[228.] Preshaw, J. et al. 2012. Would It Be Best Practice for Trainees to Learn Ultrasound Scanning in a Simulated Setting Prior to a Clinical Setting?. *International Journal of Gynecology & Obstetrics.* 119, pp. S457.

[229.] Raybourn, E. 2007. Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with Computers.* 19(2), pp. 206-214.

[230.] Rethans, J. et al. 2002. The relationship between competence and performance: implications for assessing practice performance. *Med Educ.* 36(10), pp. 901-909.

[231.] Robertson, J. and Bandali, K. 2008. Bridging the gap: enhancing interprofessional education using simulation. *Journal of Interprofessional Care* 22(5), pp. 499-508.

[232.] Roscoe, S. 1980. *Transfer and cost- effectiveness of ground based flight trainers.* Aviation Psycology Ames: Iowa State University Press, pp. 194-203.

[233.] Rosen, K. 2008. The history of medical simulation. *J Crit Care.* 23(2), pp. 157-166.

[234.] Rosenblatt, M. and Abrams, K. 2002. The use of a human patient simulator in the evaluation of and development of a remedial prescription for an anesthesiologist with lapsed medical skills. *Anesth Analg.* 94(1), pp. 149-153.

[235.] Ross, A. et al. 2013. Simulation training for improving the quality of care for older people: an independent evaluation of an innovative programme for inter-professional education. *BMJ Qual Saf.* 13(22), pp. 495-505.

[236.] Royal College of Obstetrics and Gynaecology, RCOG. 2014. Guidance examination and curriculum. [Online] Available at http://www.rcog.org.uk/education-and-exams/curriculum [Accessed 10.08.2014]

[237.] Royal College of Radiologists RCR. 2005. *Ultrasound training recommendations for medical and surgical specialists*, second edition [Online] Available at http://www.rcr.ac.uk/docs/radiology/pdf/BFCR(12)17_ultrasound_training.pdf [Accessed 16.02.2016]

[238.] Saghaei, M. 2004. Random allocation software for parallel group randomized trials, *BMC Medical Research Methodology*. 4(26), pp. 1-6.

[239.] Salvesen, K. et al. 2010. Basic European ultrasound training in obstetrics and gynecology: where are we and where do we go from here? *Ultrasound Obstet Gynecol.* 36(5), pp. 525-529.

[240.] Sanchez-Peralta, L. et al. 2012. Learning curves of basic laparoscopic psychomotor skills in SINERGIA VR simulator. *Int J Comput Assist Radiol Surg.* 7(6), pp. 881-889.

[241.] Santos-Ribeiro, S. et al. 2016. Trends in ectopic pregnancy rates following assisted reproductive technologies in the UK: a 12-year nationwide analysis including 160 000 pregnancies. *Human Reproduction Journal.* 31(2), pp. 393-402.

[242.] Satava, R. 2001. Accomplishments and challenges of surgical simulation. Dawning of the next-generation surgical education. *Surg Endosc.* 15, pp. 232-241.

[243.] Scalese, R. et al. 2008. Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med.* 23(1), pp. 46-49.

[244.] Schijven, M. and Jakimowicz, J. 2002. Face, expert, and referent validity of the Xitact LS500 laparoscopy simulator. *Surg Endosc.* 16(12), pp. 1764-1770.

[245.] Schlectre, T. et al. 1992. Computer based simulations systems and role-playing: an effective combination for fostering conditional knowledge. *J Comput Based Instruction.* 19, pp. 110-114.

[246.] Schreuder, H. et al. 2009. Face and construct validity of virtual reality simulation of laparoscopic gynecologic surgery. *Am J Obstet Gynecol.* 200(5), pp. 540 e541-548.

[247.] Schreuder, H. et al. 2014. Validation of a novel virtual reality simulator for robotic surgery. *Scientific World Journal.* 2014, p. 30.

[248.] Schulz, C. et al. 2011. Assessment of subjective workload in an anaesthesia simulator environment: reliability and validity. *Eur J Anaesthesiol.* 28(7), pp. 502-505.

[249.] Scott, D. et al. 2000. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. *Surgery.* 128(4), pp. 613–622.

[250.] Scott, D. et al. 2008. The changing face of surgical education: Simulation as the new paradigm. *J Surg Res.* 147(2), pp. 189–193.

[251.] Seixas-Mikelus, S. et al. 2010. Face validation of a novel robotic surgical simulator. *Urology.* 76(2), pp. 357-360

[252.] Seixas-Mikelus, S. et al. 2011. Content validation of a novel robotic surgical simulator. *BJU International.* 107(7), pp. 1130–1135.

[253.] Selbst, S. et al. 1989. Teaching technical skills in pediatrics. *Indian J Pediatr.* 56(1), pp. 35-54.

[254.] Sexton, J. et al. 2000. Error, stress and teamwork in medicine and aviation. *British Medical Journal*. 320, pp. 745-749.

[255.] Seymour, N. et al. 2002. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg.* 236(4), pp. 458-463.

[256.] Shah, J. et al. 2001. Simulation in urology- a role for virtual reality? *BJU Int.* 88(7), pp. 661-665.

[257.] Shah, S. et al. 2008. Development of an ultrasound training curriculum in a limited resource international setting: successes and challenges of ultrasound training in rural Rwanda. *Int J Emerg Med*. 1(3), pp. 193-196.

[258.] Shain, R. et al. 1982. Evaluation of the gynaecology-teaching associate versus pelvic model approach to teaching pelvic examination. *Medical Education* 57, pp. 646-648.

[259.] Shanmugan, S. et al. 2014. Virtual reality simulator training for laparoscopic colectomy: what metrics have construct validity? *Dis Colon Rectum.* 57(2), pp. 210-214.

[260.] Sheehan, F. et al. 2013. Echo simulator with novel training and competency testing tools. *Stud Health Technol Inform.* 184, pp. 97-403.

[261.] Shetty, S. et al. 2012. Construct and face validity of a virtual reality-based camera navigation curriculum. *Journal of Surgical Research*. 177(2), pp.191-5.

[262.] Shumard, K. et al. 2015. Development of a novel task trainer for second trimester ultrasound-guided uterine evacuation. *Simul Healthc* 10(1), pp. 49-53.

[263.] Sidhu, H. et al. 2012. Role of simulation-based education in ultrasound practice training. *J Ultrasound Med.* 31(5), pp. 785-791.

[264.] Sites, B. et al. 2004. The learning curve associated with a simulated ultrasound-guided interventional task by inexperienced anesthesia residents. *Reg Anesth Pain Med.* 29(6), pp. 544-548

[265.] Slagle, J. et al. 2002. Assessment of the intrarater and interrater reliability of an established clinical task analysis methodology. *Anesthesiology* 96(5), pp. 1129-1139.

[266.] Stather, D. et al. 2011. Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training. *Respirology.* 16(4), pp. 698-704.

[267.] Stefanidis, D. et al. 2007. Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training?. *J Am Coll Surg.* 205(2), pp. 307-313.

[268.] Stefanidis, D. et al. 2012. Simulator Training to Automaticity Leads to Improved Skill Transfer Compared with Traditional Proficiency- Based Training, A Randomised Controlled Trial. *Annals of Surgery.* 255(1), pp. 30-37.

[269.] Stefanidis, D. et al. 2009. Do metrics matter? Time versus motion tracking for performance assessment of proficiency-based laparoscopic skills training. *Simul Healthcare.* 4(2), pp. 104-108.

[270.] Sturm, L. et al. 2008. A systematic review of skills transfer after surgical simulation training. *Ann Surg.* 248(2), pp. 166-179.

[271.] Sweet, R. et al. 2004. Face, Content and Construct Validity of the University of Washington Virtual Reality Transurethral Prostate Resection Trainer. *The Journal of Urology*. 172(5), pp. 1953-1957.

[272.] Tanaka, A. et al. 2015. Robotic surgery simulation validity and usability comparative analysis. *Surgical Endoscopy*. pp. 1-10.

[273.] Tarara, D. et al. 2014. Real-Time test-retest and interrater reliability of select physical performance measures in physical active college-aged students. *Int J Sports Phys Ther.* 9(7), pp. 874-887.

[274.] Terkamp, C. et al. 2003. Simulation of abdomen sonography. Evaluation of a new ultrasound simulator. *Ultraschall Med.* 24(4), pp. 239-234.

[275.] Tetzlaff, J. 2007. Assessment of competency in anesthesiology. *Anesthesiology.* 106(4), pp. 812-825.

[276.] Tian, Y. et al. 2015. Evaluation of simulation-based training for aircraft carrier marshalling with learning cubic and Kirkpatrick's models. *Chinese Journal of Aeronautics*. 28(1), pp. 152-163

[277.] Tolsgaard, M. et al. 2012. Construct validity of the reporter-interpreter-manager-educator structure for assessing students' patient encounter skills. *Acad Med.* 87(6), pp. 799-806.

[278.] Tolsgaard, M. et al. 2013. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. *PLoS One.* 8(2), pp. e57687.

[279.] Tolsgaard, M. et al. 2014a. Which factors are associated with trainees'confidence in performing obstetrics and gynaecological ultrasound examination. *Ultrasound Obstet Gynecol.* 43(4), pp. 444-451.

[280.] Tolsgaard, M. et al. 2014b. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. *Ultrasound Obstet Gynecol.* 43(4), pp. 437-443.

[281.] Tolsgaard, M. et al. 2015a. Sustained effect of simulation-based ultrasound training on clinical performance: a randomized trial. *Ultrasound Obstet Gynecol.* 46(3), pp. 312-318.

[282.] Tolsgaard, M. et al. 2015b. The effect of dyad versus individual simulation-based ultrasound training on skills transfer. *Med Educ.* 49(3), pp. 286-295.

[283.] Tuttle, R. et al. 2007. Utilizing Simulation Technology for Competency Skills Assessment and a Comparison of Traditional Methods of Training to Simulation-Based Training. *Respiratory care.* 52(3), pp. 263-270.

[284.] van der Spek, E. et al. 2013. Introducing surprising events can stimulate deep learning in a serious game. *British Journal of Educational Technology.* 44(1), pp. 156–169.

[285.] van der Vleuten, C. and Schuwirth, L. 2005. Assessing professional competence: from methods to programmes. *Med Educ.* 39(3), pp. 309-317.

[286.] van Dongen, K. et al. 2007. Construct validity of the LapSim: can the LapSim virtual reality simulator distinguish between novices and experts? *Surg Endosc.* 21(8), pp. 1413-1417.

[287.] Van Nijlen, D. and Janssens, R. 2008. Modeling judgments in the angoff and contrasting-groups method of standard setting. *Journal of Educational Measurement.* 45(1), pp. 45-63.

[288.] Van Sickle, K. et al. 2011. A multicenter, simulation-based skills training collaborative using shared GI Mentor II systems: results from the Texas Association of Surgical Skills Laboratories (TASSL) flexible endoscopy curriculum. *Surg Endosc.* 29(9), pp. 2980-2986.

[289.] Verdaasdonk, E. et al. 2006. Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. *Surg Endosc.* 20(3), pp. 511-518.

[290.] Vick, L. et al. 2007. Face, content, and construct validities of inanimate intestinal anastomoses simulation. *J Surg Educ.* 64(6), pp. 365-368

[291.] Vnuk, A. et al. 2006. Assessing proficiency in adult basic life support: student and expert assessment and the impact of video recording. *Med Teach.* 28(5), pp. 429-434.

[292.] Walker, S. et al. 2013. Unannounced in situ simulations: integrating training and clinical practice. *BMJ Qual Saf.* 22(6), pp. 453-458.

[293.] Wang, E. et al. 2008. Developing technical expertise in emergency medicine: the role of simulation in procedural skill acquisition. *Acad Emerg Med.* 15(11), pp.1046-1057.

[294.] Ward, S. et al. 2014. An analysis of the learning curve to achieve competency at colonoscopy using the JETS database. *Gut.* 0, pp. 1-9.

[295.] Weidenbach, M. et al. 2009. Simulation of congenital heart defects: a novel way of training in echocardiography. *Heart.* 95(8), pp. 636-641.

[296.] Weller, R. and Zachmann, G. 2012. *User performance in complex bi-manual haptic manipulation with 3 DOFs vs. 6 DOFs*. IEEE Publisher. pp. 315-322

[297.] White, M. et al. 2010. Validation of a high fidelity adult ureteroscopy and renoscopy simulator. *J Urol.* 183(2), pp. 673-677.

[298.] Wijnberger, L. et al. 2000. Learning in medicine: chorionic villus sampling. *Prenat Diagn.* 2000(20), pp. 241–246.

[299.] Wilfong, D. et al. 2011. The effects of virtual intravenous and patient simulator training compared to the traditional approach of teaching nurses: a research project on peripheral i.v. catheter insertion. *J Infus Nurs.* 34(1), pp. 55-62.

[300.] Willaert, W. et al. 2011. Patient-specific endovascular simulation influences interventionalists performing carotid artery stenting procedures. *Eur J Vasc Endovasc Surg.* 41(4), pp. 492-500.

[301.] Williams, C. et al. 2013. Transvaginal ultrasound simulation and its effect on trainee confidence levels: a replacement for initial clinical training?. *Ultrasound.* 21(2), pp. 50-56.

[302.] Winkel, A. et al. 2016. Assessment of Developmental Progress Using an Objective Structured Clinical Examination-Simulation Hybrid Examination for Obstetrics and Gynecology Residents. *J Surg Educ.* 73(2), pp. 230-237.

[303.] Woodworth, G. et al. 2014. Efficacy of computer-based video and simulation in ultrasound-guided regional anesthesia training. *J Clin Anesth.* 26(3), pp. 212-221.

[304.] Wright, M. et al. 2013. Standardized assessment for evaluation of team skills: validity and feasibility. *Simul Healthcare.* 8(5), pp. 292-303.

[305.] Wulf, G. et al. 2010. Motor skill learning and performance: a review of influential factors. *Medical education.* 44(1), pp. 75-84.

[306.] Xiao, D. et al. 2014. Face, Content, and Construct Validity of a Novel Portable Ergonomic Simulator for Basic Laparoscopic Skills. *Journal of Surgical Education.* 71(1), pp. 65–72.

[307.] Xue, H. et al. 2010. Three-dimensional echocardiographic virtual endoscopy for the diagnosis of congenital heart disease in children. *Int J Cardiovasc Imaging.* 26(8), pp. 851-859.

[308.] Yanes, A. et al. 2016. Observation for assessment of clinician performance: a narrative review. *BMJ Qual Saf.* 25(2), pp. 46-55.

[309.] York, S. et al. 2016. Implementation and evaluation of a dilation and evacuation simulation training curriculum. *Contraception.* S0010-7824(15), pp. 30070-30076.

[310.] Zieky, M. and Perie, M. 2004. *A Primer on Setting Cut Scores on Tests of Educational Achievement: Excerpts From Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*.

Appendices

Appendices

Appendix : Summary of Ph.D research projects

Appendix I: Local research ethical approval (SEWREC)

Appendix II: Consent form

Appendix III: participation information form

Appendix IV: Randomised controlled trial protocol registration (ISRCTN03408765)

Appendix (3.1): Face and content validity – questionnaire

Appendix (4.1): Checklist and global rating scale
Appendix (4.2): Criteria based assessment of ultrasound skills for global rating scale

Appendix (5.1): Skill assessment checklist for simulation metrics
Appendix (5.2): Skill assessment checklist for randomised controlled trial

Appendix (7.1): Experts' survey
Appendix (7.2): Initial online for participation
Appendix (7.3): Repeated measures 'intention-to-train and subgroups' protocol analysis
Appendix (7.4): Kaplan Meier cumulative hazard estimate: supplemental results

Appendix (8.1): End of Trial survey
Appendix (8.2 and 8.3): Ultrasound simulation training and its impact on clinical skills development and barriers to access: Pilot surveys

Appendices (9.1 and 9.2): Supplemental SSC projects (2009-2010)

Summary of Ph.D projects

Summary of Ph.D Research projects

Chapter	Three
Title	Face and Content Validity of the Virtual Reality Simulator: ScanTrainer®
Aim	To determine face and content validity of TVUS ScanTrainer®
Objectives	(1) to recruit practitioners with varying levels of ultrasound experience from attendees of an international conference, and (2) for study volunteers to undertake relevant simulator tutorials and complete a structured questionnaire including statements on face and content validity.
Hypothesis	Is that the simulator is (1) realistic for the purpose of developing ultrasound skills and reflects real life scanning, and (2) the content of its structured learning approach represents the knowledge and psychomotor skills that must be learnt when scanning patients.
Research question	Is the simulator able to simulate TVUS as in real practice?
Subjects	(n=36), Non-expert (n=25) and expert (n=11)
Method	-The structured questionnaire consists of two sections; one detailed subjects' demographic information, previous ultrasound experience and any previous experience with VR simulation or ultrasound mannequins. The other section included 14 simulation-related statements. -Fourteen simulation-related statements/parameters are subjectively scored along a 10 cm visual analogue scale (VAS) line by marking the point that subjects felt most appropriate, with (0) at one end (very bad) and (10) at the other (very good). -Statements 1 to 6 assessed face validity, 7 to 12 evaluated the simulator's learning content and 13 and 14 were general statements on the value of the simulator as training and testing tool. -Ratings on the scale (10 cm which was equalised to 100 mm) were defined in "mm" as; 0-9 (very strongly disagree), 10-19 (strongly disagree), 20-29 (disagree), 30-39 (moderately disagree), 40-49 (mildly disagree), 50 (undecided), 51-59 (mildly agree), 60-69 (moderately agree), 70-79 (agree), 80-89 (strongly agree), 90-100 (very strongly agree).
Statistical analysis	1- Data not normally distributed as tested by Shapiro-Wilk, $p<0.001$ 2- Sample size (n=36), based on published studies of face and content validity 3- Median scores and box plots are constructed for each statement as rated by non-experts and experts. 4- Face validity and general statements data were stratified by expert and non-expert status. 5- Content validity data were reported for experts only. 6- Differences between experts and non-experts ratings analysed using the Mann-Whitney U test where the significance indicates p-value ≤ 0.05.
Result	**Face validity:** 1- Median scores of experts and non-experts' ratings ranged between 7.5 and 9.0 and were slightly higher by experts in two statements 2- Two statements were rated lower by experts 3- The remaining two statements were equally rated 4- No statistically significant differences between the two groups' ratings in all statements **Content validity:** 5- Experts' median scores ranged from 8.4 to 9.0 6- Median values and box-plots of the 14 statements (figures 3.1-3.3)
Limitation	A potential limitation of the study is that it did not determine in advance the sample size required to obtain a reliable result for face and content validation. In a study validating robotic simulator performance, a sample size of six participants in each of the expert and pure novice groups was deemed adequate to achieve significance at 80% statistical power on the basis of available literature data
Conclusion	This study confirms that ScanTrainer® simulator has the feel and look (face validity) and tutorial structure (content validity) to be realistic and relevant for actual TVUS scanning. This study also concurs with the notion that advancing computer technologies have been able to incorporate virtual reality into training to facilitate the practice of basic skills as well as complex procedures that leave little room for error or mistake.

Summary of Ph.D Research projects

Chapter	Four
Title	Intra- and inter-observer reliability of scoring systems for ultrasound skills assessment
Aim	Is to test the reliability of scoring systems developed for the assessment of obstetrics and gynaecology ultrasound skills.
Objectives	Are to use scored video-recordings of ultrasound scans to (1) determine intra-observer (test and re-test) absolute agreement of the scoring systems for each independent observer individually, (2) determine inter-observer reliability between two independent observers' ratings to evaluate the consistency of two scorings and (3) test the level of agreement between the checklists and GRS scores of the two observers.
Sample material	Video recordings of participants in a RCT
Method	- Observational experiment, partial (test and re-test) research design. - Two independent observers - Pilot study (n=10), test-re-test method - Large-scale study (n=144), Inter-observer reliability assessed the level of agreement and reliability between the two observers based on their independently scored video recordings using the checklist and GRS. - Checklist: seven skills in which there are three scorings: pass (1), fail (0) or (N/A). - GRS: eight performances, in which the assessment scale ranged from "not attempted" (NA=0), 1 (very poor) to 5 (excellent).
Statistical analysis	Shapiro-Wilk test was used to test normality of data distribution of the checklist and the GRS. **Pilot study (n=10):** (1) Kruskal-Wallis (for the checklists) and one-way ANOVA (for the GRS) were used for a test-retest statistical analysis with significance indicates at 0.05 of the five repetitions of each observer's scores. (2) Test-retest absolute agreement of the five repetitions by each observer was tested with the intra-class correlation coefficient (ICC), and (3) The box-plots represent median values of five attempts of checklist and GRS for each observer independently. **Large-scale study (n=144):** (1) Inter-observer agreement was tested with the percentage of absolute agreement, intra-class correlation coefficient (ICC) and (2) inter-observer reliability was tested by Cohen's Kappa (Ķ), (3) The Bland–Altman plots represent the difference of mean values between two observers scores with checklist and GRS individually, (4) Spearman correlation coefficient used to measure the degree of linear relationship "correlation" in checklist scores obtained by the two observers. (5) Pearson correlation coefficient r used to measure the degree of linear relationship "correlation" in GRS scores obtained by the two observers.
Result	**Pilot study:** 1- The test and re-test Kruskal-Wallis findings indicate no statistically significant difference in the five repetitions outcome for each observer, p-value >0.05 2- Intra-observer agreement (ICC) scores for A.A. and D.A. were 0.80 and 0.72 for checklists and 0.80 and 0.71 for GRS respectively. 3- The absolute agreement of ICC scores obtained by A.A. with checklists and GRS revealed an excellent agreement while good agreement was shown by D.A **Large-scale study:** 1- The checklist datasets for the three assignments were not normally distributed as tested by Shapiro-Wilk, p=0.001, while GRS datasets were normally distributed, p=0.5. 2- To estimate variance components in the selected sample, generalisability coefficient for variables was ranged between 0-1 Table (4.7) shows an excellent inter-rater observer absolute agreement for checklist and GRS (ICC=0.96 and 0.97 respectively) with no statistically significant difference between the two observers' checklist and GRS scores. 3- Table (4.8) shows good to excellent inter-observer reliability for checklist (Cohen's kappa Ķ); 0.83, 0.78 and 0.92 for GYN1, GYN2 and early pregnancy assignments respectively. However, the GRS results show moderate inter-observer reliability GYN1 and GYN2, with Ķ values 0.56 and 0.60 respectively, and good reliability in the early pregnancy assignment, Ķ =0.69. 4- The ICC are likely to be correlated more positively with the checklist scoring system than with GRS scores.
Limitations	1- the use of video recordings that either revealed or did not reveal the anatomy of the female pelvis. Those that did so enabled the observer to monitor the movement of the probe during the examination easily. 2-No clear instructions given to the trainee at the onset or lack or audio to the video recording, so the trainee explains her/his actions, which impact on the accuracy of rating between the two observers' results.
Conclusion	The provision of an ongoing quality assurance platform through one trustworthy tool, make the use of video recordings of significant value to medical education. This study demonstrated high agreement and reliability between two observers, and thus one (A.A) was considered as a standard human judge against simulation automated feedback in subsequent studies.

