The effect of dyad versus individual simulation-based ultrasound training on skills transfer

Martin G Tolsgaard,1,2 Mette E Madsen,1 Charlotte Ringsted,3,4 Birgitte S Oxlund,1 Anna Oldenburg,1 Jette L Sorensen,1 Bent Ottesen1 & Ann Tabor1

CONTEXT Dyad practice may be as effective as individual practice during clinical skills training, improve students’ confidence, and reduce costs of training. However, there is little evidence that dyad training is non-inferior to single-student practice in terms of skills transfer.

OBJECTIVES This study was conducted to compare the effectiveness of simulation-based ultrasound training in pairs (dyad practice) with that of training alone (single-student practice) on skills transfer.

METHODS In a non-inferiority trial, 30 ultrasound novices were randomised to dyad (n = 16) or single-student (n = 14) practice. All participants completed a 2-hour training programme on a transvaginal ultrasound simulator. Participants in the dyad group practised together and took turns as the active practitioner, whereas participants in the single group practiced alone. Performance improvements were evaluated through pre-, post- and transfer tests. The transfer test involved the assessment of a transvaginal ultrasound scan by one of two clinicians using the Objective Structured Assessment of Ultrasound Skills (OSAUS).

RESULTS Thirty participants completed the simulation-based training and 24 of these completed the transfer test. Dyad training was found to be non-inferior to single-student training: transfer test OSAUS scores were significantly higher than the pre-specified non-inferiority margin (delta score 7.8%, 95% confidence interval −3.8–19.6%; p = 0.04). More dyad (71.4%) than single (30.0%) trainees achieved OSAUS scores above a pre-established pass/fail level in the transfer test (p = 0.05). There were significant differences in performance scores before and after training in both groups (pre- versus post-test, p < 0.01) with large effect sizes (Cohen’s d = 3.85) and no significant interactions between training type and performance (p = 0.59). The dyad group demonstrated higher training efficiency in terms of simulator score per number of attempts compared with the single-student group (p = 0.03).

CONCLUSION Dyad practice improves the efficiency of simulation-based training and is non-inferior to individual practice in terms of skills transfer.

Medical Education 2015; 49: 286–295
doi:10.1111/medu.12624

Discuss ideas arising from the article at www.mededuc.com discuss.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1Department of Obstetrics, Juliane Marie Centre, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark
2Centre for Clinical Education, Capital Region and University of Copenhagen, Copenhagen, Denmark
3Department of Anaesthesia, University of Toronto, Toronto, Ontario, Canada
4Wilson Centre, University Health Network, Toronto, Ontario, Canada

Correspondence Martin G Tolsgaard, Centre for Clinical Education and Department of Obstetrics, Juliane Marie Centre, Copenhagen University Hospital Rigshospitalet, Blegdamsvej 9, Copenhagen 2100 OE, Denmark. Tel: 00 45 61 30 30 72; E-mail: martintolsgaard@gmail.com

The copyright line for this article was changed on 18 January 2016 after original online publication.
INTRODUCTION

Simulation-based training is an effective method of improving clinical skills, but is highly resource-intensive.\(^1\, 2\) The significant time and monetary costs associated with simulation-based training have led medical educators to search for methods that improve the effectiveness and efficiency of training. Training in pairs – also known as dyad practice – may allow an increased learner-to-simulator ratio and thereby improve training efficiency. The effectiveness of dyad practice has been explored in the motor skills learning literature, which demonstrates that learners perform as well or better than those who practise alone despite decreased hands-on experience.\(^3\, 6\) Dyad practice may therefore be an attractive method of reducing the costs of simulation-based medical education without compromising learning.

The conceptual theoretical framework underpinning the effectiveness of dyad practice involves peer-assisted learning theory,\(^7\) neuroscience studies,\(^8\, 9\) and cognitive load theory (CLT).\(^10\) Whereas peer learning theorists emphasise the importance of inter-subjective cognitive co-construction, shared knowledge and peer feedback,\(^11\) experimental motor learning studies involving dyad practice have attributed the benefits of dyad practice to the effects of peer observation.\(^3\, 6\) The neurophysiological basis for the effects of observation has been provided by the identification of so-called ‘mirror-neurons’ located in the prefrontal motor cortex that are activated when an individual observes and performs the same action.\(^8\, 9\) In other words, observation elicits an internal representation of the performed actions, which allows rehearsal without any actual hands-on practice. Cognitive load theory provides yet another view on the potential mechanism of action behind dyad practice.\(^12\) From a CLT perspective, dyad practice may divide the cognitive load imposed on learners by shared information processing, thereby making learning more effective and efficient. However, dyad practice may also reduce the effectiveness of learning if the task is characterised by low cognitive load.\(^13\, 14\)

The effect of dyad practice within the simulated setting has been explored in several recent studies in medical education. In a study by Tolsgaard et al.,\(^15\) medical students were randomised to dyad or single practice during patient encounter management training in a skills laboratory. A retention test using simulated patients showed the dyad group performed better than the single-practice group and demonstrated improved confidence in managing future patient encounters.\(^15\) In a study involving simulation-based lumbar puncture training, Shanks et al.\(^16\) found no differences in pre-, post- and retention test performances between students who practised alone and those who worked in dyads. However, the dyad group displayed greater pre- to post-test gains than the single-practice group without using additional practice time.\(^16\) Finally, a recent study involving virtual reality coronary angiography training did not demonstrate any significant differences in retention test performances between students who were allocated to individual or dyad practice.\(^17\) Consequently, there is increasing evidence that dyad practice improves efficiency for a wide range of clinical skills without compromising learning. However, it remains unknown whether dyad practice is as effective as individual practice in the context of skills transfer from simulated to clinical performances with patients.

The process of transfer describes the application of previously learned concepts to a novel problem.\(^18\) The transfer process is influenced by multiple factors\(^19\) and performance in a simulated setting may not necessarily reflect how learners perform in the clinical setting with patients.\(^20\, 21\) Consequently, although dyad practice is as effective as individual practice during simulation-based training, it does not guarantee equivalent skills transfer. The positive effects of dyad practice on confidence and metacognition reported in previous studies\(^15\, 17\) indicate that dyad practice may actually improve the chances of successful skills transfer according to social cognitive theory.\(^19\, 22\) However, the decreased amount of hands-on experience may reduce dyad participants’ automaticity of skills during simulation-based practice, which may impair subsequent skills transfer according to the motor skills learning literature.\(^23\, 24\)

The aim of this study was therefore to assess the effectiveness of dyad practice compared with individual practice in the context of the transfer of skills from simulation-based training in complex skills to clinical performances with patients. We chose a non-inferiority design to investigate whether dyad training was as effective as individual practice according to recent recommendations.\(^25\) In the present study, we focused on simulation-based ultrasound training because ultrasonography is a complex skill involving technical and diagnostic aspects of performance.
METHODS

This study was a randomised, parallel-group, observer-blinded, non-inferiority trial using a 1:1 allocation ratio to compare dyad and single practice during ultrasound simulation between 1 January and 1 June 2013. We hypothesised non-inferiority between dyad and single practice and used pre-planned specified boundaries according to current recommendations.25 The study is reported according to the CONSORT statement, as well as the CONSORT extension on equivalence trials.26,27 Sample size calculations for this trial were based on previously published studies (reported below). A flowchart outlining the study is shown in Fig. 1.

Ethical approval was obtained from the Regional Ethical Committee of the Capital Region (protocol no. DRVK-35596). The study protocol was registered at clinicaltrials.gov (protocol no. NCT01703975).

Participants and setting

Participants in this study were medical students in their final clinical rotations before medical licensing. The group was selected to ensure that all participants had an equal knowledge base and minimal practical ultrasound experience. The medical programme at the University of Copenhagen is a traditional 6 year curriculum which provides basic science teaching during the first half of the programme and clinical sciences teaching as well as clerkship training during the second half. During Year 4, all medical students are required to complete a 3-hour course on basic abdominal ultrasonography.