Summary of Ph.D Research projects

Chapter	Five
Title	Validation of simulator metrics
Aim	To ensure that the two metrics designs (IS and FE) in the ultrasound simulator are consistent in providing identical feedback on TVUS performance by the same subject as well as feedback that is consistent with that given by a human judge. This should determines the reliability of simulation-based assessment and its suitability for reporting the actual performance of trainees and for reflecting their gradual change in TVUS practical level during the six assessments sessions in the RCT.
Objectives	(1) to determine the reliability between the two metrics designs in the simulator: individual skill task "IS" and full examination task "FE", in providing consistent and identical feedback on TVUS performance of participants; (2) to determine the level of 'absolute' agreement between simulator metrics (IS and FE) as compared with human observer due to select the appropriate simulator metric and use it for assessing participants' TVUS performance in randomised controlled trials and finally (3) to determine absolute agreement between the simulator metric (FE) and the observer during the six assessments in the randomised controlled trial.
Subjects	**Reliability of simulation metrics: IS & FE tasks:** Sample size (n=11): expert (n=6), novice (n=5) **Agreement between simulator metric (FE) tasks and the observer:** Novices enrolled in RCT of sample of (n=63): intervention (n=34) and control (n=29), and assessed six times, to give a sample of (n=1134)
Method	**Reliability of simulation metrics: IS & FE tasks (Objective one)** 1-Self-selection method 2-The participants performed nine assignments as instructed by the PhD researcher who scored and assessed the performance using the pass/fail checklist. 3-The checklist consists of a number of skills: six from the gynaecology module and seven from the early pregnancy module. Each of these skills scores (1) for pass and (0) for a fail or not attempt (NA). 4-Each checklist sheet included a section for evaluation obtained by the observer, and another section for simulation feedback **Agreement between simulator metric (FE) tasks and the observer (Objective two)** 1-the six assessment tests in RCT are included 2-checklist consists of seven skills of each assignment in which there are three scorings: pass (1), fail (0) or not attempted (N/A). 3-Each checklist sheet included a section for evaluation obtained by the observer, and another section for simulation feedback
Statistical analysis	-Datasets are not normally distributed as tested by Shapiro-Wilk, $p<0.05$ **-Reliability of simulation metrics: IS & FE tasks (Objective one)** 1-Median values calculated for: novices and experts performances rated by simulator metrics (IS and FE) tasks and the observer. 2-The statistical significance between two groups, and between IS, FE and the observer as tested by Mann-Whitney U test. 3-Reliability and agreement were determined by using an intra-class coefficient (ICC). **Agreement between simulator metric (FE) tasks and the observer (Objective two)** 1-ICC used to estimate level of agreement between the simulator's and the observer. 2-the statistical significance between two scorings (the simulator and the observer) is considered at $p<0.05$. 3-Box-plots represent the median values as scored by the simulator metric (FE) and the observer at each point of the six assessment sessions in RCT. 4-Bland-Altman plots illustrate differences in scores between simulator metric (FE) and the observer' ratings 5- Spearman correlation coefficient used to measure the degree of linear relationship "correlation" between FE and the observer's scores.
Result	**Reliability of simulation metrics: IS & FE tasks: (Objective one)** 1-Experts scored higher than novices in TVUS performance on the simulator metrics (IS and FE), but no statistical significant differences indicated between experts and novices with IS and FE tasks. 2-ICC revealed high reliability between (IS) and (FE) tasks: ICC values are 0.96 and 0.89 for GYN and early pregnancy assignments. 3-absolute agreement between the observer's scoring and two tasks (IS) and (FE) are high: ICC values were 0.87 and 0.96 for the early pregnancy module and 0.92 and 0.77 for the gynaecology module for (IS) and (FE) tasks respectively. 4- no statistical significant differences indicated between IS, FE and the observer. **Agreement between simulator metric (FE) tasks and the observer: (Objective two)** 1-ICC values are 0.96, 0.83, and 0.86 for GYN1, GYN2 and early pregnancy assignments respectively. 2-positive correlation indicates (r=0.9, 0.8 and 0.8) for GYN1, GYN2 and early pregnancy assignments respectively
Limitations	(1) metrics-based assessment sometimes leads to flawed conclusions of correct or optimal practice, as the metric has inflexible and limited ability in comparison to human assessors. (2) small sample size of 11 subjects,(3)self-selection method used to invite participants is widely accepted, this may reflect inherent bias as participation inevitably required participants' interest in taking part.
Conclusion	The study's findings demonstrate high reliability in assessing simulated (TVUS) performance and provide accurate feedback on trainees' performance obtained by simulation metrics. This shows that simulated metrics-based performance is able to provide accurate evaluation of trainees' level of practical knowledge in gynaecology and early pregnancy ultrasound. Simulation metrics-based assessment utilises the TVUS learning approach to feedback and can provide an effective means of assisting trainees to improve their performance.

Summary of Ph.D Research projects

Chapter	Six
Title	Validation of subjects' performance on the Simulator: Construct validity
Aim	To assess whether the ScanTrainer® ultrasound simulator can discriminate between novice, intermediate and experienced-level practitioners in performing transvaginal ultrasound skills (TVUS).
Objectives	(1) to test the significance of scores achieved by the three groups in performing three simulation assignments; (2) to determine the significance different between two scores obtained by the simulator and human observer in assessing subjects' performances; (3) to standardise the performance of contrasting-groups 'pass/fail scoring' method in order to measure levels of competence of TVUS practice.
Subjects	Comparative study, includes three groups categorised as experts (n=8), intermediates (n=10) and novices (n=12)
Method	Checklist used for assessment based on pass/fail scores. Seven skills were assessed in the checklist of each assignment in which there are two scorings: (1) if pass: when the skill correctly performed, and (0) if fail: when the skill incorrectly performed or not attempted (N/A). The contrasting-groups method was used to determine cut-off (borderline) score of experts and novices subjects.
Statistical analysis	(1) Data is not normally distributed when tested by Shapiro-Wilk, $p<0.05$. (2) The median values of TVUS performances calculated for comparison between the three study groups' (experts, intermediate and novices) as rated by the simulator and the observer. (3) Non-parametric tests used to test the statistical significance of individual skills in the three assignments for the three study groups, with $p<0.05$, (Mann-Whitney U test between two groups and Kruskal-Wallis between three groups). (4) The box-plots represent median scores obtained by three study groups as rated by the simulator and the observer. (5) For contrasting group method, graph distribution of scores for experts and novices, where the intersection represents the appropriate cut-score (borderline pass score).
Result	(1) median scores as rated by the observer and the simulator reveal that experts scored higher than intermediates, and intermediates scored highly than novices. (2)There are statistical significant differences of individual skills between the three groups, however, the statistical significant difference indicated between novices and skilled subjects (experts and intermediates) but not between experts and intermediates. (3)The box-plot results for novices show a discrepancy between the simulator's and the observer's ratings and the significant different between the two observers indicated in rating novices performances. (4)For contrasting groups method, The scores obtained by the simulator were higher than the subjective scores in determining pass/fail level in the three assignments. The graph distributions of median scores reveal an intersection resulting from experts and novices scores and the cut-off represents scores of 4 and 5 for gynaecology and early pregnancy assignment respectively and score of 4 for overall competency.
Limitations	(1) lack of significant difference between intermediate and experts on all skills included in the assessment (2) small sample size
Conclusion	The study demonstrates that the ultrasound simulator is able to distinguish between subjects with different levels of transvaginal ultrasound (TVUS) experience. The study findings support existing theories of simulation construct validity and the development of competency by establishing performance standards in obstetrics and gynaecology ultrasound.

Summary of Ph.D Research projects

Chapter	Seven
Title	Assessment of learning curves: Randomised controlled trial
Aim	To determine the length of time required for trainees to acquire the skills necessary to perform transvaginal ultrasound (TVUS).
Objectives	(1) to determine the trainees' speed of acquisition of ultrasound skills with whether simulation-supported or conventional training; (2) to explore the potential factors that influence learning curves for two study groups, and (3) to explore the factors associated with each point on the learning curve; for example number of training sessions received, engagement to simulation training.
Hypothesis	There is a difference between the simulation training supplemental to conventional training and clinical training alone.
Research question	Does simulation-supported training expedite the acquisition of skills and competency? The "Null" hypothesis being that there is no difference between the two training approaches.
Subjects	RCT. Sample size of 63: intervention (n=34) and control (n=29)
Method	Checklist used for assessment and based on pass/fail scoring. Seven skills are assessed in the checklist of each assignment and in which there are two scorings: (1) if pass: when the skill correctly performed, and (0) if fail: when the skill incorrectly performed or not attempted (N/A). The six assessments arranged for each subjects.
Statistical analysis	(1)The Shapiro-Wilk test revealed that the data distribution was not normal, p-value <0.05 (2) The descriptive data (median scores) were collected from: (1) the experts' survey, (2) the logbook and (3) trainees' accounts in the simulator. (3) to estimate the time required to successfully pass the assignments, the median values for the two groups in the randomised controlled trial (RCT) were calculated according to: (i) **Repeated measures (RM):** in order to repeatedly measure the gradual changes and significance of trainees' performances at different points, from the baseline to the final assessment (sixth test), the repeated measures analysis was used to (1) test the statistically significant difference between subjects within control and intervention, as well as between the two groups as a whole, at each test, for each assignment accordingly. Also intention-to-train subgroup protocol analysis was considered. (2) Average score of three modules used for analyse the overall skill performance. Non-parametric tests: Mann-Whitney U and Kruskal-Wallis tests, were used to test the significance of gradual changes in performance/learning curves for subgroups treated as 'intention-to-train', p<0.05. (ii) **Kaplan-Meier (KM)**: non-parametric cumulative hazard to estimate the probability of reaching the highest possible score of seven, according to total skills in the checklist for the three assignments independently, also KM estimated according to attainments of borderline pass score, and statistically tested the null hypothesis using the log-rank test, which the significance level at p<0.05. in order to estimate overall skill performance for a combined three modules, time variable was analysed according to time required (1) to successfully pass (any) one module, (2) to pass three modules together, (3) to pass maximum score of (any) one module, and (4) to pass maximum score of three modules.
Result	**Repeated measure:** There are statistically significant differences among subjects for each trend, however no statistically significant between overall trends of intervention and control at the final plateau in the three assignments. For intention- training, the findings demonstrate significant differences between subjects in the control (with no clinical training, n=20) and other subjects in the control group (those received clinical training) and intervention (those received clinical and/or simulation training, at the final plateau in GYN2 and early pregnancy assignments. **Kaplan- Meier estimate:** Results of overall competency KM curves indicate a significant different between the intervention and control as assessed whether by number of tests or borderline score. However, there is no statistical significant different between two groups in the three assignments when tested separately by number of tests, result of testing the two performances with borderline scores is indicated in GYN1 and early pregnancy assignments only, p<0.05. With regard to subgroups, subjects in control group who didn't receive clinical training are statistically significant with other groups whether received clinical and/ or simulation training during the trial. Nevertheless, no statistical significant difference between those received simulation and clinical training and those received simulation training alone, the latter shows speed in skill acquisition and short learning curve.
Limitation	(1) even though this current RCT considered the largest ultrasound simulation training study to date, we acknowledge that subgroup is still a relatively small number. (2)this study conducted in un-controlled learning environment.
Conclusion	This study set out with the aim of assessing the length of time required for novice trainees to learn basic transvaginal ultrasound skills TVUS during their conventional training supported with simulation training. The findings suggest that those engaged in simulation training supplementary to their clinical training gained skill acquisition faster than those who received clinical training alone. Nevertheless, outcome of two groups were similar at the final plateau of the assessment for the three assignments.

Summary of Ph.D Research projects

Chapter	Eight
Title	Participants' perceptions of the effectiveness of simulation practice: End trial survey
Aim	To explore trainees' perceptions of simulation training as supplemental to their clinical training. The objectives of this end-of-trial survey were to investigate current ultrasound training delivered to obstetrics and gynaecology trainees
Objectives	To determine (1) the benefits and limitations of simulation training compared to another model i.e. mannequin Blue Phantom™, (2) the barriers and obstacles that have resulted in a gap in learning transvaginal ultrasound and (3) to clarify the potential solutions that help in enhancing current ultrasound training.
Subjects	The sample included subjects participated in RCT.
Method	Online survey which contents of statements that help in exploring trainees' opinions about current barriers and potential ways of overcoming those barriers while learning TVUS in obstetrics and gynaecology
Statistical analysis	The Shapiro-Wilk test revealed that the data distribution was not normal, p-value <0.05 The questions in the survey consist of categorical data that are analysed as descriptive data.
Result	A total of seventy-seven invitations were sent but only forty-four responses were received giving a response rate of 57%. Frequency of access to the simulation training centres during the trial and the obstacles faced by trainees when practising with the simulator are found in Table 8.4. Regular assessment received by trainees during the trial and the accompanying feedback provided about the TVUS performance is described in Table 8.5. In terms of assessing the quality of service provided by the ScanTrainer® ultrasound simulator, thirteen statements about the simulator were assessed and median scores were given to each statement, as shown in Table 8.6. A list of the differences between the simulator and other TVUS learning models and also a list of the reasons for engagement are detailed in Table 8.7. A list of the benefits and limitations in learning TVUS in gynaecology and early pregnancy are outlined in Table 8.9. Trainees' perceptions on possible solutions that would enhance their experience with simulation training are found in Table 8.10.
Conclusion	The findings of this study draw attention to important issues related to current ultrasound training in obstetrics and gynaecology as delivered by Welsh Deanery of obstetrics and gynaecology, by seeking trainees' perceptions of attitudes towards simulation training supplemental to clinical training during their participation in a randomised controlled trial (RCT). The response rate for the survey was high and the majority of respondents gave positive feedback about the simulation training in learning transvaginal ultrasound (TVUS).

Appendices I, II, III, IV

GIG CYMRU NHS WALES
Canolfan Gwasanaethau Busnes
Business Services Centre

South East Wales Research Ethics Committee Panel B
Telephone: 02920 376823
Facsimile: 02920 376835
Email: Carl.phillips@wales.nhs.uk

Mr Nazar Amso
Senior Lecturer &
Consultant Obstetrician & Gynaecologist
School of Medicine, Cardiff University
University Hospital of Wales
Heath Park, Cardiff
CF14 4XN

14 December 2010

Dear Mr Amso

Study Title:	**Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach**
REC reference number:	**10/WSE02/75**
Protocol number:	**SPON 896-10**

Thank you for your letter of the 13 December 2010, responding to the Committee's request for further information on the above research, and for submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation [as revised], subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

- Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

- Management permission ("R&D approval") should be sought from all NHS organisation(s) involved in the study in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System (IRAS) or at http://www.rdforum.nhs.uk.

- Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

- For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

- Sponsors are not required to notify the Committee of approvals from host organisations.

- It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Response to Request for Further Information	A Alsalamah	13 December 2010
Participant Information Sheet: Appendix VIII - Information for Participants	3.1	17 November 2010
Questionnaire: Global Rating Evaluation of Vide-Record Performance on Ultrasound Simulator	3.1	17 November 2010
Questionnaire: Ultrasound Assessment for normal female pelvis	3.1	17 November 2010
Questionnaire: Skill Assessment Checklist (Ultrasound Simulator)	3.1	17 November 2010
Questionnaire: Appendix II - Questionnaire for Experts	3.1	17 November 2010
Letter from Sponsor	Cardiff University	15 October 2015
Questionnaire: End of Trial Questionnaire for participants	3.1	17 November 2010
Letter of invitation to participant	3.11	17 November 2010
Letter confirming funding of tution fees	G Almakky	01 July 2009
Protocol	3.1	17 November 2010
Participant Information Sheet	3.2	13 December 2010
Participant Consent Form: Appendix VII	3.1	17 November 2010
Investigator CV	A Alsalamah	11 November 2010
Investigator CV	N A Amso	29 October 2010
REC application	IRAS 3.1	10 November 2010

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

10/WSE02/75	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Mrs A Dowden
Chair, Panel B,
South East Wales Research Ethics Committees

Enclosures: "After ethical review – guidance for researchers" SL- AR2

Copy to: Miss Amal Alsalamah, 197 Landmark Place, Churchill Way, Cardiff, **CF10 2HU**

R & D Office, Cardiff University

R&D Office for Cardiff & Vale University Health Board

Participant consent form

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach

Name of the researcher: Amal Alsalamah
Supervised by: Mr Nazar Amso & Dr Neil Pugh

Please tick the boxes below

1. I confirm that I have read and understood the information sheet dated **10 February 2011** (version 3.3) for the above study and have had the opportunity to ask questions ☐

2. I understand that my participation in this project will involve "assessing my skills in ultrasound scanning", and "completing the questionnaire" might require 10 minutes of my time ☐

3. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

4. I understand that the information provided by me will be held totally anonymously, so that it is impossible to trace this information back to me individually. I understand that this information may be retained indefinitely. ☐

5. I also understand that at the end of the study I will be provided with additional information and feedback about the purpose of the study. ☐

I, _____(*Participant name*) consent to participate in the study conducted by ***Amal Alsalamah*** School of Medicine, Cardiff University with the supervision of ***Mr Nazar Amso and Dr Neil Pugh***

Signed: _____ **Witness Signature:** _____

Date: _____ **Name and date:** _____

Information for Participants

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training or a conventional training approach.

Who is running the trial?
This project is being conducted by a research student called *Amal Alsalamah* as part of her Ph.D thesis. It is being run through Cardiff University School of Medicine and supervised by the Department of Obstetrics and Gynaecology. You are invited to take part in this research. Before you make your decision, if you wish to participate, it is important to be aware of the purpose of this research.

Please take time to read the following information carefully.

What is the project about?
This project is trying to find a way to help trainees to improve their skills in ultrasound. We aim to determine the trainees' ultrasound learning curves skills achieved by either simulation enhanced training or conventional training. The *ScanTrainer* or *Ultrasound Simulator* was designed by ultrasound experts to enable the trainees to acquire the necessary skills outlined in various education curricula such as the RCOG and EFSUMB level 1–2 skills through a series of tasks, assessments and feedback and was validated as a training and assessment tool by a group of experts prior to its launched..

Why have I been invited to participate?
You are invited to undertake a volunteering part in this study as you are training in ultrasound. As you work in the field of gynaecology where ultrasound is used, we assume it is one of your interests; so your input will help us to document and evaluate the training feedback after you have trained either with or without the *Ultrasound Simulator*.

Am I obliged to take part?
Your participation in the study is purely voluntary and you may withdraw at any time without giving reasons for your withdrawal.

What will happen to me and what would I be asked to do if I take part?
You will be invited to attend the induction programme which clarifies the goals of the study. It is a one-day workshop that focuses on introducing the basic knowledge of ultrasound with hands-on sessions. If you are on the simulator arm, you will have a formal simulator teaching. If you are in conventional arm, no further simulator training you will have.

You will receive an information sheet that clarifies the aims of the study and the distribution of the groups as you will be randomly allocated to one of these groups. Before you agree to take part in the

study, ensure that all your enquiries will be answered. You will also be asked to sign a written consent form to confirm your agreement. After your agreement, you will be welcome to become involved in an induction programme for a day to clarify the sequence and protocol of the study. Then we will talk about times that are convenient for you to practice on the simulator; if you randomly selected an intervention group, participate so you will get to choose the dates and times that are convenient for you.

The research is divided into three main parts. The **first part** is assessing the "baseline skill" at the onset of the study for each participant with the *Ultrasound Simulator*. The **second part** includes an induction day with practical sessions, and the **third part** includes your practical training in your hospital (conventional training) or simulator training initially followed by conventional practical training in your hospital (intervention group) as well as assessment of your skills by the "centre trainer" every two weeks after you are randomly allocated into one of the study groups and assessment of your skills by the researcher using the *Ultrasound Simulator* every four weeks. Eventually, the researcher will look at the gradual changes in the learning curves which will build up a picture of how effective the simulation enhanced training is in improving the competency of the trainee's skills.

At the end of the study, we will distribute questionnaire sheets and ask you to answer the questions and return them. Questions are related to ultrasound skills, self-evaluation and individual expectation of the *Ultrasound Simulator* as a training tool. We anticipate the time taken to answer these questions as ten minutes maximum. No personal information will be required.

What is the procedure being tested?

Introducing simulation enhanced training into conventional training to determine if there is any difference in the speed of skill acquisition and learning curves.

What are the possible disadvantages and risks of taking part?
This research is risk-free as it is a practical training. There will be no treatment, nor any medication involved. You will also need to follow the simulator's "instructions manual" to avoid any risk of repetitive strain injury. Consequently, we do not expect any associated risk if you take part.

What are the possible benefits of taking part?
After your participation, you will receive a detailed report and feedback about your skills assessment. Therefore, you would probably be made aware of the skills that need further training as well as knowing how your skills have been gradually improved by the simulator, in case you are recruited on an intervention/simulator group. If you randomly selected to be in a control group, you would be able to access ultrasound learning material throughout the study and gain access to lectures to build up your basic knowledge of ultrasound. Moreover, your help through completing the questionnaire will also greatly assist us in collecting the research data. At the end of the trial you will be able to have access to the simulator if you need to do so (see below).

Will my participation give me the option of choosing to be recruited in any study group?
Recruitment is randomised, to avoid bias in selecting the participants in each group.

What will happen to me if I have been selected to be a participant at the intervention/simulator group?
You will receive simulation training for 4–6 weeks at the very beginning of the study. After that, you will be allowed to practise on the *Ultrasound Simulator* alongside your clinical training in the hospital. Also, you will gain online access to the e-learning ultrasound module at any time.

What will happen to me if I have been selected to be a participant in the control group?
You will receive clinical training which is already set by the centre at which you work. Also, you will gain access to the online e-learning ultrasound module. Please do not access the ultrasound simulator during the trial period.

What happens when the research study terminates?

After the study is over, your participation will cease. However, you are welcome to contact the researcher if any enquiries arise. After completion of the study period, you will be able to enhance your skills on the simulator if you feel there is a need to do so. The trial researcher will set up an account for you on the simulator and guide you in the initial phase.

What will happen to the results of the research study?
The results will be compiled and written up as the thesis of a Ph.D degree. It will be submitted for peer reviewed journal publication.

What are the local arrangements for the supervision of the conduct of this research at site?
Participants can contact the local ultrasound trainer in the site where they work. The local trainer should able to provide you with information needed either for training or for any issue related to the research.
Local ultrasound trainers are:

1- Mr Nazar Amso, University Hospital of Wales, Cardiff.
2- Dr Bidyut Kumar, Wrexham Maelor Hospital, Wrexham
3- Dr Marsham Moselhi, Singleton Hospital, Swansea

Who is organising and funding the research?
The research is organised and supervised by the Department of Obstetrics and Gynaecology, Cardiff School of Medicine, University Hospital of Wales. The researcher is funded by a scholarship from her government.

What will happen if I do not want to continue with the study?
We do not expect any individual consequences in case of participants wishing to discontinue their participation in this study. We would ideally wish to have your participation until the end of the study but your decisions will be respectfully accepted.

Who has reviewed the study?
The study has been reviewed by Medical Research ethics committee at Cardiff University.

If you have any queries please do not hesitate to contact us.
Note: You will have a copy of this information sheet and a signed consent form to keep.

Contact for further information
The study is being carried out by:
Miss Amal Alsalamah, Ph.D Student
Obstetrics and Gynaecology Ultrasound
Department of Obstetrics and Gynaecology
School of Medicine, Cardiff University
Mobile: +44 (0) 7964917832
Email**:** alsalamahA@cardiff.ac.uk

Supervised by:
Mr Nazar N Amso, PhD FRCOG FHEA
Work phone: +44 (0) 29 20744448
Email: amsonn@cardiff.ac.uk

Dr Neil Pugh, PhD F.Inst P&C Phys
Work phone: 02920743547
Email: neil.pugh@cardiffandvale.wales.nhs.uk

ISRCTN registry

ISRCTN53915329 DOI 10.1186/ISRCTN53915329

The influence of a virtual simulator on the acquisition of trainees ultrasound skills

CrossMark [#]

Condition category
Not Applicable
Date applied
18/02/2013
Date assigned
22/03/2013
Last edited
22/03/2013

Prospective/Retrospective
Retrospectively registered
Overall trial status
Completed
Recruitment status
No longer recruiting

Plain English Summary

Background and study aims
Simulation training in transvaginal ultrasound (TVUS) leads to skill acquisition and opens up a new era for learning ultrasound skills. The limited opportunities of training offered to trainees lead to insufficient practice in scanning. Several studies have suggested that simulation training provides facilities to trainees to practice and learn freely. In order to improve patient safely and comfort in addition to enhancing trainee confidence, learning ultrasound skills through theory-based, structured-systematic approaches to training on the simulator will shorten the length of training time and enhance the individuals learning. Therefore, the introduction of simulation training will help to build on the trainees sonographic knowledge base and develop important ultrasound skills before they work with real patients. Thus this additional simulation training in a clinical environment would improve trainees competency in obstetrics and gynaecology ultrasound scanning.

Who can participate?
The study subjects are recruited primarily from specialist trainees (ST1- ST7 level) in Obstetrics and Gynaecology in the Welsh Deanery, other NHS staff and students of the MSc programme in Cardiff University who fulfil the inclusion criteria.

What does the study involve?
Subjects are randomly allocated into control and intervention (simulation-supported) learning groups. For the simulation group: trainees are undertaking and completing all tutorials on the simulator and are provided with simulator feedback and continued access during the trial duration. For the non-simulation (control) group: trainees are permitted to accessing the unassessed tutorial/tasks mode

before monthly assessment to familiarise themselves with the simulator. At the baseline phase of trial, the simulator is used to assess baseline skills of each subject to test their eligibility to participate. After that, subjects skills are re-assessed every month for evaluating acquisition during the trial (within 6 months). During the trial, the intervention group receive structured, self-directed simulation training while the control group are allowed to access (unassessed practice) modules on the simulator with no feedback provided. Both groups are permitted to receive conventional training in the normal way.

What are the possible benefits and risks of participating?
Study subjects will receive a detailed report and feedback about their skill assessment. They would be made aware of the skill areas that need further training as well as recognising how their skills have been gradually improved by the simulator (for intervention/simulator group). In the control group, they would be able to access ultrasound-learning material throughout the study and gain access to lectures to build up their basic knowledge of ultrasound. At the end of the trial they would be able to have access to the simulator if required to do so. This research is risk-free as it is a practical training but without involving patients. However, study subjects are advised to observe the general and professional guidelines related to avoidance of repetitive stress and injury during conventional scanning in their clinical settings or while using the simulator.

Where is the study run from?
The study runs from three training hospitals in Wales, UK
University Hospital of Wales, Cardiff
Singleton Hospital, Swansea
Wrexham Maelor Hospital, Wrexham.
Sampling selected from different Welsh Deanery training hospitals across North and South Wales, UK.

When is study starting and how long is it expected to run for?
The study started in May 2011 and will run until the required number of 60 trainees have been recruited and assessed.

Who is funding the study?
The research is organised and supervised by the Department of Obstetrics and Gynaecology, Cardiff School of Medicine, University Hospital of Wales. The study is funded by a scholarship from Saudi Arabia government.

Who is the main contact?
Miss Amal Alsalamah (Researcher), alsalamahA@cf.ac.uk
Professor Nazar Amso, amsoNN@cf.ac.uk

Trial website

[]

Contact information

Type

Scientific

Primary contact

Prof Nazar Amso

ORCID ID

[]

Contact details

University Hospital of Wales
Cardiff University School of Medicine
Obstetrics and Gynaecology department
Cardiff
CF14 4XN
United Kingdom

Additional identifiers

EudraCT number

ClinicalTrials.gov number

Protocol/serial number

SPON896-10

Study information

Scientific title

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach

Acronym

Study hypothesis

Simulation-based learning plays a significant role in the enhancement of the educational process and provides a more efficient learning environment. Recently, new devices have become available to enhance ultrasound training, ranging from physical fetal or gynaecological mannequins to the virtual reality computer-based ultrasound simulator (ScanTrainer, MedaPhor Ltd., Cardiff, UK) developed at Cardiff, UK. This Ultrasound Simulator aims to shorten the length of training time through a series of simulation tutorials encompassing a number of objectives, tasks and assessments with computer-generated individualised student feedback. It is hypothesised that such an approach to training will lead to improved technical performance in the real-life scanning environment.

The studys primary aim is to compare the learning curves for the acquisition of core ultrasound scanning skills among research subjects (trainees) undergoing conventional ultrasound training only with those undergoing conventional training supported by structured simulation training.

Ethics approval

The study has been reviewed and ethically approved by South East Wales Ethics Research Committee

Panel B. Cardiff, UK, REC Reference Number 10AA/SE02/75, 14 December 2010

Study design

Parallel randomised control trial

Primary study design

Interventional

Secondary study design

Randomised controlled trial

Trial setting

Hospitals

Trial type

Other

Patient information sheet

Condition

Improving ultrasound skills and training competency for trainees

Intervention

The intervention group were instructed to use the simulator to practice ultrasound skills in addition to receiving clinical training arranged by their hospital, whereas the control group received conventional training alone.

Number of simulation training is unlimited as the simulator is always free to access and use. However, six skill assessment sessions are booked for each participant. Duration of intervention is 6 months, starts from first day of participation.

Intervention type

Other

Phase

Not Applicable

Drug names

Primary outcome measures

Difference in ultrasound scan performance (OSATS) in obstetrics and gynaecology between

simulation- supported trained and untrained subjects

Secondary outcome measures

1. Determine factors that have affected engagement in the project and acquisition of relevant skills. This will provide information on barriers and limitations of simulation training in a clinical environment.
2. Evaluate the validity of task performance sheet for assessing relevant skills in comparison to the simulators task metrics.
3. Overall subjective ultrasound skills acquisition with/out simulation training
4. Differences in skills scores at each assessment session

Overall trial start date

15/05/2011

Overall trial end date

31/12/2013

Reason abandoned

Eligibility

Participant inclusion criteria

1. Subjects (aged 18- 50+) should have (none or limited) ultrasound experience of any kind
2. None or limited previous access to transvaginal ultrasound experience
3. Motivated to learn transvaginal ultrasound skills
4. Intends to complete the requirements of the Royal College of Obstetricians and Gynaecologists (RCOG) ultrasound training curriculum or a similar structured programme
5. Based within Wales or within a very short distance of travel to Cardiff

Participant type

Patient

Age group

Adult

Gender

Both

Target number of participants

60-80

Participant exclusion criteria

1. Individuals who have already completed any structured ultrasound training programme and certified accordingly
2. Individuals at consultant level in obstetrics and gynaecology or radiology even if they have no ultrasound experience
3. Radiographers on the ultrasound MSc programme with previous experience in transvaginal ultrasound scanning

Recruitment start date

15/05/2011

Recruitment end date

31/12/2013

Locations

Countries of recruitment

United Kingdom

Trial participating centre

University Hospital of Wales
Cardiff
CF14 4XN
United Kingdom

Sponsor information

Organisation

Cardiff University (UK)

Sponsor details

Cardiff University Registry and Academic Services
30 - 36 Newport Road
Cardiff
CF24 0DE
United Kingdom

Sponsor type

University/education

Website

http://www.cf.ac.uk [http://www.cf.ac.uk]

Funders

Funder type

Government

Funder name

Saudi Arabia Government - Scholarship funding, PhD project undertaken at Cardiff University

Alternative name(s)

Funding Body Type

Funding Body Subtype

Location

Results and Publications

Publication and dissemination plan

Not provided at time of registration

Intention to publish date

Participant level data

Not provided at time of registration

Results - basic reporting

Publication summary

Publication citations

Additional files

Editorial Notes

The influence of a virtual simulator on the acquisition of trainee's ultrasound skills	
ISRCTN	ISRCTN03408765
DOI	10.1186/ISRCTN03408765
ClinicalTrials.gov identifier	
EudraCT number	
Public title	The influence of a virtual simulator on the acquisition of trainee's ultrasound skills
Scientific title	Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach
Acronym	N/A
Serial number at source	N/A
Study hypothesis	Simulation-based learning plays a significant role in the enhancement of the educational process and provides a more efficient learning environment. Recently, new devices have become available to enhance ultrasound training, ranging from 'physical' fetal or gynaecological mannequins to the virtual reality computer-based ultrasound simulator (ScanTrainer, MedaPhor Ltd., Cardiff, UK) developed at Cardiff, UK. This Ultrasound Simulator aims to shorten the length of training time through a series of simulation tutorials encompassing a number of objectives, tasks and assessments with computer-generated individualised student feedback. It is hypothesised that such an approach to training will lead to improved technical performance in the real-life scanning environment. The study's *primary aim* is to compare the learning curves for the acquisition of core ultrasound scanning skills among research subjects (trainees) undergoing conventional ultrasound training only with those undergoing conventional training supported by structured simulation training.
Lay summary	**Background and study aims?** Simulation training in transvaginal ultrasound (TVUS) leads to skill acquisition and opens up a new era for learning ultrasound skills. The limited opportunities of training offered to trainees lead to insufficient practice in scanning. Several studies have suggested that simulation training provides facilities to trainees to practice and learn freely. In order to improve patient safely and comfort in addition to enhancing trainee confidence, learning ultrasound skills through theory-based, structured-systematic approaches to training on the simulator will shorten the length of training time and enhance the individual's learning. Therefore, the introduction of simulation training will help to build on the trainee's sonographic knowledge base and develop important ultrasound skills before they work with real patients. Thus this additional simulation training in a clinical environment would improve trainees' competency in obstetrics and gynaecology ultrasound scanning. **What does the study involve?** Subjects are randomly allocated into control and intervention (simulation-supported) learning groups. For the simulation group: trainees are undertaking and completing all tutorials on the simulator and are provided with simulator feedback and continued access during the trial duration. For the non-simulation (control) group: trainees are permitted to accessing the

	unassessed tutorial/tasks mode before monthly assessment to familiarise themselves with the simulator. At the baseline phase of trial, the simulator is used to assess baseline skills of each subject to test their eligibility to participate. After that, subjects' skills are re-assessed every month for evaluating acquisition during the trial (within 6 months). During the trial, the intervention group receive structured self-directed simulation training while the control group are allowed to access (unassessed practice) modules on the simulator with no feedback provided. Both groups are permitted to receive conventional training in the normal way. **Who can participate?** The study subjects are recruited primarily from specialist trainees (ST1-ST7 level) in Obstetrics and Gynaecology in the Welsh Deanery, other NHS staff and students of the MSc programme in Cardiff University who fulfil the inclusion criteria **What are the possible benefits and risks of participating?** Study subjects will receive a detailed report and feedback about their skill assessment. They would be made aware of the skill areas that need further training as well as recognising how their skills have been gradually improved by the simulator (for intervention/simulator group). In the control group, they would be able to access ultrasound-learning material throughout the study and gain access to lectures to build up their basic knowledge of ultrasound. At the end of the trial they would be able to have access to the simulator if required to do so. This research is risk-free as it is a practical training but without involving patients. However, study subjects are advised to observe the general and professional guidelines related to avoidance of repetitive stress and injury during conventional scanning in their clinical settings or while using the simulator. **Where is the study run from?** The study runs from three training hospitals in Wales, University Hospital of Wales, Cardiff, UK Singleton Hospital, Swansea, UK Wrexham Maelor Hospital, Wrexham, UK. Sampling selected from different Welsh Deanery training hospitals across North and South Wales, UK. **When is study starting and how long is it expected to run for?** The study started in May 2011 and will run until the required number of 60 trainees have been recruited and assessed. **Who is funding the study?** The research is organised and supervised by the Department of Obstetrics and Gynaecology, Cardiff School of Medicine, University Hospital of Wales. The study is funded by a scholarship from Saudi Arabia government. **Who is the main contact?** Miss Amal Alsalamah alsalamahA@cf.ac.uk
Ethics approval	The study has been reviewed and ethically approved by South East Wales Ethics Research Committee Panel B. Cardiff, UK
Study design	Parallel randomised control trial
Countries of recruitment	Wales, UK

Disease/condition/study domain	Improving ultrasound skills and training competency for trainees
Participants - inclusion criteria	1.Subjects should have (none or limited) ultrasound experience of any kind, 2.None or limited previous access to transvaginal ultrasound experience, 3.Motivated to learn transvaginal ultrasound skills; intends to complete the requirements of the RCOG ultrasound training curriculum or a similar structured programme; based within Wales or within a very short distance of travel to Cardiff.
Participants - exclusion criteria	1. Individuals who have already completed any structured ultrasound training programme and certified accordingly. 2. Individuals at consultant level in obstetrics and gynaecology or radiology even if they have no ultrasound experience. 3. Radiographers on the ultrasound MSc programme with previous experience in transvaginal ultrasound scanning.
Anticipated start date	15/05/2011
Anticipated end date	31/12/2013
Status of trial	On-going
Patient information material	Not available in web format, please use the contact details below to request a patient information sheet
Target number of participants	60-80
Interventions	The intervention group were instructed to use the simulator to practice ultrasound skills in addition to receiving clinical training arranged by their hospital, whereas the control group received conventional training alone.
Primary outcome measure(s)	Difference in ultrasound scan performance (OSATS) in obstetrics and gynaecology between simulation- supported trained and untrained subjects.
Secondary outcome measure(s)	1. Determine factors that have affected engagement in the project and acquisition of relevant skills. This will provide information on barriers and limitations of simulation training in a clinical environment. 2. Evaluate the validity of task performance sheet for assessing relevant skills in comparison to the simulator's task metrics. 3. Overall subjective ultrasound skills acquisition with/out simulation training 4. Differences in skills scores at each assessment session
Sources of funding	Scholarship funding of Saudi Arabia government
Trial website	
Publications	
Contact name	**Miss Amal Alsalamah**
Address	University Hospital of Wales Heath Park Obstetrics and Gynaecology department
City/town	Cardiff
Zip/Postcode	CF14 4XN
Country	UK

Sponsor	Cardiff University
Address	c/o Miss Amal Alsalamah University Hospital of Wales Heath Park Obstetrics and Gynaecology department
City/town	Cardiff
Zip/Postcode	CF14 4XN
Country	UK
Sponsor website:	www.cf.ac.uk
Date applied	11/02/2013
Last edited	
Date ISRCTN assigned	16/01/2012

Appendix (3.1)

Face Validity Questionnaire: The ScanTrainer Ultrasound Simulator

Please use Visual Analogue Score (VAS) to rate all questions
by placing an "X" where appropriate e.g.

0 (very bad) ——————————X—————————— 10 (very good)

1.	Relevance of the simulator for actual transvaginal ultrasound scanning
	0 ——————————— 10
2.	Realism of the simulator to simulate the transvaginal scan of female pelvis
	0 ——————————— 10
3.	Realism of the simulator to simulate the movements possibly required to perform in the female pelvic anatomy (uterus, ovaries/adnexa, POD)
	0 ——————————— 10
4.	Realism of the ultrasound image generated during the performance
	0 ——————————— 10
5.	Force feedback provided on the operator's hand to simulate real scan
	0 ——————————— 10
6.	Realism of simulator to provide actual action of all buttons provided in the control panel
	0 ——————————— 10
7.	Realism of the simulator to provide the endometrial thickness measurement in gynaecology task
	0 ——————————— 10
8.	Realism of the simulator to provide measurements of the ovary in gynaecology task
	0 ——————————— 10
9.	Ability to test normal gynaecological anatomy: uterus, adnexa and Pouch of Douglas
	0 ——————————— 10
10.	Ability to test early pregnancy structures: Fetus, viability and placenta
	0 ——————————— 10
11.	Realism of the simulator to provide the CRL measurement in early pregnancy task
	0 ——————————— 10

Appendix 3.1

12. Relevance of the simulator's learning resource, videos and ScanTutor function
_____ 0 10
13. Overall value of the simulator as a training tool
_____ 0 10
14. Overall value of the simulator as a testing tool
_____ 0 10

Details about yourself **(tick relevant box or provide written answers):**

Name:

Age: **Gender:** ☐ Female ☐ Male

Hospital/ Department:

1. State the grade/band /post you currently hold?

☐Consultant _____ ☐Specialist trainee _____

☐ Specialist _____ years of training ST_____

☐Nurse Hysteroscopist ☐Medical student _____

☐ Other (please specify) _____

Country of practice _____

2. State the number of years you have been practising ultrasound in clinics:

☐ Never ☐ < 6 months ☐ 6 – 11 months ☐ 1-2 years ☐ > 2 years

3. Clarify your transvaginal ultrasonography experience, you are….

☐ Independent practitioner ☐ Trainee under supervision ☐ Trainer/tutor

☐ Other (*please describe)* _____

4. How often do you scan?

☐ Never ☐ Daily ☐ Once a week ☐ Once a month ☐ Occasionally

☐ Other (*please describe)* _____

5. Previous experience with ScanTrainer ultrasound simulator? ☐Yes ☐ No

6. Previous experience with any other ultrasound models? ☐Yes ☐ No

Appendices

(4.1-4.2)

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

Date:	**Trainee code:**
Checked by main evaluator – Rater 1:	**Checked by Rater 2:**

Audit files copied to:	○ **Validation** ○ **Assessment**

Module	Skills	Rater 1- Undertook task correctly			Rater 2- Undertook task correctly		Comments
8.1 Gyencology Core skills Full examination Normal uterus AvU	Uterus correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Uterus correctly examined in the coronal plane	2	Yes	No	Yes	No	
	Left ovary correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Left ovary correctly examined in the coronal plane	4	Yes	No	Yes	No	
	Right ovary correctly examined in the sagittal plane	5	Yes	No	Yes	No	
	Right ovary correctly examined in the coronal plane	6	Yes	No	Yes	No	
	Pouch of Douglas correctly examined the sagittal plane	7	Yes	No	Yes	No	

Global Rating Scale (GRS): Evaluation of Video- record performance on Ultrasound Simulator

- Please rate the trainee on the performance statement according to the following scale
- Please use NA if statement is not relevant.

	Performance General evaluation (GYN1) (Anteverted uterus)	NA 0	Very Poor 1	Poor 2	Fair 3	Good 4	Very good 5
	Uterus seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Uterus seen in transverse plane	☐	☐	☐	☐	☐	☐
	Left ovary seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Left ovary seen in transverse plane	☐	☐	☐	☐	☐	☐
	Right ovary seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Right ovary seen in transverse plane	☐	☐	☐	☐	☐	☐
	POD visualised	☐	☐	☐	☐	☐	☐
	Perform systematic scan	☐	☐	☐	☐	☐	☐

/40

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

COMMENTS

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

Date:	**Trainee code:**
Checked by main evaluator – Rater 1:	**Checked by Rater 2:**

Audit files copied to:	○ **Validation** ○ **Assessment**

Module	Skills		Rater 1- Undertook task correctly		Rater 2- Undertook task correctly		Comments
9.2 Gyencology core skills Full examination Retroverted uterus (RvU)	Uterus correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Uterus correctly examined in the coronal plane	2	Yes	No	Yes	No	
	Left ovary correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Left ovary correctly examined in the coronal plane	4	Yes	No	Yes	No	
	Right ovary correctly examined in the sagittal plane	5	Yes	No	Yes	No	
	Right ovary correctly examined in the coronal plane	6	Yes	No	Yes	No	
	Pouch of Douglas correctly examined the sagittal plane	7	Yes	No	Yes	No	

Global Rating Scale (GRS): Evaluation of Video- record performance on Ultrasound Simulator

- Please rate the trainee on the performance statement according to the following scale
- Please use NA if statement is not relevant.

	Performance General evaluation (GYN2) (Retroverted uterus)	NA 0	Very Poor 1	Poor 2	Fair 3	Good 4	Very good 5
	Uterus seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Uterus seen in transverse plane	☐	☐	☐	☐	☐	☐
	Left ovary seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Left ovary seen in transverse plane	☐	☐	☐	☐	☐	☐
	Right ovary seen in sagittal plane	☐	☐	☐	☐	☐	☐
	Right ovary seen in transverse plane	☐	☐	☐	☐	☐	☐
	POD visualised	☐	☐	☐	☐	☐	☐
	Perform systematic scan	☐	☐	☐	☐	☐	☐

/40

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

COMMENTS

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

Date:	**Trainee code:**
Checked by main evaluator – Rater 1:	**Checked by Rater 2:**

Audit files copied to:	○ **Validation** ○ **Assessment**

Module	Skills	Rater 1- Undertook task correctly			Rater 2- Undertook task correctly		Comments
7.1 Obstetrics core skills 11 weeks early pregnancy (EP)	Gestational sac correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Foetal heart correctly examined in the sagittal plane	2	Yes	No	Yes	No	
	Foetus correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Labelling the Yolk sac	4	Yes	No	Yes	No	
	Yolk sac correctly viewed	5	Yes	No	Yes	No	
	Yolk sac correctly magnified	6	Yes	No	Yes	No	
	Placenta correctly examined in the sagittal plane	7	Yes	No	Yes	No	

Global Rating Scale (GRS): Evaluation of Video- record performance on Ultrasound Simulator

- Please rate the trainee on the performance statement according to the following scale
- Please use NA if statement is not relevant.

	Performance General evaluation of early pregnancy –11wks	NA 0	Very Poor 1	Poor 2	Fair 3	Good 4	Very good 5
	Scan GS in sagittal plane	☐	☐	☐	☐	☐	☐
	Confirm foetal viability	☐	☐	☐	☐	☐	☐
	Scan fetus	☐	☐	☐	☐	☐	☐
	Identify Yolk Sac YS	☐	☐	☐	☐	☐	☐
	Correctly viewing YS	☐	☐	☐	☐	☐	☐
	Correctly magnifying YS	☐	☐	☐	☐	☐	☐
	Identify placenta	☐	☐	☐	☐	☐	☐
	Perform systematic scan	☐	☐	☐	☐	☐	☐

/40

OSATS Skill Checklist:

Evaluation of Video- record performance on Ultrasound Simulator

COMMENTS

Appendix (4.2): Criteria based assessment of ultrasound skills

Performance	NA mark=0	Very poor mark=1	Poor mark=2	Fair mark=3	Good mark=4	Excellent mark=5
General evaluation of female pelvic anatomy:						
Uterus examined in sagittal plane Uterus examined in coronal plane	Not attempted	**Not** seen or Completely fouled up	**Seen** the uterus but scan incorrectly performed	**Scan** is correctly performed but not acceptable image optimisation	**Scan** correctly performed with acceptable image optimisation	**Scan** correctly with image optimisation and confident
Left ovary examined in sagittal plane Left ovary examined in coronal plane	Not attempted	**Not** seen the uterus or Completely fouled up	**Seen** the ovary but scan incorrectly performed	**Scan** correctly performed but not acceptable image optimisation	**Scan** is correctly performed with acceptable image optimisation	**Scan** correctly with image optimisation and confident
Right ovary examined in sagittal plane Right ovary examined in coronal plane	Not attempted	**Not** seen or Completely fouled up	**Seen** the ovary but scan incorrectly performed	**Scan** correctly performed but not acceptable image optimisation	**Scan** is correctly performed with acceptable image optimisation	**Scan** correctly with image optimisation and confident
POD examined	Not attempted	**Not** located or Completely wrong location	**Located** but incorrectly scanned	**Scan** correctly performed but not acceptable image optimisation	**Located**, Scanned correctly with acceptable image optimisation	**Located**, scanned correctly with image optimisation and confident
Perform systematic scan	Not attempted	**One** skill approached systematically	**Few** skills approached systematically (2-3 skills)	**Some** skills approached systematically (4 skills)	**Many** skills approached systematically (5-6 skills)	**All** skills approached systematically (7 skills)
General evaluation of early pregnancy –11wks:						
Examine GS in sagittal plane	Not attempted	**GS** not scanned	**Seen** the GS but scan incorrectly performed	**GS** seen, correctly scanned but not acceptable image optimisation	**GS** seen, correctly scanned with acceptable image optimisation	**Scan** correctly performed with image optimisation and confident
Confirm foetal viability	Not attempted	**Not** detected	**Not** detected	**Detected** but not acceptable image optimisation	**FH** detected and confirmed with acceptable image optimisation	**FH** detected with confident and image optimisation
Scan fetus	Not attempted	**Not** seen or completely fouled up	**Seen** but scan is incorrectly performed	**Seen** and scanned correctly but not acceptable image optimisation	**Scan** is correctly performed with acceptable image optimisation	**Scan** correctly performed with image optimisation and confident
Identify yolk sac 'YS'	Not attempted	**Not** performed or completely wrong identification	**Identified** but not labelled correctly	**Identified** and labelled but near to the exact position	**Identified** and labelled correctly	**Identified** and labelled correctly with confident
Correctly viewing YS	Not attempted	Not centralised or completely wrong identification of YS	Performed but incorrect position	Performed but not acceptable image centralisation	Performed correctly with acceptable centralisation	Performed correctly and centralised with confident
Correctly magnifying YS	Not attempted	**Not** performed or completely wrong identification of YS	**Performed** but incorrect	**Performed** but not acceptable magnification	**Performed** correctly with acceptable magnification	**Performed** correctly and with confident
Identify placenta	Not attempted	**Not** located or wrongly located	**Located** but incorrectly scanned	**Scan** correctly performed but not acceptable image optimisation	**Located**, Scanned correctly with acceptable image optimisation	**Located**, scanned correctly with image optimisation and confident
Perform systematic approach (in the order above)	Not attempted	**One** skill approached systematically	**Few** skills approached systematically (2-3 skills)	**Some** skills approached systematically (4 skills)	**Most** of skills approached systematically (5-6 skills)	**All** skills approached systematically (7 skills)

Appendices

(5.1-5.2)

Appendix (5.1)