Participants were contacted by e-mail and again during the introductory lectures to their gynaecological rotations. The inclusion criterion required participants to be less than 4 months from medical graduation. Exclusion criteria denied the participation of students who had previously undertaken ultrasound courses or had any experience with any kind of virtual reality simulation or ultrasound in obstetrics and gynaecology. Ultrasound experience was defined as any hands-on training and did not include the observation of examinations performed by other health care providers. Participants were enrolled on a consecutive basis until the required sample size was reached.

![Figure 1 Flowchart of the study](image-url)
The study was conducted at the Department of Obstetrics and Gynaecology, Copenhagen University Hospital Rigshospitalet. All participants underwent simulation-based ultrasound training. The gynaecological ambulatory unit was selected because it has a high flow of patients and thus the likelihood that participants would be able to perform transvaginal ultrasound scans was concomitantly high. The patient population included women scheduled for clinical control after medical abortion, as well as women with acute gynaecological conditions (e.g. missed abortion [missed miscarriage], lower abdominal pain). All patients provided written informed consent to their participation in this study. Eligible patients were non-pregnant or in an early stage of pregnancy and were required to have a body mass index of < 25 kg m\(^{-2}\) to ensure some degree of test condition standardisation. All patients were required to be able to read and understand Danish in order to provide informed consent.

Interventions

Participants were randomised to dyad or single practice before beginning the training. Simulations were performed using a high-fidelity transvaginal ultrasound simulator (ScanTrainer; MedaPhor Ltd, Cardiff, UK). The ultrasound simulator was equipped with a haptic device that provides realistic force-feedback to the ultrasound probe. The B-mode monitor output was obtained from real patients and presented according to the probe’s positioning, while a three-dimensional illustration simultaneously demonstrated probe orientation during the scans (Fig. S1). For this study, we selected the general gynaecological training modules, which included nine basic training exercises focused on probe orientation and systematic examination of a normal female pelvis.

Both groups were provided with a brief introduction (10 minutes) in order to familiarise group members with the simulated environment and the equipment. Participants were then instructed to complete a pre-test to assess baseline performance before training, as well as a post-test once they had finished training. They were instructed to perform a systematic ultrasound examination of the female pelvis which included the measuring of endometrial thickness within a time limit of 10 minutes. The 3-D illustration was not enabled during the pre- and post-tests and case difficulty was low (normal pelvic anatomy, module 1.1). All simulated and clinical ultrasound examinations were recorded to allow for subsequent performance assessment. All tests were completed individually; participants in the dyad group were not allowed to observe one another’s performances.

Once the pre-test was completed, participants were instructed to begin training using nine general gynaecological simulator modules. The training time was fixed for both groups at 2 hours, upon which the post-test was completed. Participants in the dyad group were instructed to take turns as the active practitioner and to attempt each module once and then to move on to the next module. The participants in the single group, practising alone, were instructed to attempt each module twice. If any of the participants managed to complete all nine modules within 2 hours, they were instructed to repeat the modules from the beginning. Discussion was allowed between the participants working in pairs, but only one participant at a time was involved in hands-on practice.

All participants were provided with two optional 10-minute feedback sessions with one of two simulator instructors (MGT or MEM). The feedback was structured according to a framework for the assessment of ultrasound skills, the objective structured assessment of ultrasound skills (OSAUS) (Appendix S1). After 2 hours of training, the sessions were ended and participants were briefed on the transfer test scheduled for the following day.

Outcomes

The main outcome measure was clinical performance during ultrasound scans performed the day after the completion of simulation-based training (transfer test). All participants were assessed on one transvaginal scan by one of two blinded raters (AO or BSO) using the OSAUS scale. Participants were instructed to perform a systematic examination of the uterus and lateral pelvic wall, as well as to assess free fluid in the pouch of Douglas (the recto-uterine pouch), and to measure endometrial thickness. The rater provided the participant with a short oral introduction to the equipment and setting before the participant performed the scans. During the scans, the rater assisted with technical problems, but provided limited verbal instructions unless patient discomfort prompted intervention. Technical assistance included help with operating the buttons on the ultrasound dashboard (i.e. handling ‘knobology’). The participant was allowed to scan for a maximum of 10 minutes before the rater took over the examination. Participants were excluded from the analysis of transfer test performances if no eligible
patients were booked or if no patients provided consen
t on the day following simulator training.

Secondary outcome measures included the simula-
tor-based pre- and post-tests, which were assessed
using the OSAUS by the same two raters who
assessed the transfer tests. Whereas only one of
the two raters assessed the transfer tests, both raters
assessed all pre- and post-tests. Finally, simulator
metrics (i.e. the automated feedback provided by
the simulator) were used to assess efficiency during
training (Appendix S2).

Validity of assessment instruments and rater training

The OSAUS framework evaluates ultrasound compe-
tence and consists of five items pertaining to equip-
ment knowledge, image optimisation, systematic
examination, image interpretation, and documenta-
tion of findings (Appendix S1). Each item is rated
on a 5-point Likert scale (1 = poor performance,
5 = excellent performance). The OSAUS scale has
demonstrated content validity,\textsuperscript{28} construct validity,
and high inter-rater reliability and internal consis-
tency,\textsuperscript{29,30} as well as evidence of structural validity.\textsuperscript{31}
Finally, criterion-based pass/fail standards have
been established for the OSAUS scale in a previous
study.\textsuperscript{30} The validity evidence of simulator metrics
used in the assessment of simulated performances
has been examined in a previous study.\textsuperscript{32} Only met-
rics that demonstrated construct validity were
included in the analyses of the present study.

The two raters were specialty registrars with an inter-
est in gynaecological ultrasound. The raters com-
pleted comprehensive training in assessing pre-
recorded ultrasound performances until rating
consensus was reached, which occurred after three
videos. The raters were instructed to assess perfor-
mances according to the level that can be expected
from a recently certified obstetrics and gynaecology
consultant.

Randomisation and blinding

The participants were informed and enrolled by
MEM and randomised to either dyad or single prac-
tice by an independent research fellow (TT) from
another department within the same hospital (the
Centre for Clinical Education). Randomisation was
performed in blocks of two using a web-based
randomiser. The dyads were paired in the order in
which they were randomised; hence they had no
influence on partner choice. The two raters were
blinded to the allocation of participants. Blinding
was ensured by training either one dyad or two con-
secutive single-practice participants at a time so that
two participants were always available for the trans-
fer test on the following day.

Sample size calculations

The primary outcome’s non-inferiority margin was
based on the smallest difference considered to be
clinically or educationally meaningful. We regarded
a difference in OSAUS scores of $< 4.6\%$ to be
within the non-inferiority margin. This was based on
previously published data, in which trainees demon-
strated a mean ± standard deviation (SD) increase
in OSAUS scores of $4.60 ± 3.74\%$ for every
2 months of clinical training.\textsuperscript{30} Hence, dyad train-
ing was considered non-inferior to single-student
training if differences in performance scores were
above the non-inferiority margin of $-4.6\%$. Using a
power of 90\% and an $\alpha$-value of 0.05, 24 partici-
pants were needed to establish non-inferiority
between the two groups based on the procedure
used to determine sample size in a non-inferiority
trial.\textsuperscript{33}

Statistical analysis

All simulator and performance scores were calcu-
lated into percentages of maximum score. The
mean OSAUS scores were calculated for the pre-,
post- and transfer test performances. Confidence
intervals (CIs) for the differences in transfer test
OSAUS scores between the two groups were calcu-
lated to assess the non-inferiority hypothesis. The
numbers of trainees who achieved transfer test
OSAUS scores above a pre-established criterion-ref-
erenced pass/fail level of 50.0\%\textsuperscript{30} were calculated
and compared between the two groups using chi-
squared statistics.

A mixed-design ($2 \times 2$) repeated-measures analysis
of variance (ANOVA) was performed using the pre-
ard post-test scores to assess the main effects of
training and to test for interactions with type of
training (dyad or single). The simulator scores were
calculated as the sum of metrics with established
validity evidence. Training efficiency was assessed by
dividing the simulator score on the final attempt on
the nine modules by the total number of attempted
modules. Simulator and efficiency scores were com-
pared between the two groups using Mann–Whitney
$U$-tests.

Internal consistency for the OSAUS items was calcu-
lated using Cronbach’s alpha; inter-rater reliability
for the pre- and post-test assessments was calculated using intraclass correlation coefficients (ICCs).