Assessing ultrasound skills using ScanTrainer simulator: evaluation simulation metric (Skill checklist)

Date	Participant code/group
Checked by (assessor name)	• Validation

Module TVS-G-CS001.4 (Assignment)	Core skills Gynaecology module (Assignment- Task)	Individual Assignments IS					Full assignment: 8.2 FE			
		Rater		Simulator			Rater		Simulator	
		P	F	P	F		P	F	P	F
6.1	6.1 T3 SK1:Uterus correctly examined in the sagittal plane	Yes	No	Yes	No		Yes	No	Yes	No
	6.1 T5 SK2:Uterus correctly examined in the coronal plane	Yes	No	Yes	No		Yes	No	Yes	No
7.2	7.2 T2 Scan in CORONAL SK3:Right ovary correctly examined in the sagittal plane *(1-Split image- identify-coronal scan-locate in largest diameter-capture image- 2-Press split image – rotate to sagittal)*	Yes	No	Yes	No		Yes	No	Yes	No
	7.2 T3 Scan in SAGITTAL SK4:Right ovary correctly examined in the coronal plane *(Identify- examine- locate in largest diameter- freeze)*	Yes	No	Yes	No		Yes	No	Yes	No
7.3	7.3 T2 Scan in CORONAL SK5:Left ovary correctly examined in the sagittal plane	Yes	No	Yes	No		Yes	No	Yes	No
	7.3 T3 Scan in SAGITTAL SK6:Left ovary correctly examined in the coronal plane	Yes	No	Yes	No		Yes	No	Yes	No

Module TVS-O-CS-001.4 (Assignment)	Core skills 'Early pregnancy' module (Assignment- Task)	Individual Assignments IS					Full assignment: 7.2 FE			
		Rater		Simulator			Rater		Simulator	
		P	F	P	F		P	F	P	F
4.1	4.1 T2 SK1: GS correctly examined in the sagittal plane *GS in Sagittal* *1-Scan in sagittal* *2-Measurements in Sagittal* *3-Take picture*	Yes	No	Yes	No		Yes	No	Yes	No
	4.1 T4 *GS in Coronal* *1-Scan in coronal* *2-Measurement in coronal* *3-Take picture*	Yes	No	Yes	No					
5.1	5.1 T2 SK2: Fetus correctly examined in the sagittal plane	Yes	No	Yes	No		Yes	No	Yes	No
	5.1 T3 SK3: Detect heart beat	Yes	No	Yes	No		Yes	No	Yes	No
	5.1 T3 Labelling fetal heart	Yes	No	Yes	No					
6.1	6.1 T2 SK4: Yolk sac correctly viewed	Yes	No	Yes	No		Yes	No	Yes	No
	6.1 T2 SK5: Yolk sac correctly labelled	Yes	No	Yes	No		Yes	No	Yes	No
	6.1 T2 SK6: Yolk sac correctly magnified	Yes	No	Yes	No		Yes	No	Yes	No
6.2	6.2 T2 SK7: Placenta correctly examined in the sagittal	Yes	No	Yes	No		Yes	No	Yes	No
	6.2 T2 Placenta correctly labelled	Yes	No	Yes	No					

P= skill achieved correctly (pass), F= skill achieved incorrectly (Fail).
IS=individual skill, FE=full examination task

Appendix (5.2)

Skill Checklist

Date:	Centre:	Trainee code:	Test no:
Checked by evaluator/Rater:			

Audit files copied to:	○ Validation ○ Assessment

Module	Skills	Rater feedback			Simulator feedback		Comments
GYN1 Anteverted Gynaecology Assignment	Uterus correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Uterus correctly examined in the coronal plane	2	Yes	No	Yes	No	
	Left ovary correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Left ovary correctly examined in the coronal plane	4	Yes	No	Yes	No	
	Right ovary correctly examined in the sagittal plane	5	Yes	No	Yes	No	
	Right ovary correctly examined in the coronal plane	6	Yes	No	Yes	No	
	Pouch of Douglas correctly examined the sagittal plane	7	Yes	No	Yes	No	
GYN2 Retroverted Gynaecology Assignment	Uterus correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Uterus correctly examined in the coronal plane	2	Yes	No	Yes	No	
	Left ovary correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Left ovary correctly examined in the coronal plane	4	Yes	No	Yes	No	
	Right ovary correctly examined in the sagittal plane	5	Yes	No	Yes	No	
	Right ovary correctly examined in the coronal plane	6	Yes	No	Yes	No	
	Pouch of Douglas correctly examined the sagittal plane	7	Yes	No	Yes	No	

Appendix (5.2)

Skill Checklist

Date:	Centre:	Trainee code:	Test no:
Checked by main evaluator/rater:			

Audit files copied to:	○ Validation ○ Assessment

Module	Skills	Rater feedback			Simulator feedback		Comments
Early pregnancy assignment	Gestational sac correctly examined in the sagittal plane	1	Yes	No	Yes	No	
	Fetal heart correctly examined in the sagittal plane	2	Yes	No	Yes	No	
	Fetus correctly examined in the sagittal plane	3	Yes	No	Yes	No	
	Labelling the Yolk sac	4	Yes	No	Yes	No	
	Yolk sac correctly viewed	5	Yes	No	Yes	No	
	Yolk sac correctly magnified	6	Yes	No	Yes	No	
	Placenta correctly examined in the sagittal plane	7	Yes	No	Yes	No	

Appendices

(7.1-7.2)

Questionnaire for Experts

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with conventional training approach

Dear Participant,

Miss A Al-Salamah is Ph.D student in Cardiff University and is undertaking research as part of her Ph.D project in Obstetrics and gynaecological sonography. This research is supported and supervised by *Mr N. Amso* and *Dr Neil Pugh*.

This research project aims to determine the trainee's ultrasound learning curves of skills achieved by either conventional or simulation enhanced training. Also to look at the benefit of introducing simulation learning into clinical training and evaluating the new ultrasound simulator "Scan Trainer" as a valid assessment tool.

If you found this research fit your interests, your responses to my questionnaire will be highly valued. You are under no obligation to fill this questionnaire as your involvement is completely voluntary. If you would like to take part, I would very much appreciate your time and effort in filling out this questionnaire, which should take few minutes to complete.

Your answers will be kept strictly confidential, private and anonymous and the data will be kept securely according to the University's policy of academic integrity. If you are unhappy answering any of the questions, please leave them blank.

Thank you in advance for your help.

If you have any queries about this questionnaire, please contact Amal Alsalamah, Ph.D student, Department of Obstetrics and Gynaecology, Cardiff School of Medicine, Heath Park, CF14 4XN Cardiff, UK
Email: alsalamahA@cardiff.ac.uk

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using skills using simulation-supported training with conventional training approach

Questionnaire

We are interested in collecting information about your views on whether ultrasound training supported by simulation can improve or expedite the trainee's skills or even has an effect on the learning curve. This questionnaire should only take you few minutes to complete. Please tick one or more boxes or provide written answers where appropriate. Please complete all sections.

Section 1: Details about yourself **(tick relevant box or provide written answers):**

Your name (optional) _____

1. You are: ☐ Female ☐ Male

2. State the grade/band /post you currently hold?

☐ Consultant ☐ Specialist trainee

☐ Clinical assistant ☐ Associate specialist

☐ Radiographer ☐ Nurse

☐ Midwife ☐ Other (please specify) _____

3. State the number of years you have been using ultrasound in clinics:

☐ Never ☐ Less than 6 months ☐ more than 6m ☐ 1-2 years ☐ More than 2 years

4. Clarify your ultrasonography background and how often do you scan?

☐ Daily ☐ Once a week ☐ Once a month ☐ Other (*please describe*) _____

Further comments

5. How strongly do you agree/disagree that the conventional training (i.e. adhoc or locally organised training as per local practice) which is supported by the addition of an Ultrasound Simulator could improve the quality of training (better trainee's skills and expedited attainment of skills)?

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

6. How strongly do you agree/disagree that conventional training alone would improve the trainee's skills over a long period of time?

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

7. How strongly do you agree/disagree that beginners undertaking their training in the conventional way would acquire core ultrasound skills in a relatively short period!

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

8. Could you indicate how many conventional training sessions (per day/per week/per month) are needed by a trainee to acquire the core skills in a relatively short period time without additional learning material, mannequin or simulator support?

☐ _____session per a day

☐ _____session per a week

☐ _____session per a month

☐ Other (*please describe*) _____

Further comments

9. Could you indicate how many conventional training sessions (per day/per week/per month) are needed by a trainee to acquire the core skills in a relatively short period time if supported by additional learning material, mannequin or simulator?

☐ _____session per a day

☐ _____session per a week

☐ _____session per a month

☐ Other (please describe) _____

Further comments

10. Have you ever heard about the effect of introducing the simulation to support the learning curve of ultrasound skills in short time?

☐ Yes ☐ No

11. If you were a trainee, which of the training methods you prefer to have for your training purposes and why?

☐ Conventional training with long-term practice.

☐ Conventional training supplemented with ultrasound simulation training (e-learning and simulator).

☐ Simulation- based training (using simulator only) especially for core ultrasound skills prior to undertaking scans on a patient/volunteer.

Further comment

12. Taking your own institution circumstances and as an expert in ultrasound, how long would a trainee in your institution take to achieve competency in core ultrasound skills (*see figure 1 for learning outcomes*) utilising current conventional training resources?

☐ Three months

☐ Six months

☐ Nine months

☐ A year

☐ Other (*please describe*) _____

Further comments

13. In your opinion as an expert, how long should it take for the trainee to achieve competency in core ultrasound skills when conventional training is supplemented with simulation training?

☐ Three months

☐ Six months

☐ Nine months

☐ A year

☐ Other (*please describe*) _____

Further comments

14-After you have used the ultrasound simulator, kindly rate the following:

Performance	1 Strongly disagree	2 disagree	3 Neutral	4 agree	5 Strongly agree
1. The simulator provides easy access to practice endovaginal ultrasound scanning					
2. Training with the simulator helps in familiarising the trainee with core ultrasound skills					
3. The simulator can be used as training tool for endovaginal ultrasound scan					
4. The simulator can enhance training quality in a short time					
5. The ultrasound image appeared realistic in this Ultrasound Simulator					
6. The instructions and the 3D depiction of anatomy have a very useful role in understanding the ultrasound image					
7. All buttons in the simulator were handy and well explained					
8. Providing force feedback on the operator's hand helps in simulating the real scanning					
9. Training with the simulator reflects a similar or very close experience of a patient/volunteer scan					
10. The confidence in scanning is enhanced after training with the simulator					
11. The simulator has the potential to be used as an assessment tool for endovaginal ultrasound scan					

I would thank you for completing this questionnaire

Figure 1: *ultrasound skills listed depend on RCOG module for intermediate vaginal scanning.*

Image optimisation
Ability to utilise the image depth when it is necessary
Apply the overall gain
Apply the Time Gain Compensation (TGC)
Place the focus correctly at the level of interest
Utilise image zooming when it is necessary
Apply the image scrolling on cine-loop
Apply orientation-convention on the scan image
Correct image annotation
Identify the female pelvic anatomy (uterus)
Demonstrate the uterus in sagittal and coronal planes
Demonstrate the thickest part of endometrium
Identify any fibroid(s) and measure it
Identify the female pelvic anatomy (adenxa and ovaries)
Identify the right and left ovary in sagittal and coronal planes
Demostarte the normal morphology of the ovary and the cycle variation
Measure any ovary within three diameters
Evaluate right and left adenxa properly
General organ knowledge
Identify the urinary bladder
Identify the urethra and anterior wall of vagina
Demonstrate any free fluid located in pouch of Douglas
Other machine related skills
Machine set-up (switch-on/off), enter patient's data, selecting proper exam and choosing correct probe
Image orientation convention
Ultrasound probe handling and manoeuvre

Participant information form

Validation and determination of the influence of a virtual simulator on the acquisition of ultrasound skills and comparison of learning curves of those using simulation-supported training with a conventional training approach

Name of the researcher: Amal Alsalamah

Supervised by: Professors Nazar Amso & Neil Pugh

Details about yourself **(tick relevant box or provide written answers):**

Your name _____

Hospital title where you currently work at _____

1. You are: ☐ Female ☐ Male

2. State the grade/band /post you currently hold?

☐ Consultant ☐ Specialist trainee

☐ Clinical assistant ☐ Associate specialist

☐ Radiographer ☐ Nurse

☐ Midwife ☐ Other (please specify)_____

Please circle your year of training:

Year [1] [2] [3] [4] [5] [6] [7]

3. State the number of years you have been practising ultrasound in clinics:

☐ Never ☐ Less than 6 months ☐ 6 – 11 months ☐ 1-2 years ☐ More than 2 years

4. Clarify your ultrasonography background and how often do you scan?

☐ Daily ☐ Once a week ☐ Once a month ☐ Other (*please describe*) _____

5. Do you often practise TVS in?

☐ Gynaecological pt. ☐ EP ☐ Both

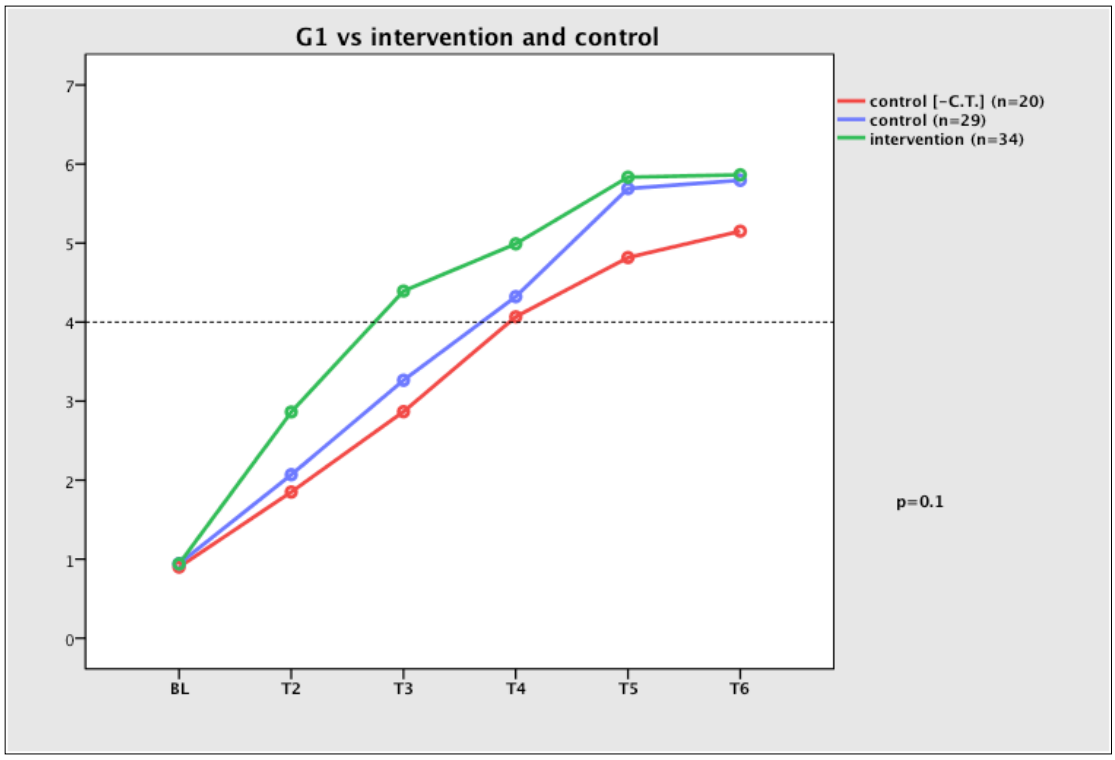
*Please provide further information*_____

Appendix (7.3)

Repeated measures protocol analysis results

Repeated measures data analysis: intention-to-train and subgroups protocol analysis

(Chart 1):G1: control-CT vs intervention and control

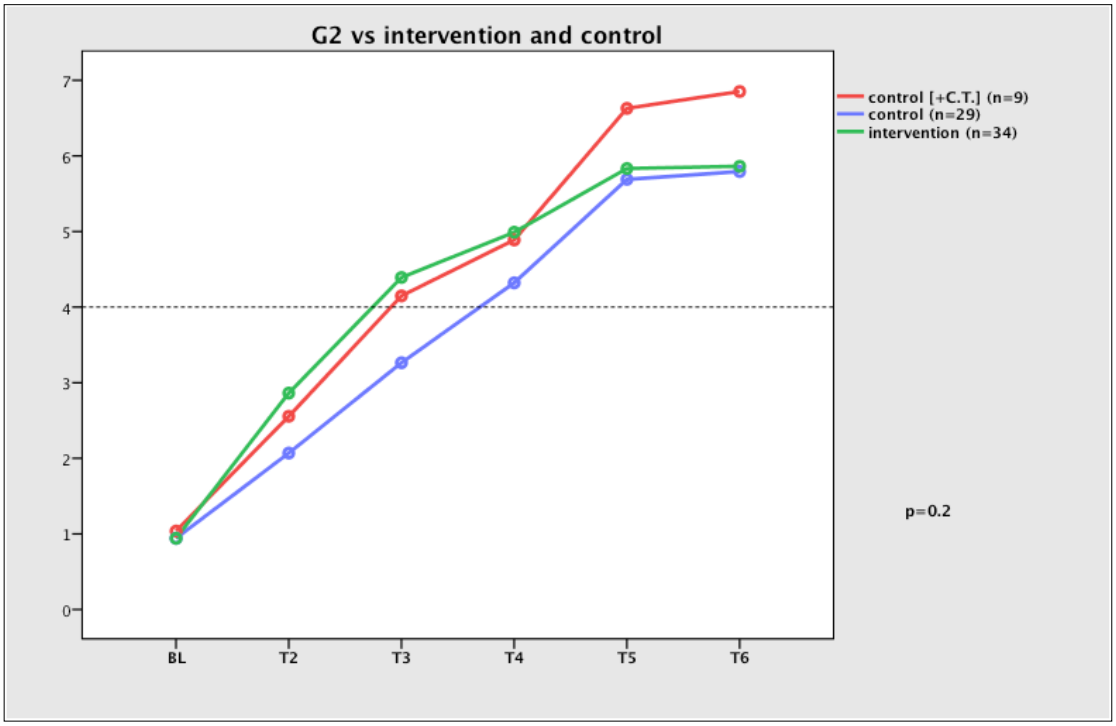


Significance differences tested by repeated measures ANOVA

	Within three groups	G1 vs intervention	G1 vs control
Baseline - test 2	0.01	0.01	0.6
test 2- test 3	0.5	0.2	0.7
test 3- test 4	0.2	0.08	0.7
test 4- test 5	0.5	0.8	0.3
test 5- test 6	0.4	0.2	0.4
Overall	0.1	0.02	0.3

Repeated measures data analysis: intention-to-train and subgroups protocol analysis

(Chart 2): G2: control +CT vs intervention and control

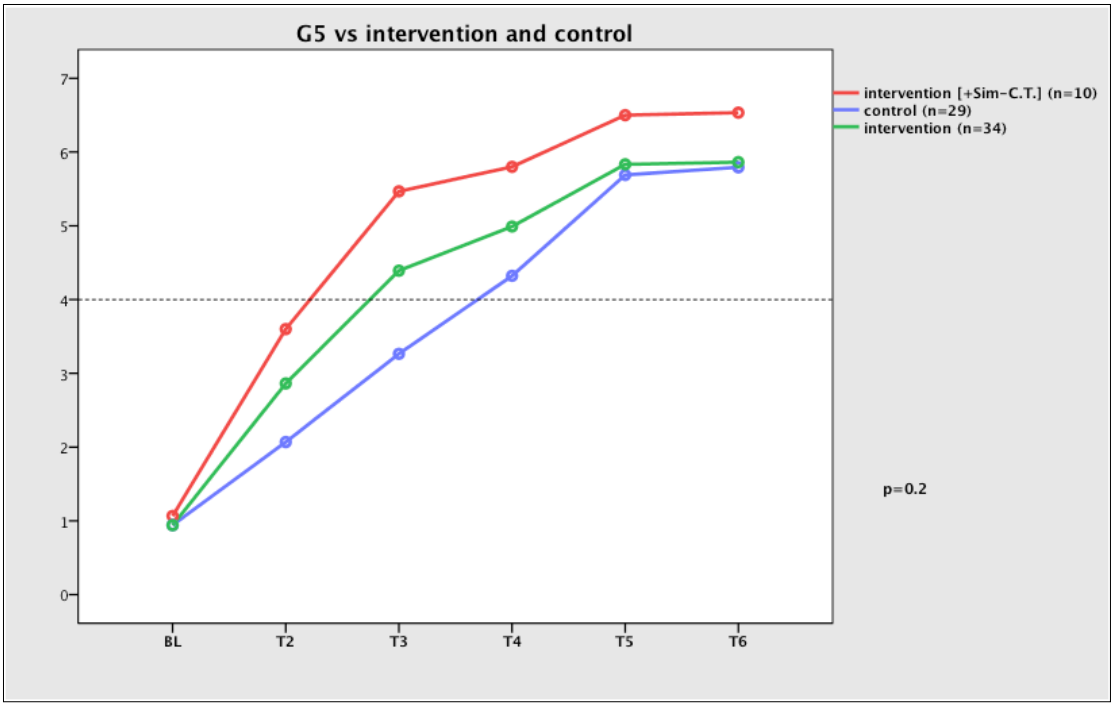


Significance differences tested by repeated measures ANOVA

	Within three groups	G2 vs intervention	G2 vs control
Baseline - test 2	0.07	0.4	0.4
Test 2- test 3	0.6	0.9	0.5
Test 3- test 4	0.2	0.6	0.5
Test 4- test 5	0.4	0.2	0.6
Test 5- test 6	0.8	0.5	0.7
Overall	0.2	0.6	0.2

Repeated measures data analysis: intention-to-train and subgroups protocol analysis

(Chart 3): G5: intention (+Sim-CT) vs intention and control

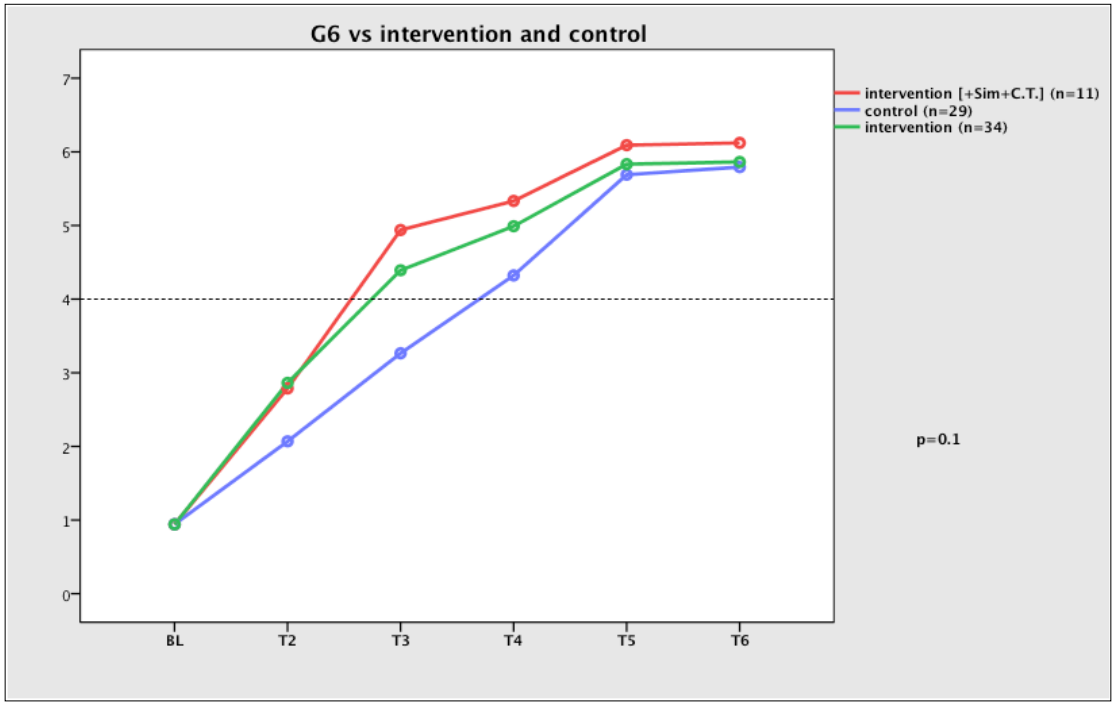


Significance differences tested by repeated measures ANOVA

	Within three groups	G5 vs intervention	G5 vs control
Baseline - test 2	0.009	0.2	0.003
Test 2- test 3	0.5	0.5	0.2
Test 3- test 4	0.1	0.8	0.1
Test 4- test 5	0.5	0.8	0.3
Test 5- test 6	0.9	0.9	0.8
Overall	0.06	0.1	0.03