RESULTS

Thirty participants were randomised and completed the simulation-based training; 24 of these completed the transfer test. Baseline characteristics of the participants and mean OSAUS scores on the pre-, post- and transfer tests are shown in Table 1. There were no differences in the distribution of case pathology between groups $\chi^2(2) = 0.14, p = 0.93$ (Table 1).

The 95% CI for the differences in transfer test OSAUS scores between the two groups included zero but not the non-inferiority margin (delta score 7.8%, 95% CI –3.8–19.6%). Hence, the clinical performances of the dyads were not inferior to those of the single-practice students ($p = 0.04$), but neither were they superior ($p = 0.19$). However, a significantly higher number of dyads than single-practice students achieved scores above the pre-specified pass/fail level in the transfer test (dyads 71.4%, singles 30.0%; $\chi^2(1) = 4.03, p < 0.05$). The distribution of scores is shown in Fig. 2.

There were no significant interactions between effects of training and type of training (dyad or single) (d.f.(1,28) = 0.72, $p = 0.41$). There was a significant main effect of simulation-based training in terms of performance improvements from pre- to post-test (d.f.(1,28) = 509.933, $p < 0.01$), with large effect sizes (Cohen’s $d = 3.85$). There was no statistical difference between the two groups in final simulator scores (dyads: mean ± SD: 82.1 ± 9.6%; singles: mean ± SD 72.1 ± 21.3%; $p = 0.22$). However, the dyad group demonstrated significantly higher training efficiency with a mean ± SD of 5.88 ± 1.13 points/attempt compared with the single-practice group, which achieved a mean ± SD of 2.79 ± 0.92 points/attempt ($p < 0.01$) (Fig. 3). The ICCs for inter-rater reliability on the pre- and post-test assessments were 0.66 and 0.40, respectively. Those for internal consistency on the pre- and post-test assessments were 0.77 and 0.84, respectively.

| Characteristics                              | Dyad group (n = 16) | Single group (n = 14) |
|----------------------------------------------|---------------------|-----------------------|
| Men, n                                        | 1                   | 0                     |
| Women, n                                     | 15                  | 14                    |
| Age, years, mean (range)                     | 28.8 (27–34)        | 27.2 (23–31)          |
| Feedback sessions during training, mean ± SD | 0.9 ± 0.6           | 1.1 ± 0.7             |
| Pre-test OSAUS score, mean ± SD (%)          | 28.5 ± 4.2          | 24.9 ± 3.7            |
| Post-test OSAUS score, mean ± SD (%)         | 55.5 ± 6.3          | 49.3 ± 6.0            |
| Transfer test OSAUS score, mean ± SD (%)     | 56.3 ± 14.8         | 48.4 ± 11.8           |
| Distribution of cases, %                     |                     |                       |
| Incomplete abortion                          | 50.0                | 50.0                  |
| missed abortion or                            |                     |                       |
| complete abortion                             |                     |                       |
| Bleeding in relation to early pregnancy      | 33.3                | 25.0                  |
| Imminent abortion                             | 16.7                | 25.0                  |

OSAUS = objective structured assessment of ultrasound skills; SD = standard deviation.

DISCUSSION

Participants in the dyad group had only half the amount of hands-on time, but their performances with patients were non-inferior to those of participants who practised alone. In fact, a significantly higher number of participants in the dyad group compared with those in the single-practice group achieved performance scores above a pre-established pass/fail level. Consequently, training efficiency was doubled without any sign of negative effects on transfer of learning. The fact that the dyad group needed significantly fewer attempts to achieve a certain simulator score supports the suggestion that dyad participants were able to learn from each other’s errors and avoid them during subsequent performance. This confirms existing theories on the mechanisms underlying the efficacy of dyad practice from the motor skills learning and peer learning literature. However, in the present study we did not attempt to explore the relative contributions of observation, feedback, shared knowledge or cognitive co-construction on the observed performances in the dyad group.

Of the 30 participants included, 24 completed the transfer test, thereby fulfilling the study’s targeted sample size. Although this study was powered to demonstrate only non-inferiority, the dyad group
demonstrated superiority with regard to simulator training efficiency and transfer test pass rate. This study used assessment instruments for which validity evidence is documented, rater blinding, and thorough rater training prior to its commencement.