Repeated measures data analysis: intention-to-train and subgroups protocol analysis

(Chart 4): G6: intervention (+Sim +CT) vs intervention and control



Significance differences tested by repeated measures ANOVA

	Within three groups	G6 vs intervention	G6 vs control
Baseline - test 2	0.09	0.8	0.1
Test 2- test 3	0.2	0.2	0.1
Test 3- test 4	0.1	0.4	0.1
Test 4- test 5	0.5	0.9	0.4
Test 5- test 6	0.9	0.9	0.8
Overall	0.2	0.5	0.1

Appendix (7.4)

Kaplan Meier analysis results:

supplemental report

Kaplan Meier primary analysis:
Attainment of borderline score and maximum score in GYN1, GYN2 and early pregnancy modules
control (n=29) VS intervention (n=34)

GYN1
Attainment of Borderline score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	2.759	1	0.09

GYN1
Attainment of Maximum score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	0.771	1	0.3

GYN2
Attainment of Borderline score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	2.178	1	0.1

GYN2
Attainment of Maximum score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	1.605	1	0.2

Early pregnancy
Attainment of Borderline score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	0.609	1	0.4

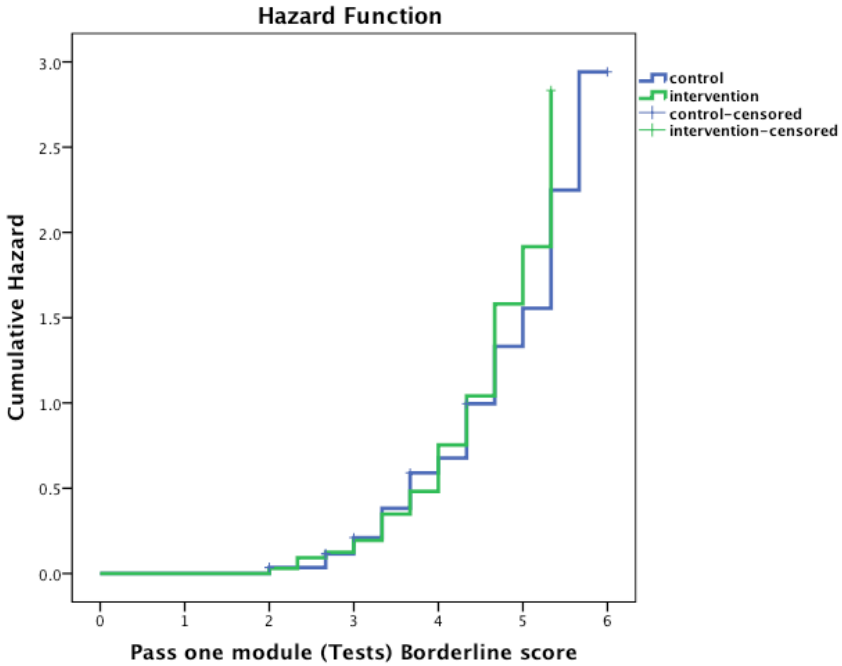
Early pregnancy
Attainment of Maximum score of 7 by number of (Tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	0.612	1	0.4

Kaplan Meier (primary and secondary) analysis
Attainment of primary and final end-points to pass (one) and (three) modules

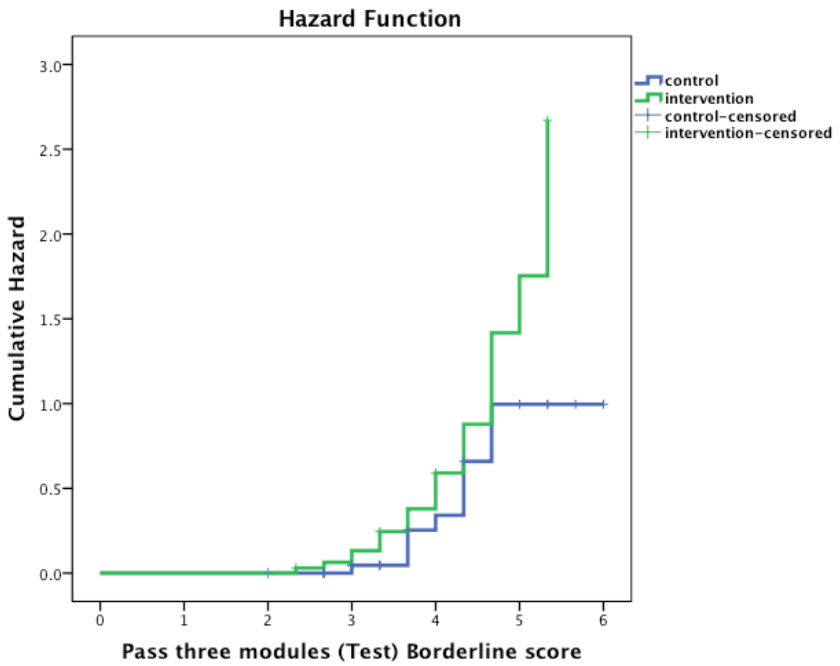
Pass one module
Attainment of Borderline score (4 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	0.221	1	0.6



Pass three modules
Attainment of Borderline score (4) by number of (tests)

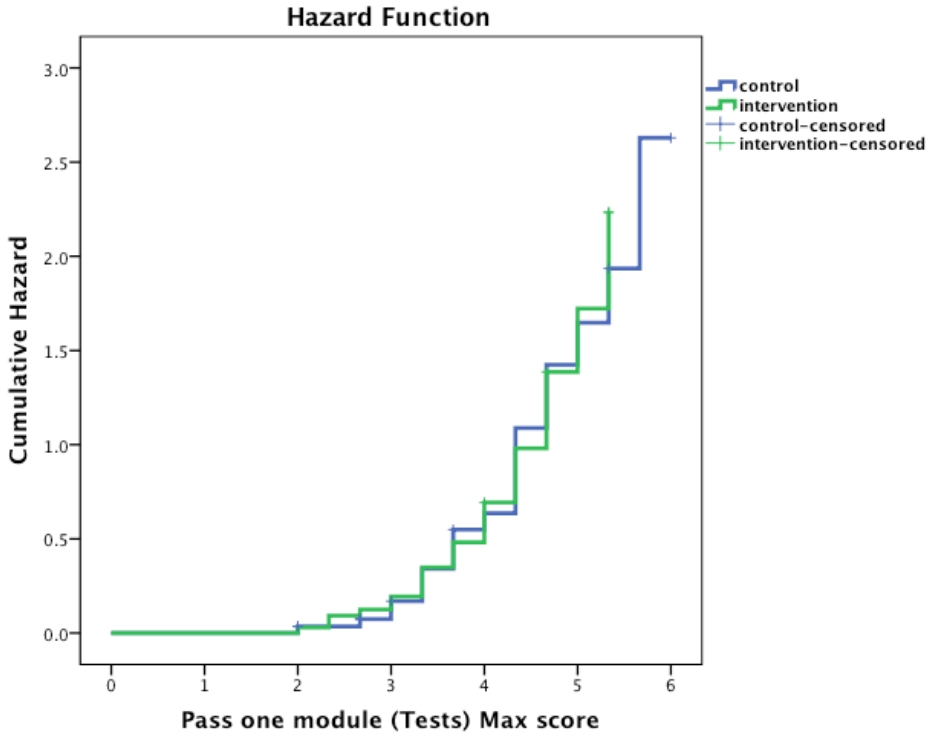
	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	4.425	1	0.03



Kaplan Meier (primary and secondary) analysis
Attainment of primary and final end-points to pass (one) and (three) modules

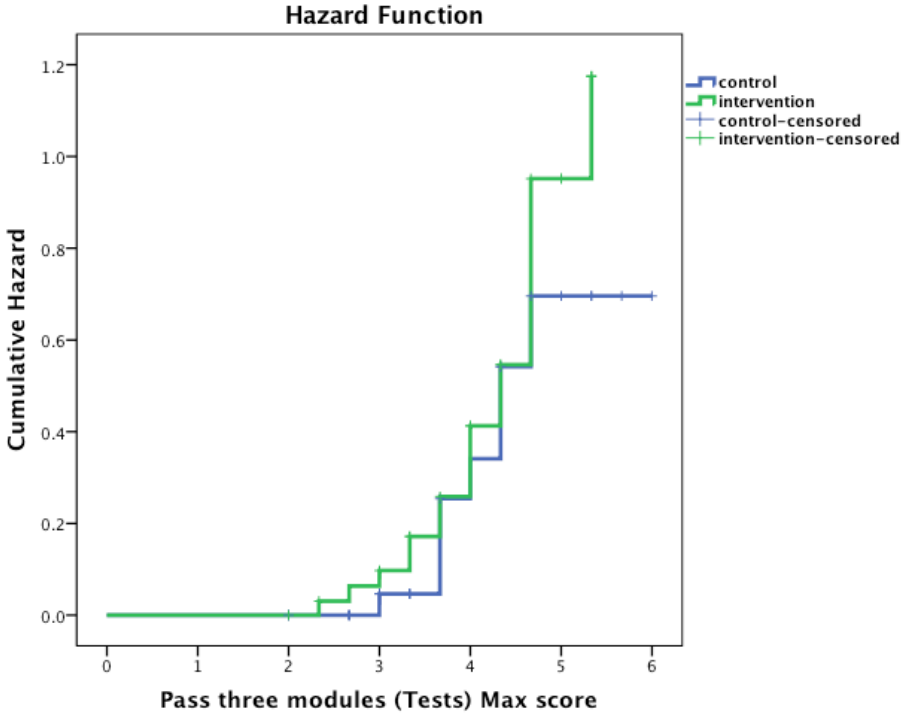
Pass one module
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	0.71	1	0.7



Pass three modules
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
Control (n=29) vs Intervention (n=34)	1.199	1	0.2



Kaplan Meier (primary and secondary) analysis
Attainment of primary and final end-points to pass (one) and (three) modules

Pass one module
Attainment of Borderline score (4) in (days)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	12.443	3	0.006

Pass three modules
Attainment of Borderline score (4) in (days)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	29.772	3	0.0001

Pass one module
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	4.829	3	0.18

Pass three modules
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	15.341	3	0.002

Pass one module
Attainment of Maximum score of 7 in (days)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	12.812	3	0.005

Pass three modules
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	18.955	3	0.0001

Pass one module
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	5.514	3	0.13

Pass three modules
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group (four subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	9.236	3	0.026

Kaplan Meier (primary and secondary) analysis
Attainment of primary and final end-points to pass (one) and (three) modules

Pass one module
Attainment of Borderline score (4) in (days)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	8.964	3	0.03

Pass one module
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	2.830	1	0.4

Pass one module
Attainment of Maximum score of 7 in (days)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	9.301	3	0.02

Pass one module
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	2.895	3	0.4

Pass three modules
Attainment of Borderline score (4) in (days)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	24.377	3	0.0001

Pass three modules
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	14.181	3	0.003

Pass three modules
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	12.727	3	0.005

Pass three modules
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group (six subgroup)	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	6.482	3	0.09

Kaplan Meier (primary and secondary) analysis
Attainment of primary and final end-points to pass (one) and (three) modules

Kaplan Meier secondary analysis: FOUR subgroup
G1: control-CT, G2: control+CT, G3:intervention (+Training), G4:intervention (no Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	23.413	3	0.0001

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	15.268	3	0.002

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	7.901	3	0.048

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	5.539	3	0.1

Kaplan Meier secondary analysis: FOUR subgroup
G1: control-CT, G2: control+CT, G3:intervention (+Training), G4:intervention (no Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	13.882	3	0.003

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	13.774	3	0.003

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	13.459	3	0.004

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	14.354	3	0.002

Kaplan Meier secondary analysis: FOUR subgroup
G1: control-CT, G2: control+CT, G3:intervention (+Training), G4:intervention (no Training)

Early pregnancy
Attainment of Borderline score (5) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	12.303	3	0.006

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	13.790	3	0.003

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	4.530	3	0.21

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G3:Intervention (+training), (n=21) G4:Intervention (no training), (n=13)	5.415	3	0.14

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G3: intervention (+ Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	22.823	1	0.0001

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	14.866	1	0.0001

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	6.807	1	0.009

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	4.117	1	0.04

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G3: intervention (+ Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	11.392	1	0.001

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	10.845	1	0.001

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	5.247	1	0.022

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	5.893	1	0.015

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G3: intervention (+ Training)

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	10.846	1	0.001

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	12.515	1	0.0001

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	2.612	1	0.1

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G3:intervention (+training),(n=21)	5.324	1	0.012

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G4: intervention (no Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.949	1	0.1

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.838	1	0.1

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	2.204	1	0.1

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	2.069	1	0.1

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G4: intervention (no Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.786	1	0.1

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	3.633	1	0.057

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	7.791	1	0.005

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	8.590	1	0.003

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G4: intervention (no Training)

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.026	1	0.3

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.498	1	0.2

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	1.458	1	0.2

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G4:intervention (no training),(n=13)	2.231	1	0.1

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G3: intervention (+ Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.952	1	0.3

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.447	1	0.5

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.055	1	0.8

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.952	1	0.3

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G3: intervention (+ Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.077	1	0.7

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.325	1	0.5

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	2.863	1	0.09

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	2.863	1	0.09

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G3: intervention (+ Training)

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	1.190	1	0.2

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	1.225	1	0.2

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.877	1	0.3

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G3:intervention (+ Training),(n=21)	0.002	1	0.9

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G4: intervention (no Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	1.325	1	0.2

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.852	1	0.3

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.529	1	0.4

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.609	1	0.4

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G4: intervention (no Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	2.464	1	0.1

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	2.464	1	0.1

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.131	1	0.7

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.131	1	0.7

Kaplan Meier secondary analysis (four subgroup)
G2: Control +CT VS G4: intervention (no Training)

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.748	1	0.3

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.491	1	0.4

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.516	1	0.4

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G4:intervention (no training),(n=13)	0.038	1	0.8

Kaplan Meier secondary analysis (four subgroup)
G3: intervention (+Training) VS G4: intervention (no Training)

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	7.691	1	0.006

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	4.745	1	0.29

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	0.596	1	0.4

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	0.152	1	0.6

Kaplan Meier secondary analysis (four subgroup)
G3: intervention (+Training) VS G4: intervention (no Training)

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	4.340	1	0.037

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	2.759	1	0.09

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	1.688	1	0.1

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	1.688	1	0.1

Kaplan Meier secondary analysis (four subgroup)
G3: intervention (+Training) VS G4: intervention (no Training)

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	4.334	1	0.03

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	4.507	1	0.034

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	0.003	1	0.9

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G3:intervention (+Training),(n=21) G4:intervention (no training),(n=13)	0.100	1	0.7

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G2: Control +CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	8.031	1	0.005

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	6.206	1	0.013

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	4.428	1	0.035

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	4.467	1	0.035

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G2: Control +CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	8.932	1	0.003

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	12.398	1	0.0001

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	9.469	1	0.002

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	10.505	1	0.001

Kaplan Meier secondary analysis (four subgroup)
G1: Control -CT VS G2: Control +CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	1.928	1	0.1

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	2.825	1	0.09

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	3.413	1	0.06

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G2:Control (+CT),(n=9)	2.401	1	0.1

Kaplan Meier Secondary analysis: Six subgroup
G1: control-CT, G2: control+CT, G5:+SIM-CT, G6:+SIM+CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	26.484	3	0.0001

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	20.667	3	0.0001

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	26.484	3	0.011

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	11.034	3	0.012

Kaplan Meier Secondary analysis: Six subgroup
G1: control-CT, G2: control+CT, G5:+SIM-CT, G6:+SIM+CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	13.975	3	0.003

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	13.528	3	0.004

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	12.117	3	0.007

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	11.853	3	0.008

Kaplan Meier Secondary analysis: Six subgroup
G1: control-CT, G2: control+CT, G5:+SIM-CT, G6:+SIM+CT

Early pregnancy
Attainment of Borderline score (5) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	11.278	3	0.01

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	11.463	3	0.009

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	4.597	3	0.2

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (no training), (n=20) G2:Control (+training), (n=9) G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	5.136	3	0.16

Kaplan Meier secondary analysis (six subgroup)
G1: Control -CT VS G5: +SIM-CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	13.648	1	0.0001

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	4.413	1	0.03

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	2.402	1	0.12

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	0.188	1	0.6

Kaplan Meier secondary analysis (six subgroup)
G1: Control -CT VS G5: +SIM-CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	17.726	1	0.0001

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	14.394	1	0.0001

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	6.291	1	0.01

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G5:Intervention (+Sim-CT), (n=10)	5.955	1	0.01

Kaplan Meier secondary analysis (six subgroup)
G1: Control -CT VS G5: +SIM-CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	12.092	1	0.001

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	8.954	1	0.003

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	1.792	1	0.18

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	2.636	1	0.1

Kaplan Meier secondary analysis (six subgroup)
G1: Control +CT VS G6: +SIM+CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	21.474	1	0.0001

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	17.650	1	0.0001

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	7.891	1	0.005

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	7.695	1	0.006

Kaplan Meier secondary analysis (six subgroup)
G1: Control +CT VS G6: +SIM+CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	5.340	1	0.021

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	6.593	1	0.010

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	2.503	1	0.1

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	3.967	1	0.046

Kaplan Meier secondary analysis (six subgroup)
G1: Control +CT VS G6: +SIM+CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	5.566	1	0.08

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	9.377	1	0.002

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	1.061	1	0.3

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G1:Control (-CT),(n=20) G6:Intervention (+Sim+CT), (n=11)	4.115	1	0.04

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G5: +SIM-CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.001	1	0.9

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.525	1	0.4

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	1.319	1	0.2

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	1.974	1	0.1

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G5: +SIM-CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	17.726	1	0.0001

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.257	1	0.6

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.508	1	0.4

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.508	1	0.4

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G5: +SIM-CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	1.293	1	0.2

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.575	1	0.4

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.568	1	0.4

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G5:Intervention (+Sim-CT), (n=10)	0.060	1	0.8

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G6: +SIM+CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	2.176	1	0.4

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	1.927	1	0.1

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.480	1	0.4

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.497	1	0.4

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G6: +SIM+CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.234	1	0.6

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.234	1	0.6

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	4.291	1	0.03

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	4.292	1	0.038

Kaplan Meier secondary analysis (six subgroup)
G2: Control +CT VS G6: +SIM+CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.282	1	0.5

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.685	1	0.4

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	1.520	1	0.2

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G2:Control (+CT),(n=9) G6:Intervention (+Sim+CT), (n=11)	0.004	1	0.9

Kaplan Meier secondary analysis (six subgroup)
G5: +SIM-CT VS G6: +SIM+CT

GYN1
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	2.498	1	0.1

GYN1
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	4.208	1	0.04

GYN1
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	2.463	1	0.1

GYN1
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	4.179	1	0.041

Kaplan Meier secondary analysis (six subgroup)
G5: +SIM-CT VS G6: +SIM+CT

GYN2
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	1.262	1	0.2

GYN2
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	0.254	1	0.6

GYN2
Attainment of Borderline score (4) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	1.192	1	0.2

GYN2
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	0.060	1	0.8

Kaplan Meier secondary analysis (six subgroup)
G5: +SIM-CT VS G6: +SIM+CT

Early pregnancy
Attainment of Borderline score (4) in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	1.493	1	0.2

Early pregnancy
Attainment of Maximum score of 7 in (Days)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	0.455	1	0.5

Early pregnancy
Attainment of Borderline score (5) by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	0.92	1	0.7

Early pregnancy
Attainment of Maximum score of 7 by number of (tests)

	Log Rank (Mantel-Cox)		
Group	Chi-Square	df	Significance
G5:Intervention (+Sim-CT), (n=10) G6:Intervention (+Sim+CT), (n=11)	0.177	1	0.6

Appendices

(8.1-8.3)

Appendix (8.1)

End trial Questionnaire

Validation and determination of the influence of a virtual simulator on the learning curves for the acquisition of ultrasound skills using simulation-based training or conventional training approach

Dear Participant,

I, Amal Alsalamah, would like to take this opportunity to thank you for taking part in the project titled above. This PhD Ultrasound Training Research Project promises that valuable outcomes will arise from your efforts, whether you enrolled in intervention or control groups. As you have reached the end of the trial, I would like to ask you to share your experience about simulation training by completing this questionnaire. This survey aims to investigate the benefits and barriers of training by ScanTrainer Ultrasound simulator. Participation in completing this survey is strictly voluntary. This should take no more than five minutes of your time to complete.

As you were a valued participant, an individual report of your progress during the trial will be provided, as well as a summary of the project's result when it is completed and written in final thesis format.

If you are still interested in receiving further information about the project, please don't hesitate to contact me, or simply show your interest by ticking the relevant box (Part 3) at the end of the questionnaire.

Thank you in advance for your help.

If you have any queries about this questionnaire, please contact Miss Amal Alsalamah, PhD student, Department of Obstetrics and Gynaecology, Cardiff School of Medicine, Heath Park, CF14 4XN Cardiff, UK
Email: alsalamahA@cardiff.ac.uk

Part 1 – General information

1. **Name:**
2. **Hospital:**

3. **Age:**
   - ○ 18–23
   - ○ 24–30
   - ○ 31–40
   - ○ 41–50
   - ○ 51+

4. **Gender:**
   - ○ Male
   - ○ Female

5. **Position:**
   - ○ Consultant/Senior clinician
   - ○ Specialist trainee
   - ○ Postgraduate student
   - ○ Midwife/Nurse
   - ○ Academic
   - ○ Other *(please specify)*: _____

6. **What type of ultrasound training have you received in the last two years?**
   - ○ Transvaginal
   - ○ Trans-abdominal
   - ○ Both
   - ○ None. *Please go to part2*

7. **Where did you receive ultrasound training?** *Select as many as apply*
   - ▪ Gynaecology clinics
   - ▪ Antenatal clinics
   - ▪ Early pregnancy unit
   - ▪ Other *(Please specify)*: _____

Part 2 – Perceived advantages and disadvantages of simulation in ultrasound training

1. **Had you received any other ultrasound training prior to your participation in the trial? such as practising with …** *please select as many as apply*
   - ○ Mannequins
   - ○ Ultrasound simulator
   - ○ Both
   - ○ Other *(please specify)*: _____
   - ○ None. *Please go to part 3–Question 11*

2. **The factors that influenced you to engage in ultrasound training were:** *Select as many as apply*
   - ▪ RCOG modules requirements
   - ▪ Tutor/ trainer's request
   - ▪ Attending ultrasound workshop
   - ▪ Other *(please specify)*: _____

3. **Where were you able to access a ScanTrainer simulator?** *Select as many as apply*
   - ○ Labour ward at University Hospital of Wales (UHW)
   - ○ Cardiff Medicentre
   - ○ Swansea
   - ○ Wrexham
   - ○ Other *(please specify)*: _____

4. **In contrast to other methods of training (e.g. mannequins with an ultrasound machine, without a trainer, or clinical training sessions with patients and trainer), how do you feel the simulator compared in terms of its effectiveness in familiarising the trainee with the core skills of transvaginal scan?**
   - ○ Better
   - ○ Equivalent
   - ○ Worse

5. **Are there any particular differences that you have noticed when you practised on the ScanTrainer simulator and mannequin, in terms of their advantages and disadvantages? If so, please describe below**

_____

_____

6. **In general, do you think that you have benefited from the use of the ScanTrainer simulator, in addition to any other training that you have received (e.g. practising on mannequins)?**
   - ○ Yes
   - ○ No

7. **Do you feel that the simulator has particular benefits in terms of enhancing trainees' competence in learning the core skills of transvaginal scan? If so, please describe below.**
   - ○ Yes
   - ○ No
   - **Benefits: ----------**

8. **Do you feel that the simulator has a particular limitation in terms of enhancing trainees' competence in learning the core skills of transvaginal scan? If so, please describe below.**
   - ○ Yes
   - ○ No
   - **Limitations: ----------**

9. **Were there any aspects of your training that you felt the ScanTrainer simulator was particularly good at teaching or assessing? If so, please describe below**
   - ○ Yes
   - ○ No

_____

_____

10. **Overall, do you feel that utilisation of the ScanTrainer simulator was a useful and helpful addition to your clinical training?**
    - ○ Yes
    - ○ No

Part 3 – Barriers to engagement with the ScanTrainer simulator

1. **How easy did you find it was to access the simulator?** (Please mark a cross at an appropriate point on the scale below.)