Although simulator instructor blinding was not possible, we used standardised instruction and feedback protocols to limit any instructor bias. The fact that the effects of dyad training generalise across several practice domains supports the trustworthiness of the

Figure 2 Pre-, post- and transfer test performances of participants trained in pairs (dyads) or alone (singles). OSAUS = objective structured assessment of ultrasound skills

Figure 3 Efficiency of simulation-based training for participants who worked in pairs (dyads) and alone (singles), respectively. The number of attempts during the 2-hour training session is plotted against the simulator scores obtained on the final attempt at the nine modules (i.e. the sum score of all passed simulator metrics)
present results. However, the present study included only novice learners, who had limited clinical experience and no prior ultrasound training. By contrast with our findings, Shanks et al.\textsuperscript{16} hypothesised that novice learners may not benefit from dyad practice because they lack the knowledge and skills necessary for effective peer teaching. This was supported by an experimental study showing that peer-assisted learning in pairs was inferior to computer-assisted learning and teacher-assisted learning.\textsuperscript{24} However, the training provided in our study included a combination of computer- and teacher-assisted learning, and the effects of dyad practice for novice learners may therefore depend on some element of guidance during training.

In the present study, we examined the effect of dyad practice during complex skills training, but similar results have been demonstrated for simple skills training in a previous study involving lumbar puncture.\textsuperscript{16} Hence, the effectiveness of dyad practice may not exclusively rely on task complexity, as proposed by Kirschner et al.,\textsuperscript{13,14} as the cognitive load imposed during initial simple skills training may be equivalent to the load after several hours of complex skills training. As cognitive load decreases with increasing levels of skills automativeness,\textsuperscript{10,24} the benefits of dyad practice may be reduced and this reduction may occur independently of skill complexity. Interestingly, the time allocated to dyad practice during simulation-based training in the study by Shanks et al.\textsuperscript{16} (24 minutes), the present study (2 hours), and the study by Räder et al.\textsuperscript{17} (3.5 hours) seems to be inversely related to the immediate gains in learning of the dyad participants. In the study by Shanks et al.,\textsuperscript{16} the dyad group demonstrated significantly greater pre-to post-test gains than the single-practice group, whereas there were no differences between groups in the present study and slightly worse performances in the dyad group in the study by Räder et al.\textsuperscript{17} The trade-off between gains from dyad practice in terms of peer observation and shared information processing against the reduction in hands-on experience may help to explain the reduced effects of dyad practice beyond initial training. Future qualitative studies may help to explore how these issues affect learning.

Limitations

The limitations of this study include its small sample size and the use of single-case assessments for the transfer test. During the transfer tests, only one rater assessed performances, which may have impacted the reliability of ratings. Although the validity of the OSAUS scale has been explored in multiple previous studies involving different trainees and institutions,\textsuperscript{28–31} the validity of a score does not necessarily transfer to a new setting with a different sample of participants. Furthermore, because of the large differences between pre- and post-test performances, the raters may have been able to distinguish between the types of assessment in the simulated setting. Although this may have resulted in the overestimation of the immediate training effects for both groups, it does not affect the interpretation of the main outcome or affect the overall conclusions of this study as the raters were blind to the allocation of participants in the pre-, post- and transfer tests. To limit the case-mix effect, we aimed for strict selection criteria for the transfer test cases and found no significant differences between groups in the types of pathology participants encountered. To ensure the standardisation of the transfer test conditions between the two groups, we included only scans performed on the day after training. Hence, six of the 30 participants included were unable to complete the transfer test because we lacked eligible patients. We chose to limit the transfer test to include a single scan because aiming for a higher number of scans would have limited the completion rate further.

The use of different methods for assessing pre-, post- and transfer test performances may have resulted in bias as there is some evidence that live performances are rated more highly than video-recorded performances.\textsuperscript{35} Hence, the extent of skills transfer to the clinical setting may have been overestimated for both groups of participants as a result of rater leniency. Finally, we examined only the immediate effects of dyad practice on skills transfer and did not attempt to evaluate effects on skills retention. This has, however, been examined in several recent studies, in which no significant differences were found between dyad and single practice in the simulated setting after delays of 2–3 weeks.\textsuperscript{15–17,36}

CONCLUSIONS

The finding that dyad practice was non-inferior to individual practice with regard to skills transfer is remarkable as dyad participants had only half the hands-on training of single-practice participants. Dyad practice may be used in multiple clinical disciplines to lower the time and monetary costs associated with early simulation-based training. However,
the type of guidance needed to facilitate the positive effects of dyad practice, as well as the effectiveness of dyad practice during different stages of learning are subjects for further studies.

**Contributors:** MGT contributed to the study design, provided instructions during simulator training, wrote the statistical analysis plans, monitored data collection throughout the trial, analysed the data and contributed to data interpretation. MEM contributed to the study design, provided instructions during simulator training and contributed to data interpretation. CR and AT contributed to the conception and design of the study, the statistical analysis plans and data interpretation. BSO and AO contributed to the study design, the statistical analysis plans and data interpretation. JLS and BO contributed to the critical revision of the paper and approved the final manuscript. All authors are responsible for all aspects of the work.

All authors contributed to the critical revision of the paper and approved the final manuscript. All authors are responsible for the integrity of the data and are accountable for all aspects of the work.

**Acknowledgements:** the authors are grateful to Tobias Todsen at the Centre for Clinical Education, Denmark, for conducting the concealed computer-based randomisation of participants.

**Funding:** this study presents independent research funded by a grant from the University of Copenhagen.

**Ethical approval:** this study complied with the Helsinki II Declaration on biomedical research. Ethical approval was obtained from the Regional Ethical Committee of the Capital Region (protocol no. DRVK-35596).

**REFERENCES**

1. Zendejas B, Wang AT, Brydges R, Hamstra SJ, Cook DA. Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery* 2013;153:160–76.

2. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003–2009. *Med Educ* 2010;44:50–63.

3. Shea CH, Wulf G, Whitacre C. Enhancing training efficiency and effectiveness through the use of dyad training. *J Mot Behav* 1999;31:119–25.

4. Sanchez-Ku ML, Arthur W Jr. A dyadic protocol for training complex skills: a replication using female participants. *Hum Factors* 2000;42:512–20.

5. Arthur W Jr, Day EA, Bennett W Jr. Dyadic versus individual training protocols: loss and reacquisition of a complex skill. *J Appl Psychol* 1997;82:783–91.

6. Granados C, Wulf G. Enhancing motor learning through dyad practice: contributions of observation and dialogue. *Res Q Exerc Sport* 2007;78:197–203.

7. Topping KJ. Trends in peer learning. *Educ Psychol* 2005;25:631–45.

8. Rizzolatti G, Craighero L. The mirror-neuron system. *Annu Rev Neurosci* 2004;27:169–92.

9. Mattar AAG, Gribble PL. Motor learning by observing. *Neuron* 2005;46:153–60.

10. Sweller J. Cognitive load during problem solving. *Cogn Sci* 1988;12:257–85.

11. Topping KJ, Eby SW. Peer-assisted learning: a framework for consultation. *J Educ Psychol Consult* 2001;12:113–32.

12. Van Merriënboer JJJG, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;44:85–93.

13. Kirschner F, Paas F, Kirschner PA. A cognitive load approach to collaborative learning: united brains for complex tasks. *Educ Psychol Rev* 2009;21:31–42.

14. Kirschner F, Paas F, Kirschner PA. Task complexity as a driver for collaborative learning efficiency: the collective working-memory effect. *Appl Cognit Psychol* 2011;25:615–24.

15. Tolsgaard MG, Bjerck S, Rasmussen MB, Gustafsson A, Ringsted C. Improving efficiency of clinical skills training: a randomised trial. *J Gen Intern Med* 2013;28:1072–7.

16. Shanks D, Brydges R, den Brok W, Nair P, Hatala R. Are two heads better than one? Comparing dyad and self-regulated learning in simulation training. *Med Educ* 2013;47:1215–22.