   Very difficult [0] ——————————————— **[10] Very easy**

2. **Were there any particular obstacles that made it difficult for you to access the simulator?** *(Please select as many as apply)*
   - ○ Being in control group
   - ○ Work/duties and other commitments
   - ○ Wasn't given 'protected training time'
   - ○ Had difficulties in logging-in
   - ○ Had difficulties in completing tasks
   - ○ Kept failing tasks and lost interest
   - ○ Location of simulator not suitable
   - ○ Travel distance
   - ○ Not interested
   - ○ Other *(please specify)*: _____

3. **How often did you gain access to the ScanTrainer for ultrasound training?**
   - o 1–2 times/week
   - o 1–2 times/ month
   - o Once / 3 months
   - o Less than ONE access / 6 months
   - o Never
4. **Please state the overall time you usually spend on *"un-interrupted practice"* with the ScanTrainer simulator in a single session.**
   - o Less than half an hour (< 0–30 minutes)
   - o Less than an hour (31–60 minutes)
   - o More than an hour (> 60 minutes)
   - o Never

5. **How often were you assessed by the PhD student during the trial?**
   - o Every 4 weeks
   - o Every 4–6 weeks
   - o Every 6–8 weeks
   - o Every 2–3 months
   - o Every >3 months

6. **Six months assessment as a length of time to gain competence in the core skills in ultrasound are:**
   - o Too long
   - o Too short
   - o Adequate
   - o Other *(please specify)*: _____

7. **Did you receive feedback from the PhD student after the monthly assessment?**
   - o Yes
   - o No

   If yes, how useful did you find the feedback? *(Please mark a cross at an appropriate point on the scale below)*

   Not at all useful [0] _____ **[10] Very useful**

8. **Do you think that the PhD student's feedback was helpful in enhancing your understanding towards the simulator?**
   - o Very helpful
   - o Somewhat helpful
   - o Not helpful at all
9. **How useful do you think it is to have the physical presence of an instructor available with the simulator throughout your training?**
   - o Very helpful
   - o Somewhat helpful
   - o Not helpful at all
10. **How many times did you need to approach the Trainer/instructor to help you during the simulation session?**
    - o I needed help every session I had.
    - o I needed help only once
    - o I didn't need help at all.

11. **Specify your status with the PhD trial?**
    - o Withdrawn
    - o Completed the participation

12. **Regardless of your status as participant in the trial (completed or withdrawn), have you used the simulator since then (after completion or withdrawal)?**
    - o Yes
    - o No

13. **Do you intend to use the simulator in the future?**
    - Yes *(please go to Qestion15)*
    - No
14. **If not, please specify why you don't intend to use the simulator in the future:**
    - Hospital doesn't own an ultrasound simulator
    - I am not authorised to use the simulator
    - Work commitments
    - Simulator is not realism
    - Not interested *(please go to Question17)*
    - Others _____

15. **What would have improved your experience of using the simulator?**

_____

_____

16. **If you have completed the trial, are you interesting in receiving a summary report about your progress during this trial?**
    - Yes
    - No
17. **Are you interesting in being informed of the trial's result?**
    - Yes
    - No

Part 4 – General statements about ultrasound simulator: (Quality of service given by the simulator)

Please rate your opinion about your performance with ScanTrainer Ultrasound Simulator:

Statements	Rate your satisfaction [0]----------[10] not satisfied satisfied
1. The simulator provides easy access to practise endovaginal ultrasound scanning	[0]----------[10]
2. The simulator is excellent for training beginners in ultrasound scanning skills	[0]----------[10]
3. It is a good process for teaching and learning a systematic approach in scanning	[0]----------[10]
4. Training with the simulator helps in familiarising the trainee with core ultrasound skills (endovaginal scanning skills)	[0]----------[10]
5. The ultrasound image appeared realistic in the simulator	[0]----------[10]
6. The instructions and the 3D depiction of anatomy have a very useful role in understanding the ultrasound image and orientation	[0]----------[10]
7. All buttons on the simulator were handy and well explained	[0]----------[10]
8. Providing force feedback on the operator's hand helps in simulating real scanning	[0]----------[10]
9. Training with the simulator reflects a similar or very close experience to a patient/volunteer scan	[0]----------[10]
10. My confidence in scanning is enhanced after training with the simulator	[0]----------[10]
11. Training is better at simulation training centres	[0]----------[10]
12. Training is better at clinical stations e.g. ANC, Labour ward	[0]----------[10]
13. The simulator helps overcome the current shortage of learning capacity in hospitals and training centres	[0]----------[10]

----End of questionnaire----
Thank you for participation

Project title: Ultrasound simulation training and its impact on clinical skills development

Nature of project: Research

Tutor: Prof. Nazar Amso

Name: Grace Langan

Student number: 0901000

Word count: 1604

Transvaginal ultrasound simulation training and its impact on clinical skills development

Abstract

Objective: To determine the benefits and limitations of ultrasound simulation training as perceived by trainees.

Method: A questionnaire was distributed to individuals participating in an on-going RCT investigating TVUS simulation training. Responses were obtained via e-mail, telephone and paper copies from 15 out of the 26 possible participants. Name, age, gender and clinical occupation were recorded, as were the trainees own experiences of and opinions concerning their training in TVUS.

Results: Qualitative data was collected and analysed in order to establish the general opinion of the participants with regards to their use of the TVUS simulator.

Conclusion: Simulation training in TVUS is a useful addition to any other training received on the subject, and is particularly effective at improving basic clinical skills to a competent level.

Introduction

Transvaginal ultrasound is used gynaecologically to assess the female pelvic organs and also in obstetric examination during the first trimester of pregnancy[1,2,3]. It is an essential procedure for any trainees in obstetrics and gynaecology to become familiar with and competent at, and is currently taught via several methods. These include clinical sessions with a trainer, the utilisation of mannequins[4], and practice with simulators[5].

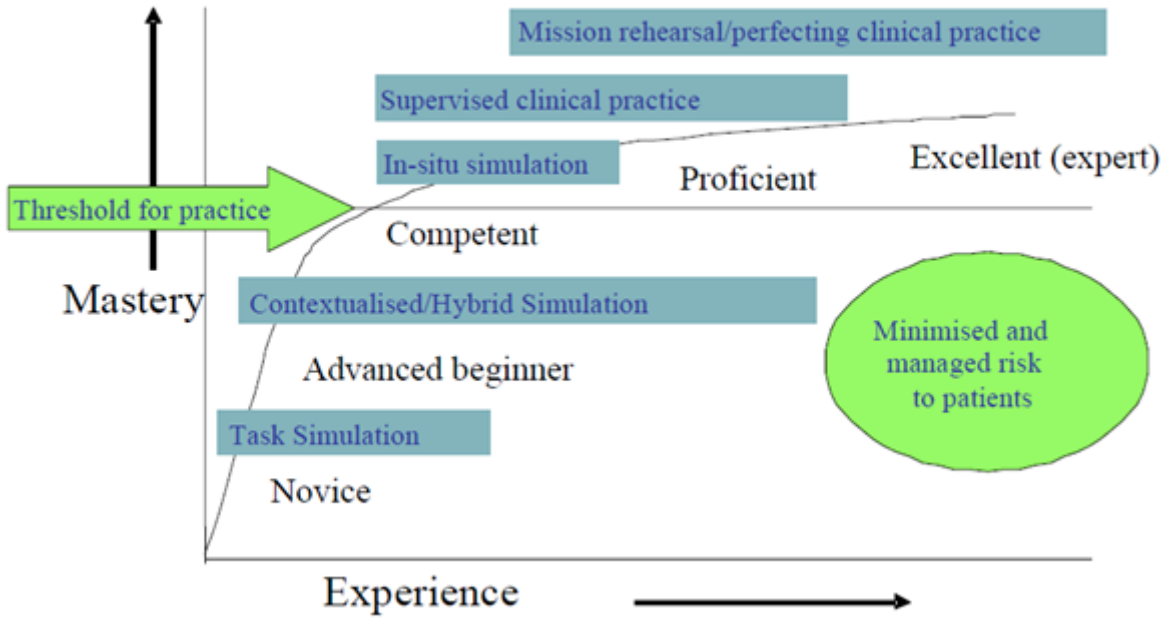
According to a recent report released by the Department of Health, "technology has an important role to play in the continuum of managed learning processes"[6]. The technique of combining a variety of different educational methods is commonly referred to as "blended learning"[6]. It can involve simulation, face to face teaching and e-learning[6,7,8,9].

Simulation is a technique whereby an environment is created that allows a virtual representation of a real process to be experienced[6]. It can be used for learning and assessing competence in this process, and has many benefits in terms of time and cost effectiveness[6,8].

Simulation can result in skills being acquired in a more efficient and comprehensive manner, but is best utilised with other clinical learning methods to achieve a blended approach[6,10].

As shown in the Figure 1, simulation training is particularly suited to improving skills to a basic level until the trainee is competent and confident enough to practice clinically. As experience and capability increases, there is minimised risk to the patient[11].

Figure 1: A simulation-enhanced learning trajectory[6]



The use of TVUS simulation training theoretically has several benefits. As it does not require the presence of a qualified trainer and a real patient, it reduces the cost of training in terms of both time and money[6,12]. It can be accessed at trainees' convenience and is intended to improve their skills and technique prior to patient contact, resulting in a better outcome for both the patient and the trainee[6,13].

The ScanTrainer ultrasound simulator was designed and created in Cardiff[14]. It aims to "provide curriculum-based training in a pre-clinical setting".[14] It utilises haptic feedback to provide a real feel for the user and employs actual ultrasound images in order to try and make the simulation as realistic and effective as possible.[14]

Objective

This report aims to establish the benefits and limitations of TVUS simulation training as perceived by trainees.

Method

A questionnaire was distributed to individuals participating in an on-going RCT investigating the speed of skill acquisition in TVUS with and without the use of the simulator. Prior to the trial, all had received very limited training in TVUS.

The participants were contacted via telephone and e-mail, and the sample to be analysed was determined to include all those who completed the questionnaire. Responses were obtained in the form of paper copies, e-documents returned via e-mail, and telephone interviews. The questionnaire was comprised of three sections, two of which are examined in this report.

There are 3 centres where the trainees could access the TVUS simulator – Cardiff (University Hospital of Wales and the Medicentre), Wrexham Maelor Hospital and Singleton Hospital.

Out of the 26 possible individuals, responses were collected from 15. Information regarding their completion status within the RCT and their designation in either the control or intervention groups was provided beforehand from the student running the RCT.

In the first section of the questionnaire, the participants were asked to state their name, gender, age and clinical position.

The second part comprised of 7 questions in both tick-box and open answer formats. In the open answer questions, any previous training was established, along with benefits and limitations of the simulator, and any aspects of their training that they felt it was good or bad at assessing. It also asked them to describe any perceived differences between the simulator and other methods of TVUS training in terms of advantages and disadvantages. Three tick-box questions were present, concerning the general benefit of the simulator, its comparative effectiveness in contrast to other methods, and whether the simulator was a useful and helpful addition to their clinical training overall.

In the last section, which is not be analysed here, the questions were directed at discovering any barriers to access that the trainees had regarding their use of the TVUS simulator.

The full questionnaire can be found in the Appendix.

Results

Of the 15 individuals who answered the questionnaire, 2 were in the control group and the remainder belonged to the intervention arm of the aforementioned RCT. Regarding their current status within the trial at the time of the questionnaire, 2 had withdrawn from it, 5 were still in progress and 8 had fully completed the project. Most of the responses were obtained via a phone interview (7 in total), and of the others 3 completed and returned a paper copy and 4 returned the form via e-mail. All of the data analysed in this report is of a qualitative nature.

A number of ages were given for the trainees to select, and the results ranged from 24 to over 51 years old.

Full details of the responses to the questionnaire can be found in the results table below.

Table 1: Full list of questionnaire responses

				Part 1			Part 2	
No.	**Group**	**Status**	**Form**	**Q2.**	**Q3.**	**Q4.**	**Q1.**	**Q2.**
1	Intervention	In progress	Phone	24-30	M	Specialist trainee	None	Yes
2	Control	Withdrawn	Phone	31-40	F	Specialist trainee	Few courses	Yes
3	Intervention	Withdrawn	Phone	41-50	M	Specialist trainee	None	Yes
4	Intervention	Completed	Phone	31-40	M	Specialist trainee	Clinical sessions with trainer but didn't actually practise on patient by self	Yes
5	Intervention	Completed	Phone	51+	F	Specialist trainee	None	Yes
6	Intervention	Completed	Phone	51+	F	Associate specialist	RCOG US module and practised on patients	Yes
7	Intervention	In progress	Phone	24-30	F	Specialist trainee	None	Yes
8	Intervention	Completed	Phone	24-30	F	Postgrad US student and radiologist	Clinical sessions with trainer and patients	Yes
9	Intervention	In progress	Paper	24-30	F	Postgrad US student	None	Yes
10	Intervention	Completed	E-mail	31-40	F	Senior clinician	None	Yes
11	Intervention	In progress	E-mail	31-40	F	Specialist trainee	Formal TVUS training	Yes
12	Intervention	Completed	E-mail	41-50	F	Specialty doctor	Clinical sessions with trainer and patients	Yes
13	Intervention	Completed	E-mail	41-50	M	Postgrad US student	Yes	Yes
14	Control	Completed	Paper	41-50	F	Clinical fellow/honorary lecturer	Practical US sessions once a week	Yes
15	Intervention	In progress	Paper	31-40	F	Specialist trainee	Course many years ago	No

No.	**Q3a.**	**Q3b.**
1	Only opportunity to practise. Instant feedback at end of each module	Only transvaginal - no transabdominal aspect
2	Good for beginners. Mimics actual US machine in parameters provided	Not complete package unless with a clinical session
3	Convenient. Precision in areas previously unsure about. Good feedback	None
4	Visual is present by side of the scan on the screen	Very rigid marking gets irritating eg. Can't deviate from planes
5	Good for development of hand-eye coordination, knowledge of US anatomy, understanding of orientation and basic controls of US machine	Very fine tolerance for passing certain skills. Time to access simulator
6	Can practise as much as you want. Basic experience	Irritating when a few mm away from being correct but simulator says you are incorrect. Should be able to see iliac vessels to enable location of ovaries
7	Can familiarise self with equipment in a stress-free environment. Nice to practise pressure that can be used without hurting the patient	Access for research students is difficult
8	Very good at beginning for learning layout of organs and manipulating the probe	Not very realistic as don't have to set up machine
9	Very good for basic skills if never had training before	Not real. Straightforward scan without any difficulties
10	Only way of obtaining training with current resources of limited training with sonographers	Only beneficial if receive feedback
11	Improves skills with systematic approach to scanning. User-friendly. Can spend more time practising as compared to when with a patient	Not easily accessible. Feedback not explanatory enough and has no references to reading material. Some settings are not similar to those on an actual US machine
12	Builds confidence in TVUS	None
13	Aids in learning and mastering pelvic US quickly. Very close to real-time pelvic scanning	Exclusion of common pelvic pathologies
14	N/A	Was in control group initially, so was not allowed full assessment with feedback
15	Opportunity to practice in non-clinical situation away from the pressures of clinic	Same scenarios all the time. Always tested on the same "patient" so know where things are without having to look for them. Difficult to access simulator

No.	Q4.	Q5.	Q6.	Q7.
1	None	N/A	N/A	Yes
2	None	Worse	Feel is not exactly the same as in a real patient	Yes
3	None	N/A	N/A	N/A
4	None	Equivalent	N/A	Yes
5	Good at developing systematic approach as had to go through individual steps in order to pass	N/A	N/A	Yes
6	Good for basic skills. Bad for advanced skills	Better (for basic skills)	N/A	Yes
7	Not very good for measurements, accuracy of labelling and zooming in. Good for learning different components of each scan in O&G	N/A	N/A	Yes
8	None	Worse	Reality	Yes
9	None	Better	Not noticed	Yes
10	This was the only training received	N/A	N/A	Yes
11	Best aspect is improving with each time task is undertaken. Each assessment doesn't have feedback in order to not repeat mistake, therefore there is very little interactive teaching	Better than mannequins. Worse than clinical training sessions	Clinical sessions are more interactive with instant feedback from skilled people. There were more opportunities to practice with the simulator	Yes
12	After half of sessions felt more confident to use simulator. PhD Student was useful to correct mistakes	Better	Yes - can come in own time. Good access to venue and opportunities to book sessions. Car parking available	Yes
13	Good at demonstrating the pelvic US sectional anatomy	Equivalent	Mannequins give a virtual impression of real life scan, therefore boost student's confidence compared to simulator, which appears more like playing a video game	Yes
14	Beneficial for learning systematic approach for scanning	Equivalent	Simulator has evaluation and results with feedback on performance, compared to mannequins and brief feedback from trainer in clinical sessions	Yes
15	Very rigid on measurement e.g. "out by 0.4mm". Does this make a difference in clinical practice?	Worse	Regularity	No

Discussion

The most common age group amongst the participants was that of 31 to 40 years. 4 of the individuals were male and 11 were female. The most common occupation was “specialist trainee”, and others included “senior clinician”, and “postgraduate ultrasound student”.

Firstly, in Part 2 Q1, the participants were asked if they had received any previous or concurrent training in TVUS other than their use of the simulator. 40% had none and 20% have had clinical sessions with trainers and patients present.

Q2 asked if, in general, they thought that they had benefited from use of the simulator in addition to any other training received, 93% answered “yes”.

The next question (Q3) enquired as to whether the trainees felt the simulator had any particular benefits or limitations in regards to their development of competency in the core skills of TVUS. 93% described benefits of the simulator, and the most common was that of the development of basic skills. 87% of the participants stated specific limitations. The 3 most common limitations described were that the simulator was not easily accessible, non-realistic, and that the rigidity of marking used by the program was frustrating. Other particular benefits and limitations commented upon can be found detailed in Table 1.

In Q4, the participants were asked if there were any aspects of their training that they felt the simulator was particularly good or bad at teaching or assessing, and 60% responded in the affirmative. The positive and negative aspects described are in Table 1.

Q5 gave 3 options and asked the trainees to choose how they felt the simulator compared to other methods of training in terms of its effectiveness. 20% said that the simulator was "better", 20% stated that it was "worse" and 20% felt that the simulator was "equivalent" to other methods of ultrasound simulation. One participant thought that it was better than mannequins but worse than clinical training sessions. The remaining trainees did not answer.

In Q6, participants were asked to describe any differences that they had noticed in terms of advantages and disadvantages between the simulator and other forms of training. 47% did not answer, and the responses of those who did can be found in Table 1.

Finally, Q7 asked if overall they felt that utilisation of the simulator was a useful and helpful addition to their clinical training. 87% of the trainees answered "yes". One individual said "no", as they could not access it enough for their own personal needs, and one did not answer.

The results of this study suggest that the addition of TVUS simulation training to current teaching in obstetrics and gynaecology would be useful.

Simulation training has been used in other specialties with much success and has proven to be efficacious when practised alongside other teaching methods also[15, 16, 17, 18].

However, as the RCT that the sample population was drawn from is only very small, any implications of this study would need further investigation to make any firm conclusions on the topic. A greater number of participants and a higher response rate would have made the results more reliable, but perhaps there will be opportunities for this to occur in future research. Ideally, a nationwide randomised control trial would be performed to determine the speed of TVUS skill acquisition with and without use of the simulator, and the participants questioned to analyse its efficacy.

Conclusion

The results show that TVUS simulation training has specific benefits and limitations, but is overall considered to be a worthwhile addition to the training methods used for this technique. It is a valuable resource in the acquisition of clinical skills, and would be best utilised with other educational techniques such as mannequins and actual clinical practice on patients in order to achieve a blended learning approach. When combined, these methods should lead to greater competency and confidence in the performance of TVUS by trainees, and therefore a better outcome for patients.

References

1. Nasri MN, Setchell ME, Chard T. 1990. Transvaginal ultrasound for diagnosis of uternine malformations. *British Journal of Obstetrics and Gynaecology.* 97. Pp. 1043-1045.
2. Haylen BT, Frazer MI, Sutherst JR, West CR. 1989. Transvaginal Ultrasound in the Assessment of Bladder Volumes in Women: Preliminary Report. *British Journal of Urology*. 63. Pp. 149-151.
3. Nasri MN, Shepherd JH, Setchell ME, Lowe DG, Chard T. 1991. The role of vaginal scan in measurement of endometrial thickness in postmenopausal women. *British Journal of Obstetrics and Gynaecology*. 98. Pp. 470-475.
4. Cooper JB, Taqueti VR. 2008. A brief history of the development of mannequin simulators for clinical education and training. *Postgraduate Medical Journal.* 84. Pp. 563-570.
5. Ennen CS, Satin AJ. 2010. Training and assessment in obstetrics: the role of simulation. *Clinical Obstetrics and Gynaecology.* 24. Pp. 747-758.
6. Department of Health. 2011. A framework for technology-enhanced learning. Available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131061.pdf [Accessed 29 June 2012].
7. Kirkley SE, Kirkley JR. 2004. Creating next generation blended learning environments using mixed reality, video games and simulations. *TechTrends.* 49. Pp. 42-53.
8. Ginns P, Ellis R. 2007. Quality in blended learning: exploring the relationships between on-line and face-to-face teaching and learning. *The Internet and Higher Education.* 10. Pp. 53-64.
9. Psotka J. 1995. Immersive training systems: Virtual reality and education and training. *Instructional Science.* 23. Pp. 405-431.
10. Ruiz JG, Mintzer MJ, Leipzig RM. 2006. The impact of e-learning in medical education. *Academic Medicine.* 81. Pp. 207-212.
11. Kunkler K. 2006. The role of medical simulation: an overview. *International Journal of Medical Robotics and Computer Assisted Surgery.*2. Pp. 203-210.
12. Ennen CS, Satin AJ. 2010. Training and assessment in obstetrics: the role of simulation. *Clinical Obstetrics and Gynaecology.* 24. Pp. 747-758.
13. Sidhu HS, Olubaniyi BO, Bhatnagar G, Shuen V, Dubbins P. 2012. Role of simulation-based education in ultrasound practice training. *Journal of Ultrasound in Medicine.* 31. Pp. 785-791.
14. ScanTrainer. MedaPhor Ltd. Cardiff UK. 2011 Available at: http://www.scantrainer.com [Accessed: 30 June 2012].
15. Haque S. 2006. A meta-analysis of the training effectiveness of virtual reality surgical simulators. *Information Technology in Biomedicine*. 10. Pp. 51-58.

16. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. 2008. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *British Journal of Surgery.* 95. Pp. 1088-1097.
17. Lee Gordon D, Issenberg SB, Gordon MS, LaCombe D, McGaghie WC, Petrusa ER. 2005. Stroke training of prehospital providers: an example of simulation-enhanced blended learning and evaluation. *Medical Teacher.* 27. Pp. 114-121.
18. Reznek M, Harter P, Krummel T. 2002. Virtual reality and simulation: training the future emergency physician. *Academic Emergency Medicine.* 9. Pp. 78-87.