17. Räder SB, Henriksen A-H, Butrymowich V, Sander M, Jørgensen E, Lönn L, Ringsted CV. A study of the effect of dyad practice versus that of individual practice on simulation-based complex skills learning. *Acad Med* 2014;89:1287–94.

18. Kufasegaram K, Min C, Ames K, Howey E, Neville A, Norman G. The effect of conceptual and contextual familiarity on transfer performance. *Adv Health Sci Educ Theory Pract* 2012;17:489–99.

19. Burke LA, Hutchins HM. Training transfer: an integrative literature review. *Hum Resource Dev Rev* 2007;6:263–96.

20. Shea JB, Morgan RL. Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *J Exp Psychol 1979;5:179–87.*

21. Cook DA, West CP. Perspective: reconsidering the focus on ‘outcomes research’ in medical education: a cautionary note. *Acad Med* 2013;88:162–7.

22. Bandura A. *Self-Efficacy: The Exercise of Control.* New York, NY: Freeman 1997.

23. McGill RA. *Motor Learning and Control.* Concepts and Applications. Boston, MA: McGraw-Hill College 2007.

24. Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomised controlled trial. *Ann Surg* 2012;255:30–7.

25. Tolsgaard MG, Ringsted C. Using equivalence designs to improve methodological rigour in medical education trials. *Med Educ* 2014;48:220–1.
26 Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG, CONSORT Group. Reporting of noninferiority and equivalence randomised trials: extension of the CONSORT 2010 Statement. JAMA 2012;308:2594–604.
27 Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Ann Intern Med 2010;152:726–32.
28 Tolsgaard MG, Todsen T, Sorensen JL, Ringsted C, Lorentzen T, Ottesen B, Tabor A. International multispecialty consensus on how to evaluate ultrasound competence: a Delphi consensus survey. PLoS ONE 2013;8 (2):e57687.
29 Todsen T, Tolsgaard MG, Olsen BH, Henriksen BM, Hillingsø JG, Konge L, Jensen ML, Ringsted C. Reliable and valid assessment of point-of-care ultrasonography. Ann Surg 2014;261:309–15.
30 Tolsgaard MG, Ringsted C, Dreisler E, Klemmensen A, Loft A, Sorensen JL, Ottesen B, Tabor A. Reliable and valid assessment of ultrasound operator competence in obstetrics and gynecology. Ultrasound Obstet Gynecol 2014;43:437–43.
31 Tolsgaard MG, Rasmussen MB, Tappert C, Sundler M, Sorensen JL, Ottesen B, Ringsted C, Tabor A. Which factors are associated with trainees’ confidence in performing obstetric and gynecological ultrasound examinations? Ultrasound Obstet Gynecol 2014;43:444–51.
32 Madsen ME, Konge L, Nørgaard LN, Tabor A, Ringsted C, Klemmensen AK, Ottesen B, Tolsgaard MG. Assessment of performance and learning curves on a virtual reality ultrasound simulator. Ultrasound Obstet Gynecol 2014;44:693–99.
33 Julious SA. Sample sizes for clinical trials with normal data. Stat Med 2004;23:1921–86.
34 Walsh CM, Rose DN, Dubrowski A, Ling SC, Grierson LEM, Backstein D, Carnahan H. Learning in the simulated setting: a comparison of expert-, peer-, and computer-assisted learning. Acad Med 2011;86:12–6.
35 Konge L, Annema J, Clementsen P, Minddal V, Vilmann P, Ringsted C. Using virtual-reality simulation to assess performance in endobronchial ultrasound. Respiration 2013;86:59–65.
36 Bjerrum AS, Eika B, Charles P, Hilberg O. Dyad practice is efficient practice: a randomised bronchoscopy simulation study. Med Educ 2014;48:705–12.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Monitor output from the transvaginal ultrasound simulator.

Appendix S1. The objective structured assessment of ultrasound skills.

Appendix S2. Simulator metrics with validity evidence in terms of ability to discriminate between differences in performance in a group of competent performers (obstetrics and gynaecology consultants) and a group of non-competent performers (medical students).

Received 12 June 2014; editorial comments to author 29 July 2014, 8 September 2014; accepted for publication 11 September 2014