Appendix: Full questionnaire distributed to participants

Part 1 – General Information

1. **Name: _____**

2. **Age:**
   - **18-23**
   - **24-30**
   - **31-40**
   - **41-50**
   - **51+**

3. **Gender:**
   - **Male**
   - **Female**

4. **Position:**
   - **Consultant/Senior clinician**
   - **Specialist trainee**
   - **Postgraduate US student**
   - **Midwife/Nurse**
   - **Academic**
   - **Other (please specify):**

     _____

     _____

Part 2 – Perceived advantages and disadvantages of the simulator

1. **Have you received any other training in transvaginal ultrasound (TVUS)? If so, please describe below.**

_____
_____
_____
_____

2. **In general, do you think that you have benefited from the use of the simulator, in addition to any other training that you have received?**
   - **Yes**
   - **No**

3. **Do you feel that the simulator has any particular benefits or limitations in regards to your development of competency in the core skills of TVUS? If so then please describe below.**

   - **Benefits:**

_____
_____
_____
_____

   - **Limitations:**

_____
_____
_____
_____

4. **Were there any aspects of your training that you felt the simulator was particularly good or bad at teaching or assessing?**

_____
_____
_____
_____

5. **In contrast to other methods of training (e.g. mannequins with an ultrasound machine, without a trainer, or clinical training sessions with patients and trainer), how do you feel the simulator compared in terms of its effectiveness?**
   - ☐ **Better**
   - ☐ **Equivalent**
   - ☐ **Worse**

6. **Are there any particular differences that you have noticed between the simulator and other forms of training in terms of their advantages and disadvantages?**

_____
_____
_____
_____

7. **Overall, do you feel that utilisation of the simulator was a useful and helpful addition to your clinical training?**
   - ☐ **Yes**
   - ☐ **No**

Part 3 – Barriers to engagement with the simulator

1. Where were you able to access a simulator? (Select as many as apply.)
   - € Labour ward at UHW
   - € Cardiff Medicentre
   - € Swansea
   - € Wrexham

2. How easy did you find it was to access the simulator? (Please mark a cross at an appropriate point on the scale below.)

   [0]——————————————————————————[10]———

   Very difficult Very easy

3. Were there any particular obstacles that made it difficult for you to access the simulator? (Please select as many as apply.)
   - € Being in control group
   - € Work/duties and other commitments
   - € Wasn't given 'protected training time'
   - € Had difficulties in logging-in
   - € Location of simulator is not suitable
   - € Travel distance
   - € Not interested
   - € Other (please specify):_____
     _____

4. How often did you gain access to the ScanTrainer for ultrasound training?
   - € 1-2 times/week
   - € 1-2 times/ month
   - € Once / 3 months
   - € less than ONE access / 6 months
   - € Never

5. Please state the overall time you usually spend on *"un-interrupted practice"* with the ScanTrainer simulator in a single session.
   - € Less than half an hour (< 0- 30 minutes)
   - € Less than an hour (31- 60 minutes)
   - € More than an hour (> 60 minutes)
   - € Never

6. How often were you assessed by the PhD student?
   - € Every 4 weeks
   - € Every 4-6 weeks
   - € Every 6-8 weeks
   - € Every 2 -3 months
   - € Every >3 months

7. Did you receive feedback after the assessment?
   - € Yes (please answer question a)
   - € No (please answer question b)

   a) If Yes how useful did you find the feedback? (Please mark a cross at an appropriate point on the scale below.)

   [0]——————————————————————————[10]

   Not at all useful Very useful

   b) If No would you like to have received feedback?
      - € Yes
      - € No

8. Have you completed the programme?
   - € Yes
   - € No

9. If not, are you still participating?
   - € Yes
   - € No

10. If not, why not?

_____
_____
_____
_____

11. How useful did you find the simulator was useful when you used it? (Please mark a cross at an appropriate point on the scale below.)

   [0]——————————————————————————[10]

   Not at all useful Very useful

12. What would have improved your experience of using the simulator?

_____
_____
_____
_____

Ultrasound Simulation training and its impact on clinical skills development – barriers to access

Word Count: 1491

Sophie Mullins

Student Number: 0902973

Tutor: Prof Nazar Amso

Other Tutors: Miss Amal Al-Salama/Mrs Helena Clarke

Institution: University Hospital of Wales

Ethical Approval Gained from South East Wales Research Ethics Committee Panel B (dated 14 December 2010) as part of the work of PhD student Amal Al-Salamah

Ultrasound Simulation training and its impact on clinical skills development barriers to access

Ultrasound Simulation Training and its Impact on Clinical Skills Development – Barriers to Access

Abstract

Objective: To evaluate the trainees' views on the barriers to and reasons for engagement with the ultrasound simulator (ScanTrainer) for training in core transvaginal ultrasound (TVUS) skills.

Design: Retrospective questionnaires in paper, electronic or phone interview format were analysed qualitatively and quantitatively.

Setting: University Hospital of Wales, with participants from Singleton Hospital and Wrexham Maelor Hospital

Participants: Participants in the randomised controlled trial (RCT) investigating the benefit of the simulator including those who have withdrawn.

Results: The main barriers to access were conflict of commitments, lack of protected training time and location of the simulator. There were no statistically significant difference between the access and perceived ease of access between the centres or different levels of training.

Conclusions: It is likely that the most effective method for improving utilisation of the simulator would be to implement protected training time, however there are staffing and time constraints to this.

Introduction

Clinical training in trans-vaginal ultrasound (TVUS) costs clinical time and money, and may cause the patient discomfort and embarrassment. Alternatives to this have been proposed to aid the acquisition of the core skills in TVUS necessary for obstetric and gynaecological examination: such as mannequins[1], which, however do not offer any feedback on the pressure applied of TVUS; or practising ultrasound on those undergoing operations when they are under general anaesthetic,[2] which could lead to an increase in theatre time and thus cost, and there would be no feedback on the discomfort caused to the patient. Assessment of TVUS skills includes not causing undue discomfort to the patient,[3] therefore mannequins and practise under general anaesthesia are insufficient methods alone.

Blended learning is the use of a combination of different learning approaches, typically combining face-to-face instruction and technology.[4] The use of blended learning has been shown to be effective in other areas of clinical training[5,6].The Department of Health is encouraging the use of simulators to enable trainees to reach a level of safety and proficiency, prior to training on patients, as demonstrated in Fig. 1.[7] An ultrasound simulator (ScanTrainer) has been developed to try to fulfil this and aid the acquisition of the core skills of TVUS in a situation requiring less patient contact and less supervision.[8] The ScanTrainer has modules corresponding to the Royal Society of Obstetrics and Gynaecology curriculum

and gives immediate feedback. It uses a haptic feedback device, and a touch-screen displaying virtual anatomy, ultrasound image and controls similar to an ultrasound machine.[8]

At present the simulator is being evaluated in a randomised controlled trial (RCT) that is assessing for any difference in the rate of acquisition of core TVUS skills between those receiving standard clinical training and simulator training, and a control group receiving clinical training alone.[9] This project was undertaken as part of this evaluation.

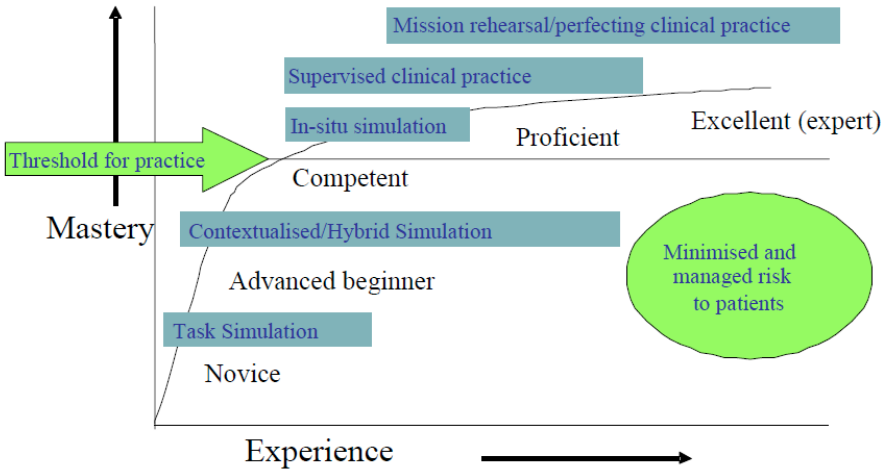


Fig. 1The use of simulation in training of clinical skills, with learning trajectory. Taken from Department of Health (2011).[7]

Aims

This study was undertaken to evaluate the participants' views on the barrier to and reasons for engagement with the ultrasound simulator (ScanTrainer) for training in the TVUS skills required in obstetrics and gynaecology, by means of questionnaire.

Methods

Questionnaires (paper, electronic or phone interview format) were distributed to assess the barriers to and reasons for engagement with the ultrasound simulator (ScanTrainer) for training in the TVUS skills required in obstetrics and gynaecology. The questionnaire covered the following topics: generic information, where the trainee accessed the simulator, frequency and ease of access and perceived benefit of the simulator. To reduce inconvenience to the participants the questionnaire was combined with another that assessed the benefits and limitations of the simulator from the perspective of the trainee, however topics covered were analysed separately. Questions were in multiple choice, global rating scale and open answer format. [See Appendix for a copy of the questionnaire (questions analysed in bold)].

The sample group of 26 was identified as the subjects in the RCT that is evaluating the simulator. Prior to partaking in this RCT the trainees had no prior TVUS experience. The sample included those who withdrew from the study or were lost to follow up, but were still willing to complete a questionnaire.

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

Where possible participants were contacted by both e-mail and phone. They were given the option of completing a paper, electronic or interview format version of the questionnaires. If a paper or electronic copy was chosen participants were contacted to ensure they had received the questionnaire.

Quantitative results were analysed using SPSS 16.0, and Chi-squared and Kruskal-Wallis tests. Qualitative results were reviewed for patterns.

Ethical approval received from SEW REC B* as part of the PhD project this research contributes to.

Results

Of the 26 initial participants in the RCT, 15 responses were received (see Fig. 1 for more detail). 3 participants accessed the ScanTrainer in Singleton Hospital, 4 in Wrexham Maelor Hospital, 1 in University Hospital of Wales (UHW) and 6 in Cardiff Medicentre. 1 participant had access in both UHW and the Cardiff Medicentre.

Of those who responded 3(20%) accessed the simulator 1-2 times/week, 7(47%) 1-2 times/month, 2(13%) once every 3 months, 2(13%) less than once every 6 months and 1(7%) never accessed the simulator. The median frequency of access was 1-2 times per month.

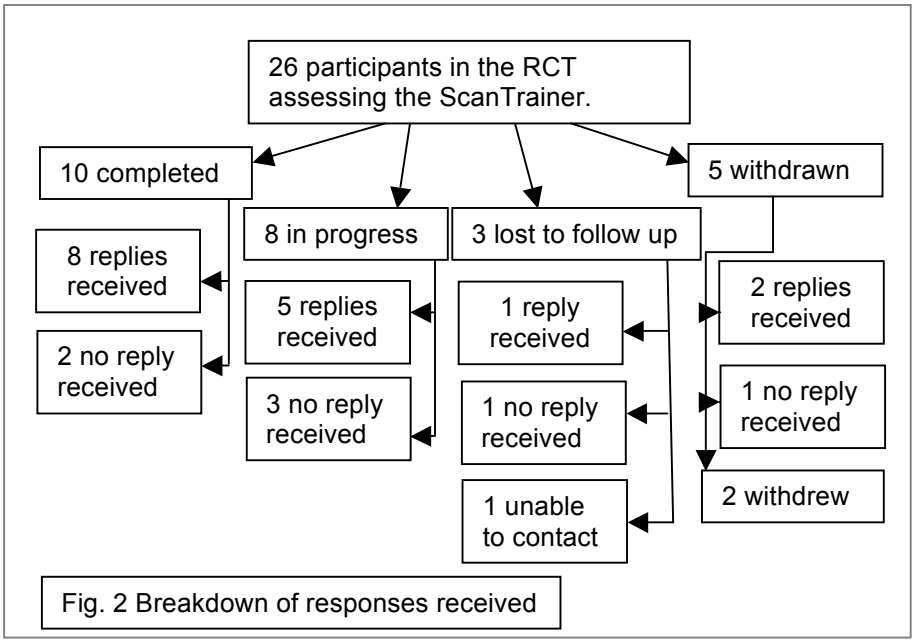


Fig. 2 Breakdown of responses received

When asked about specific barriers to training 10 participants stated that work duties and other commitments were a barrier, 5 felt that not having 'protected training time' was an issue, 3 thought the location was not appropriate, 2 stated that travel distance was a problem and 1 began the project in the control group, which limited their access to the simulator. Individual comments made about access were: other people using the simulator reduced accessibility; session length was limited by the haptic device overheating, even when using reduced force feedback; and 1 participant stated that they were not in clinical training at the time and thus found it more difficult to use the simulator around their timetable. 1 stated that adequate staffing to enable protected training time would improve access. However, 3 participants felt that there were no specific barriers to accessing the simulator.

* SEW REC B: South East Wales Research Ethics Committee Panel B, dated 14 December 2010

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

Upon statistical analysis of the global rating scales it was found that there was no statistically significant difference between the location and the ease of access score (p=0.794) and no statistically significant difference between the location and the frequency of engagement with the simulator (p=0.530). However, 2 participants stated that the labour ward would be a more appropriate location, as in one case the

Average (mean) ease of access to simulator score (0=very difficult, 10=very easy)	
Singleton Hospital	5.0
Wrexham Maelor Hospital	6.6
Labour Ward at UHW	4.4
Cardiff Medicentre	6.4
Overall	5.9

Frequency of access to scan trainer					
	1-2 times/ week	1-2 times/ month	Once every 3 months	Less than once every 6 months	Never
Singleton Hospital	1	1	0	1	0
Wrexham Maelor Hospital	1	1	1	0	1
Labour Ward at UHW	0	1	0	1	0
Cardiff Medicentre	1	5	1	0	0
Overall	3	7	2	2	1

Amount of time spent per session in "uninterrupted practice"					
	Less than half an hour (<30 minutes)	Less than an hour (31- 60 minutes)	More than an hour (>60 minutes)	Never	n/a
Singleton Hospital	0	1	1	1	0
Wrexham Maelor Hospital	1	0	2	0	1
Labour Ward at UHW	0	0	2	0	0
Cardiff Medicentre	0	2	5	0	0
Overall	1	3	9	0	1

Frequency of assessment by PhD student					
	Every 4 weeks	Every 4-6 weeks	Every 6-8 weeks	Every 2-3 months	More than 3 months between assessments
Singleton Hospital	0	2	0	1	0
Wrexham Maelor Hospital	0	2	0	1	0
Labour Ward at UHW	0	1	0	1	0
Cardiff Medicentre	4	3	0	0	0
Overall	4	7	0	3	0

Average(mean) perceived usefulness of feedback given after assessments by PhD student (0=not very useful, 10=very useful)	
Singleton Hospital	8.7
Wrexham Maelor Hospital	7.7
Labour Ward at UHW	7.8
Cardiff Medicentre	7.8
Overall	8.0

Average (mean) perceived usefulness of simulator(0=not very useful,10=very useful)	
Singleton Hospital	7.5
Wrexham Maelor Hospital	7.2

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

Labour Ward at UHW	8.4
Cardiff Medicentre	7.2
Overall	7.4

Fig. 3 Results by location and question

simulator was in a room that was locked out of hours and on the opposite side of the hospital to the obstetrics and gynaecology department.

Additionally there is no statistically significance difference between the perceived usefulness of and the frequency of engagement with the simulator (p=0.567). There was no statistically significant difference between frequency of engagement with the simulator and occupation (those who were practising medical professionals and those who were postgraduate ultrasound students) (p=0.753).

1 participant stated that “better access” would have improved their experience of the simulator.

Results are summarised in Fig 3.

Discussion

Qualitative comments of trainees and a median access of just 1-2 times per month show that access is insufficient. Therefore if proven to be effective in improving speed of acquisition of the core skills of TVUS and increasing patient safety by allowing this to occur prior to patient contact, it is essential to its success that any barriers to access are rectified.

Though the access at each centre was different, there was no statistically significant difference between the centres and either frequency of access or perceived ease of access. This would imply that the locations and their different access systems did not affect the engagement with the simulator, despite the location being reported as a barrier to access by 3 individuals and 2 stating that moving the simulator to the labour ward would aid engagement. However, the sample size is small, meaning the power of the results are reduced, so type 2 statistical errors are very likely. However, factors other than location must be looked into when considering the barriers to engaging with the simulator.

With 10 out of 15 participants (67%) stating it as a specific barrier to engagement with the simulator, work duties and other commitments was the largest barrier to access from the perspective of the

participant. The secondary perceived barrier was lack of protected training time. It is a possibility that correcting this may also reduce competition of commitments. However, as 1 participant stated there was insufficient staffing to allow for protected training time.

Conclusions

Simulation has been proven effective in many other areas of clinical training.[5,6] The median access to the simulator of 1-2 times per month is probably insufficient to rapidly learn the basic skills of TVUS that should preferably be learnt prior to patient contact,[7] as such there are barriers to receiving sufficient training on the simulator. From the responses received it is likely that the most effective method for improving utilisation of the simulator would be to implement protected training time. However, staffing and time constraints may make the implementation impossible.

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

Appendix

Dear Participant,

Our names are Grace Langan and Sophie Mullins and we are 3rd year medical students at Cardiff University. We are currently undergoing a project regarding the benefits, limitations and barriers to accessing the ScanTrainer from the perspective of those who have used it in their ultrasound training. This project is under the supervision of Prof. Nazar Amso and is contributing to the work of PhD student Miss Amal Alsalamah.

If possible, we would like to take a few minutes of your time to interview you in order to complete a questionnaire regarding your experience with the ScanTrainer.

Thank you for your time,

Grace Langan and Sophie Mullins

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

Part 1 – General Information

1. **Name:** _____

2. **Age:**
   - **18-23**
   - **24-30**
   - **31-40**
   - **41-50**
   - **51+**

3. **Gender:**
   - **Male**
   - **Female**

4. **Position:**
   - **Consultant/Senior clinician**
   - **Specialist trainee**
   - **Postgraduate US student**
   - **Midwife/Nurse**
   - **Academic**
   - **Other (please specify):** _____
     _____

Part 2 – Perceived advantages and disadvantages of the simulator

1. Have you received any other training in transvaginal ultrasound (TVUS)? If so, please describe below.

_____
_____
_____
_____

2. In general, do you think that you have benefited from the use of the simulator, in addition to any other training that you have received?
   - Yes
   - No

3. Do you feel that the simulator has any particular benefits or limitations in regards to your development of competency in the core skills of TVUS? If so then please describe below.

   - Benefits:

_____
_____
_____
_____

   - Limitations:

_____
_____
_____
_____

4. Were there any aspects of your training that you felt the simulator was particularly good or bad at teaching or assessing?

_____
_____

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

5. In contrast to other methods of training (e.g. mannequins with an ultrasound machine, without a trainer, or clinical training sessions with patients and trainer), how do you feel the simulator compared in terms of its effectiveness?
   - Better
   - Equivalent
   - Worse
6. Are there any particular differences that you have noticed between the simulator and other forms of training in terms of their advantages and disadvantages?

7. Overall, do you feel that utilisation of the simulator was a useful and helpful addition to your clinical training?
   - Yes
   - No

Part 3 – Barriers to engagement with the simulator

1. **Where were you able to access a simulator? (Select as many as apply.)**
   - **Labour ward at UHW**
   - **Cardiff Medicentre**
   - **Swansea**
   - **Wrexham**
2. **How easy did you find it was to access the simulator? (Please mark a cross at an appropriate point on the scale below.)**

   [0]——————————————————————— [10]
   Very difficult Very easy
3. **Were there any particular obstacles that made it difficult for you to access the simulator? (Please select as many as apply.)**
   - **Being in control group**
   - **Work/duties and other commitments**
   - **Wasn't given 'protected training time'**
   - **Had difficulties in logging-in**
   - **Location of simulator is not suitable**
   - **Travel distance**
   - **Not interested**
   - **Other (please specify):**
4. **How often did you gain access to the ScanTrainer for ultrasound training?**
   - **1-2 times/week**
   - **1-2 times/ month**
   - **Once / 3 months**
   - **less than ONE access / 6 months**
   - **Never**
5. **Please state the overall time you usually spend on *"un-interrupted practice"* with the ScanTrainer simulator in a single session.**
   - **Less than half an hour (< 0- 30 minutes)**

- € **Less than an hour (31- 60 minutes)**
- € **More than an hour (> 60 minutes)**
- € **Never**

6. **How often were you assessed by the PhD student?**
   - € **Every 4 weeks**
   - € **Every 4-6 weeks**
   - € **Every 6-8 weeks**
   - € **Every 2 -3 months**
   - € **Every >3 months**

7. **Did you receive feedback after the assessment?**
   - € **Yes (please answer question a)**
   - € **No (please answer question b)**

   a) **If Yes how useful did you find the feedback? (Please mark a cross at an appropriate point on the scale below.)**

   [0]——————————————————————— [10]
   Not at all useful **Very useful**

   b) **If No would you like to have received feedback?**
      - € **Yes**
      - € **No (please answer part a)**

8. **Have you completed the programme?**
   - € **Yes (please move on to question 9)**
   - € **No (please answer part a)**

   a. **If not, are you still participating?**
      - € **Yes (please move on to question 9)**
      - € **No (please answer part b)**

   b. **If not, why not?**

   _____
   _____
   _____
   _____

9. **How useful did you find the simulator was useful when you used it? (Please mark a cross at an appropriate point on the scale below.)**

   [0]——————————————————————— [10]
   Not at all useful **Very useful**

10. **What would have improved your experience of using the simulator?**

_____
_____
_____
_____

Ultrasound Simulation training and its impact on clinical skills development
barriers to access

References

[1] Cooper JB, Taqueti VR. A brief history of the development of mannequin simulators for clinical education and training. *Postgraduate Medical Journal*. 2008; 84: 563-570

[2]Shamsuddin L, Malone S, Chamberlain J. Teaching Trans-Vaginal Ultrasound under General Anaesthetic – The Patient's Perspective. *Presented at the British Society of Gynaecological Imaging 2012 Annual Scientific Meeting.* London, 2012

[3] Royal College of Obstetricians and Gynaecologists. OSATS - Generic Technical Skills Assessment. London: Royal College of Obstetricians and Gynaecologists; 2010 [accessed 26 Jun 2012]. Available from: http://www.rcog.org.uk/files/rcog-corp/uploaded-files/Ed-Ultrasound_Generic_Technical_Assessment.pdf

[4] Graham C. *Blended Learning Models*. In: Khosrow-Pour M. Encyclopedia of Information Science and Technology. 2nd Ed. New York: IGI Global; 2009. 375-382

[5] Wayne DB, Butter J, Siddall VJ, Fudala MJ, Wade LD, Feinglass J, McGaghie WC. Mastery learning of advanced cardiac life support skills by internal medicine residents using simulation technology and deliberate practice. *Journal of General Internal Medicine*. 2006; 21: 251–256.

[6] Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM. Virtual reality training improves operating room performance. *Annals of Surgery*. 2002; 236: 458–64.

[7] Department of Health. *A Framework for Technology Enhanced Learning,* 16787. London: Stationery Office; 2011.

[8] Medaphor. *What is ScanTrainer?* Cardiff: Medaphor [accessed 26 Jun 2012]. Available from: http://www.scantrainer.com/what-is-scan-trainer

[9] Al-Salamah A, *Validation and determination of the influence of a virtual simulator on the*

acquisition of ultrasound skills and comparison of learning curves of those

using simulation-supported training or a conventional training approach. PhD protocol, Cardiff University; 2010

Appendices

(9.1-9.2)

Peer training of transvaginal ultrasound scanning

Co-authors

Nazar Amso
Professor in Obstetrics & Gynaecology
University Hospital of Wales

Meena Murugan
Medical Student
School of Medicine
Cardiff University

Amal Al Salamah
PhD student
School of Medicine
Cardiff University

Abstract

Objective: To assess the effectiveness of peer education in the teaching of transvaginal ultrasound scanning amongst a group of medical students.

Design: A medical student was taught the basics of transvaginal ultrasound scanning using a mannequin and then went on to teach other medical students the same skill. Their improvement was monitored via weekly assessments over a time frame of five weeks and feedback was given on how to improve their scanning. A questionnaire was filled out by the students at the end of the study to see their views.

Participants: eight third year medical students from Cardiff University doing 9 week placements in the department of Obstetrics and Gynaecology.

Main outcome measures: mark sheet devised by consultant to assess weekly improvement. Questionnaire devised by the medical student peer tutor to look into the students thoughts on peer training.

Main results: the results showed that all the medical students improved weekly through the weekly feedback and assessment sessions. The students themselves thought that receiving teaching from a peer was useful and benefits included the sessions being informal and the students feeling less pressured than when in front of a senior member of staff.

Conclusions: peer education is an effective way of teaching practical techniques to medical students. In the future it could be considered as a method of learning transvaginal ultrasound scanning and could have a place in medical school when students learn how to perform clinical examinations.

Introduction

The project entails peer training of transvaginal ultrasound (TVUS) scanning to other medical students undertaking placements in the department of Obstetrics and Gynaecology.
TVUS along with transabdominal ultrasound are types of gynaecological ultrasonography used to assess the female pelvic organs, especially the uterus, ovaries and Fallopian tubes as well as the bladder, adnexa and pouch of douglas. TVUS in particular refers to the application of an ultrasound probe into the vagina and compared to transabdominal imaging utilizes a

higher frequency giving a better resolution[1]. Its scope and importance is of huge value in diagnosing and managing gynaecological pathology such as cancer, endometriosis, leiomyoma, adenomyosis and ovarian cysts, and in identifying adnexal masses including ectopic pregnancies. It is also used extensively in fertility treatments to track the response of ovarian follicles to fertility medication[2].

According to the WHO peer education is "a process whereby well trained and motivated young people undertake informal or organized educational activities with their peers (as defined by age, background or interests) over a period of time, aimed at developing their knowledge, attitudes, beliefs and skills and enabling them to protect and be responsible for their own health"[3].

In an article written by Gopee at Coventry University about the role of peer assessment and peer review in nursing, it was stated that "although self-assessment provides each individual with a medium for ascertaining his/her own level of performance and, therefore, identifying his/her learning needs, peer review and peer assessment provide healthy means for obtaining feedback and external perceptions"[4]. The benefits of peer education are widely recognised and peer education has a strong emphasis on personal development and can be particularly effective in allowing low achieving students to fully participate and succeed in a wider range of educational and health promoting activities[5].

A literature search was carried out and found that little research into the effectiveness of peer teaching of ultrasound has been done as all the work is relatively new. However there has been a lot of interest in peer education in general and one area where peer education has been found to be of high value is in promoting healthy behaviour for example in regards to sexuality, violence and substance abuse. A study by Sloane and Zimmer found that people are more likely to believe messages, therefore changing their attitudes and behaviours, if they believe the messenger is similar to them and faces the same concerns and pressures as they do[6]. As a result peer education draws on the credibility that young people have with their peers and leverages the power of role modelling making it a very effective way of teaching[7]. The WHO concurs by saying youth peer educators are less likely to be seen as authority figures "preaching" about how others should behave and instead the process of peer education is perceived more like receiving advice from a friend "who is in the know"[3].

A study by Cheeseman, Clack et al looked into the feasibility of medical students being involved in sex education in secondary schools by adopting the role of lesson leaders in peer group discussions. The results found that 94% of teachers and 93% of pupils were in favour of medical student involvement in schools. Furthermore pupils identified teacher embarrassment in certain issues as a barrier to communication, and 89% found communication to be easier with medical students[8]. Consequently this is further evidence that peer education is effective.

Regarding the methods of ultrasound teaching, there are 4 main ways:

1. no formal teaching; the learner simply observes and then performs
2. completion of a programme run by the RCOG through supervised teaching or an apprenticeship
3. attendance of a short course
4. enrolment on a higher postgraduate education programme achieving a degree such as Certificate, Diploma or Masters in a field of ultrasound

Therefore at present peer training is not used as a method of teaching transvaginal ultrasound scanning.

The main aims of the project are to:

1. Train a medical student to carry out a TVUS scan using a mannequin and act as a peer tutor
2. Teach other medical students the skill
3. Assess the students at regular intervals to see how they improve with peer teaching and feedback

This will help evaluate whether peer education is an effective way of teaching TVUS scanning. It was predicted that the students' ability to perform a TVUS scan will improve weekly with the peer teaching.

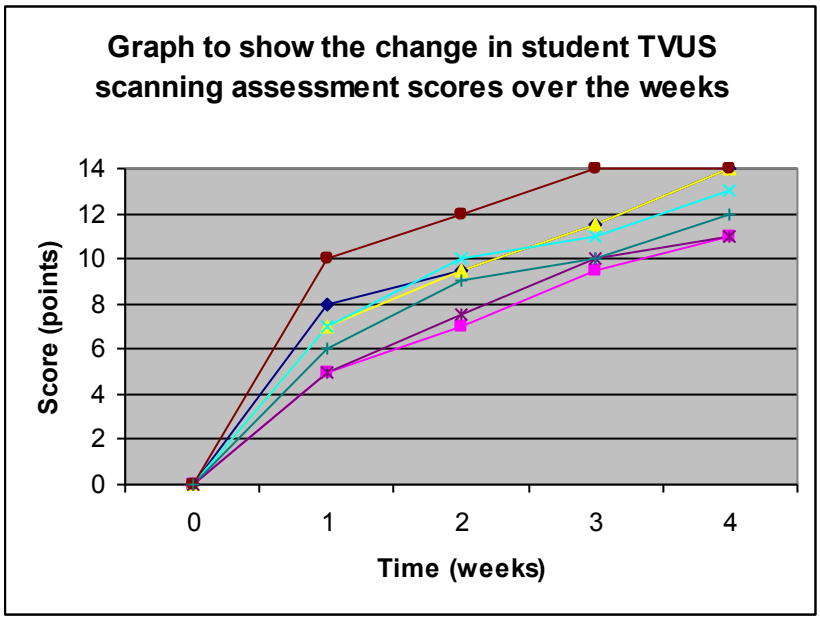
Methods

For the first three weeks of the study a medical student (who later become the peer tutor) was taught the basics of TVUS examination using a mannequin resembling the female anatomy. The student observed a masters student scanning the mannequin and was introduced to the probe and image on the screen. She was taught the importance of checking for the marker on the probe to orientate myself as to where anterior is and learnt about the different echogenicities on the screen: black areas being anechoic, grey hypoechoic and white hyperechoic. The use the ultrasound machine was taught as well as how to optimise the image on the screen by using the overall image gain, time gain, focus and depth facilities. On inserting the probe into the model vagina the student familiarised herself with the appearance of the uterus with its hyperechoic endometrium and serosa and hypoechoic myometrium. Knowledge was gained of the two fundamental views, sagittal and transverse, in which the female pelvic organs are viewed. The student also learnt how to take different measurements including the endometrium at maximal thickness and width/length of the uterus. Finally she progressed to looking at and assessing the adnexa (right and left ovaries). Throughout the three weeks the student came in on the days that she was not being taught to practise on the mannequin until she was confident enough to teach other students. Furthermore the username and password to Medaphor, an online ultrasound training website was also given.

To build on her knowledge and understanding the student attended scanning sessions which were part of the UKCTOCS study and observed both transvaginal and transabdominal scans on postmenopausal women over the age of 55 years. This helped show how sonography is used in real patients and how scanning in reality differs from the 'perfect' mannequin.

During this time other third year medical students doing SSCs were recruited to take part in my project. In total seven students were recruited, who in the fourth week were split into three groups to facilitate teaching. On the initial occasion the groups came in a baseline assessment of their knowledge was carried out to see if they had any of sonography skills prior to any peer training with them. The peer tutor then taught the students how to perform a TVUS scan using the sequence she had been taught previously. Immediatley after the teaching session the students took it in turns to scan the mannequin themselves individually. They then came in individually at weekly intervals and were assessed by the peer tutor using the mark sheet already devised (see appendix). A point was given for each basic skill that they could demonstrate (for example if they could identify the right ovary), consequently every student was given an overall score out of a maximum of 14 points. During the assessment the students were not told what to do but were stopped at various points and given feedback on where to improve in order to enhance their scanning. This allowed them to see where they would need to focus their time practising before the next assessment session. In between the weekly assessments the students were given the opportunity to come into the ultrasound room to use the mannequin and machine to practice on as well as access to Medaphor. In total one baseline and four further assessments were carried out for each student.

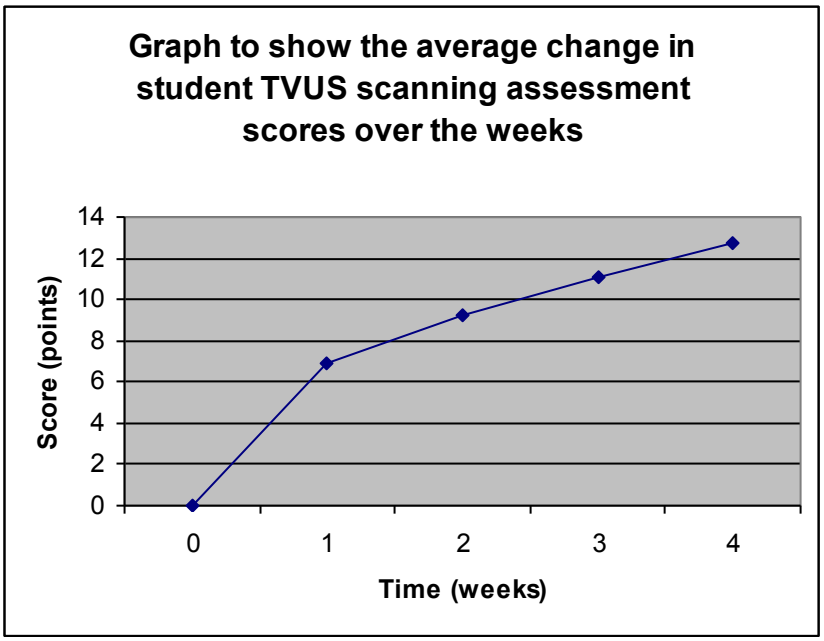
Results



In the graph above the gradient and curve of each line (showing each individual) represents student improvement in TVUS scanning. At the baseline assessment (week zero), all the students scored zero points showing that they had no prior knowledge whatsoever of how to perform a TVUS scan. However as the weeks went on, the students improved as shown by the positive gradient of each line. There are no dips in any of the lines meaning that no student ever got a weekly assessment score that was lower than their score the previous week.

The biggest improvement in scores for all students shown by the steepest gradient was between weeks zero and one, this was therefore a week on from the initial teaching session. The improvement then slowed down at each weekly assessment illustrated by the lines flattening off.

Although the students all improved no two lines on the graph are the same, showing their pattern improvement was unique. However after the initial teaching session the gradients of the lines seem to be very similar suggesting that improvement after this is alike in all students.



The graph above shows the mean of all the student scores for each week giving an overall trend-line. The average score after the teaching session at the first weekly assessment was seven and improvement then slowed with each student gaining an average of an extra two points each week relating to performing two more skills on the mark sheet.

To assess how useful the teaching was to the students themselves a questionnaire was also devised (see appendix). From their answers it was found that all the students believed the mannequin was an adequate way to learn how to scan as it allowed them to familiarise themselves with the equipment and allowed them to practice their skills without worrying about causing discomfort to a patient: one student said "I wouldn't want an actual person to go through what we put the models through in the practice!". When asked if the initial teaching session to be adequate to then be able to perform a scan, most students replied that although it was not detailed enough to then go ahead and scan a real patient it gave them a basic outline of what to do when practising on the model. All students felt that they improved throughout the weeks with the assessment sessions as the feedback allowed them to learn from their mistakes and improve on the things they did wrong for the next session. The questionnaire showed that all the students believed peer training was a good way of learning how to scan and the advantages of peer education included that the sessions were informal and having a peer as a tutor meant that the students did not feel embarrassed or as under pressure as might be felt when in front of a member of senior staff. This therefore supports the study by Sloane and Zimmer on why peer education is effective.

When asked if there were any parts of their medical degree where they thought peer education would be helpful and the main response was in the teaching of clinical skills. Students said that they often have different experiences on placement and learn different methods of examinations depending on the consultant they are with. Therefore they thought peer training would be of use as individual students could become familiar with the standard format of a certain procedure (e.g. how to perform a cardiovascular examination) and then teach it to other medical students in small groups. As for the disadvantages of peer training most of the students said that as the tutor is a peer and not an expert they might not be able to fully answer any questions asked, but a way to resolve this would be to have a senior member of staff to contact about any queries or extra teaching sessions by a qualified radiologist. Other ways to improve peer training included having more sessions that were less far apart in time so skills were not forgotten and having a step-by-step leaflet with instructions on that could be followed.

Discussion

The main finding of the project was that the medical students improved week on week throughout the 5 week period with the peer training sessions. This concurs with what was predicted prior to starting the project and also with the literature that claimed peer education is an effective way of teaching. It was expected that the baseline scores would be zero as the students had not yet undertaken an obstetrics and gynaecology placement and so the majority of third year students would not have yet seen or carried out a scan before. At week 1 the scores ranged from 5 to 10 points suggesting that some students picked up and remembered more from the initial teaching sessions than others. This could be due to the fact that they practised more on the mannequin in between the sessions, they made better use of the e-modules or they could simply have a better memory. The findings could be improved if the students filled in a weekly questionnaire asking how much practice they had done and whether they had been through any of the e-modules. This could then be correlated with the scores and we would expect to find that the students incorporating both practice with online training to achieve higher scores.
In general the results showed that the higher the starting score for a student, the higher their final score. However we believe if the project was to be carried on for a longer period of time all the students eventually with enough practice and assessment sessions would reach the maximum mark. Therefore the study could be improved by increasing the time frame. To strengthen the results further we could also have more students take part in the project as this would give me average results more representative of the population.

Conclusion

Overall the project has been successful in fulfilling its aims as a peer tutor along with 7 other medical students have learnt how to perform a basic TVUS scan on a mannequin. Furthermore through the assessment and feedback sessions the results showed that the students' scores improved weekly, therefore supporting the hypothesis that peer training is a useful method of teaching a practical skill such as how to perform an ultrasound scan. However it is not possible with the project design to see how effective peer training is in the teaching of TVUS scanning as we did not compare it with any other teaching method. Therefore to validate the effectiveness of peer education we would need to have more students allocated to different groups each with a different method of teaching for example one with a peer tutor, one with a consultant and one with no formal method of teaching. Then I could compare the results to see which method of teaching is most useful.
From my questionnaire, it was found that all the students found the use of the mannequin and teaching sessions to be useful and that peer education can be incorporated into the medical course especially when learning how to perform clinical skills such as taking blood pressure and clinical examinations.

Acknowledgments

(not sure)

Disclosure of interests

Non declared?

Contribution to authorship

Nazar Amso – project design and oversaw/ran entire project
Meena Murugan – peer tutor
Amal Alsalamah– Masters student acting as mentor to peer tutor

References

1. Woo J. A short history of the developmental of ultrasonography in obstetrics and gynaecology. 2002. http://www.ob-ultrasound.net/history1.html. Accessed 23/06/09.
2. Radiological Society of North America. Ultrasound Imaging of the Pelvis. 2009. http://www.radiologyinfo.org/en/info.cfm?pg=pelvus. Accessed 23/06/09/
3. WHO. Does peer education work in Europe? *The European Magazine for Sexual and Reproductive Health*. No 56. 2003.
4. Gopee N. Coventry University, School of Health and Social Sciences, Gulson Campus, Coventry The role of peer assessment and peer review in nursing.
5. Learning and teaching Scotland. Health promoting school practioners. Peer education. 08/05/08. http://www.ltscotland.org.uk/healthpromotingschools/practitioners/topics/emotionalwellbeing/peereducation.asp. Accessed 25/06/09.
6. Sloane BC, Zimmer CG. The power of peer health education. *Journal of American College Health* 1993; 41:241-245
7. National Hemophilia Foundation. Peer-to-Peer Health Education Programs for Youth: Their Impact on Comprehensive Health Education. New York: The Foundation, 1994.
8. Clack AR, Cheeseman GJ et al. A feasibility study of adolescent sex education: medical students as peer educators in Edinburgh schools. *BJOG: An International Journal of Obstetrics & Gynaecology*. Volume 107 Issue 3, Pages 436 – 436. 12/08/05.

APPENDIX

Mark sheet

Skill – Image optimization	**Inadequate 0**	**Adequate 1**
Overall image gain		
Time gain control		
Use of focus facility		
Depth		
Technique: Systemtic approach		
Adjust correct orientation on the screen		
Correct orientation- laterality		
Annotation		
Examine the uterus		
Show uterus in sagittal plane		
Show uterus in transverse plane		
Measure endometrium at the maximum thickness correctly		
Adnexal assessment		
Identify right ovary		
Identify left ovary		
Demonstrate right ovary in sagittal & transverse plane		
Demonstrate left ovary in sagittal & transverse plane		
Total score /14		

Questionnaire

1. Do you think using a mannequin is a useful way to learn how to perform a scan?
2. Did you find the initial teaching session to be adequate to then be able to perform a scan?
3. Do you think you improved throughout the weeks with the assessment and feedback sessions?
4. Do you think peer training was a good way to learn how to do an ultrasound scan?
5. What do you think are advantages of peer education?
6. Are there any parts of our medical degree where you think peer education/training would be useful
7. Do you feel there are any disadvantages to peer education?
8. Are there any ways the peer training of ultrasound scanning could be improved?
9. Any other comments

Appendix (9.2)

Comparison of Two Different Methods of Teaching Ultrasound Scanning to Students

Holly Morgan, Amal Alsalamah, Nazar Amso*

*Cardiff University School of Medicine, Academic Section of Obstetrics & Gynaecology

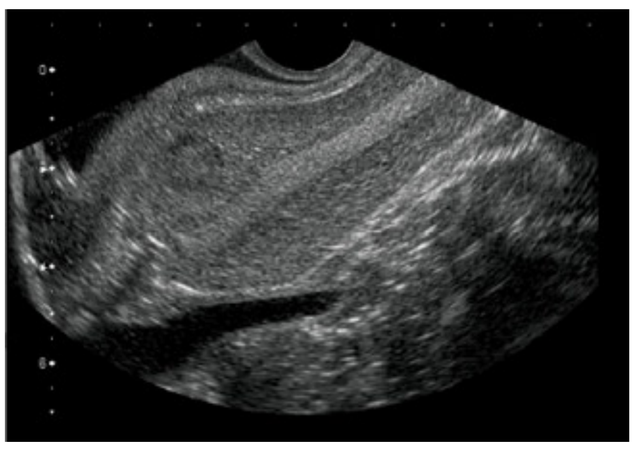
BACKGROUND

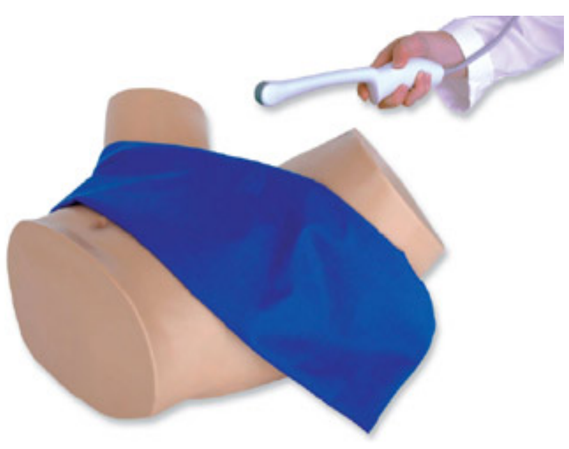
Trans-vaginal ultrasound simulation became available in 2010. The developers hope to be able to include this simulator into future obstetric and gynaecological training programmes across the UK to supplement existing training, which is often restricted by lack of mannequin equipment or volunteers. Simulation training has become very popular in recent years with many studies finding that they can be a very useful tool to aid learning whilst reducing one to one teaching costs.

AIMS

The project aimed to compare teaching trans-vaginal ultrasound to students on the simulator versus the conventional training of being taught on a mannequin.

Simulator Group: Trans-vaginal Ultrasound Simulator and Computer Interface

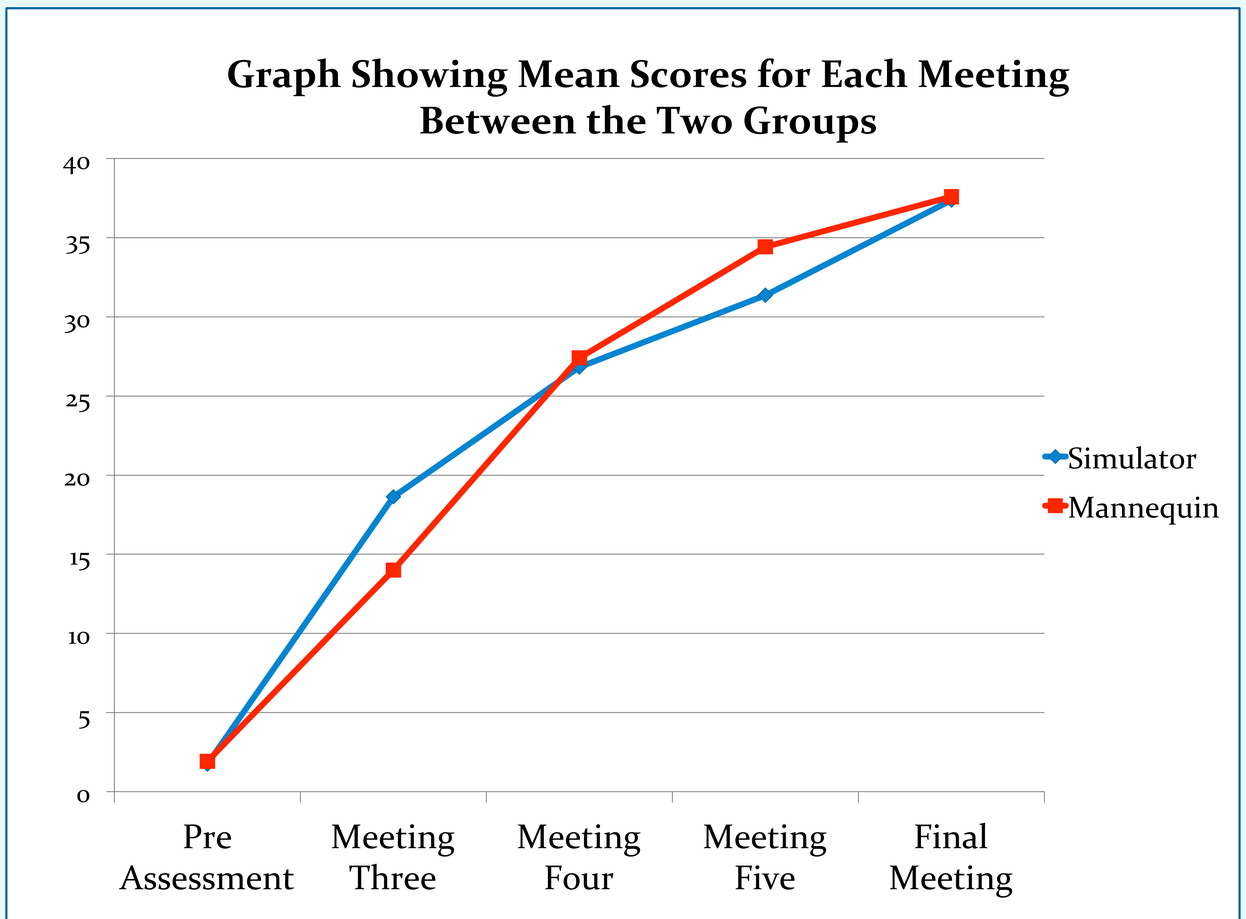




Mannequin Group: Typical Image Seen; Mannequin with Probe

METHODS

All third year Cardiff medical students were invited to take part via email, and 30 were initially recruited and randomised via computerised random number generation. The students were then taught at six once weekly sessions using one of the two methods. Students were assessed before each thirty minute tutorial using a standard form which was used for both groups. Both groups were also given access to an e-learning module which they could access for the duration of the study.



RESULTS

23 students completed the study. The mean mark achieved, out of 38, was 37.4 in the simulator group, and 37.6 for the mannequin group (p=0.93). The median score for both groups was 38. All final scores were found to be non significant between the two groups. The simulator group showed a steeper increase in scores for the first two meetings, (p=0.043), suggesting they found it easier to pick up the very basic skills. These results show that simulator training can lead to a similar level of practical skill acquisition as one to one sessions.

CONCLUSIONS

Preliminary evidence suggests that further studies with larger numbers and different categories of trainees are required to validate these results. Potentially the skills gained at the end of the study period could be assessed by the successful completion of a scan on a patient.

REFERENCES

- **Medaphor Simulation Technology (2010)** Available at: http://www.medaphor.com [Accessed on 18.11.10]
- IM Heer et al. (2004) Ultrasound training: the virtual patient. *Ultrasound in Obs and Gynae.* 24 4 440-4 Sep

CORRESPONDENCE EMAIL:
morganhp@cardiff.ac.uk