# Educational effectiveness of gynaecological teaching associates. A multi-centre randomised controlled trial.

Australian New Zealand Clinical Trial Registry: 363283

James M N Duffy[1], Samuel Chequer[2], Aaron Braddy[3], Sophie Mylan[2], Ana Royuela[4], Javier Zamora[4], Jacey Ip[5], Shelia Hayden[2], Marian Showell[6], Paul Kinnersley[7], Rashna Chenoy[8], Olwyn M Westwood[2], Khalid S Khan[8], Annie Cushing[2].

[1]Nuffield Department of Primary Health Sciences, University of Oxford, Oxford.
[2]Clinical and Communication Skills Learning Unit, Barts and the London School of Medicine and Dentistry, London.
[3]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield.
[4]Clinical Biostatistics Unit, Hospital Ramón y Cajal, Madrid.
[5]Department of Obstetrics and Gynaecology, Homerton University Hospital NHS Foundation Trust, London.
[6]Cochrane Menstrual Disorders and Subfertility Group, University of Auckland, Auckland.
[7]Institute of Medical Education, Cardiff University, Cardiff.
[8]Women's Health Research Unit, Queen Mary, University of London, London.

Correspondence to:
Dr James M N Duffy MBChB MRes BSc (Hons) PG HCL
Nuffield Department of Primary Health Sciences
Radcliffe Observatory Quarter
University of Oxford
Oxford OX2 6GG
United Kingdom
Telephone: +44 18 6528 9298
Facsimile: +44 18 6528 9297
E-mail: james.duffy@balliol.ox.ac.uk

Running title: Educational effectiveness of gynaecological teaching associates.

## Abstract

**Objective:** To evaluate, among medical students learning the female pelvic examination, the added benefits of training by gynaecological teaching associates compared to training involving a manikin only.

**Design:** Randomised controlled trial.

**Setting:** Seven university teaching hospitals.

**Population:** 94 medical students recruited prior to commencing a four-week obstetrics and gynaecology rotation.

**Methods:** The control training consisted of lectures, demonstration of the pelvic examination on a manikin, and opportunities to practice on this low fidelity simulation (n=40). The experimental group received additional gynaecological teaching associate training, delivered by pairs of experienced associates to groups of four medical students (n=54).

**Main Outcome Measure:** Outcomes measured at the end of the rotation included knowledge of the correct order of examination components (yes/no), and student comfort (Likert scales anchored between 1 [very uncomfortable] and 4 [very comfortable] on 4 items) and confidence (Likert scales anchored between 1 [No] and 3 [Yes] on 6 items). The primary outcome, measured at the end of the academic year, was the objective structured clinical examination of a female pelvis (score range, 0-54).

**Results:** At baseline, the groups were similar in age, gender, and ethnicity. At the end of the clinical rotation the experimental intervention had an impact on knowledge (difference 29.9% [95% CI 11.2 to 48.6%]; p=0.002), and student confidence

(difference 1 [95% CI 0 to 2]; p<0.001) and comfort (difference 1.8 [95% CI 0.6 to 3.0]; p=0.004) compared to control. At the end of the academic year, the experimental intervention had no impact on skills compared to the control (difference 2 [95% CI -1 to 4]; p=0.26).

**Conclusions:** Among medical students taught the female pelvic examination by low fidelity simulation, additional training by gynaecology teaching associates improved knowledge, comfort, and confidence at the end of the clinical rotation, but did not improve examination skills at end of the academic year.

**Trial Registration:** Australian New Zealand Clinical Trial Registry: 363283 (https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=363283)

**Keywords:** Pelvic examination, speculum examination, gynaecological teaching associates, lay person training, medical examination

## Introduction

Pelvic examination is an essential component of the care women receive in primary and secondary care. Papanicolaou smears alone account for 4% of all healthcare visits by women in the United States [1]. Learning to perform the pelvic examination is difficult. Medical students are required to acquire these skills as a core competency. Typical training strategies involve didactic sessions, audio-visual demonstrations, and instruction involving low fidelity simulation including manikins. Gynaecological teaching associates (GTAs) are lay women trained to teach the pelvic examination with themselves being examined. They usually work in pairs, one acting as an instructor with the other as a patient. GTAs are trained in providing immediate and constructive feedback during and after the examination with regards to technical and interpersonal skills.

The vast number of medical schools in Canada, The Netherlands, and The United States employ GTAs but this approach is not universally adopted. The educational effectiveness of GTA-delivered training has been evaluated in four single-centre randomised controlled trials (RCTs) [2-5]. These studies suffered several limitations: choice of an inferior comparator [2], limited statistical power [3-5], lack of assessment of the retention of learning over time [3,4], incompleteness of participant follow up through the study [3,5], lack of clarity concerning intention to treat analysis [2-5], attrition and reporting bias [3-5], and limited generalisability [2-5].

We conducted a high quality, multi-centre RCT evaluating the educational effectiveness of GTA delivered training over the short and medium term.

## Methods

### Ethical Approval and Registration

Approval for the study was obtained from the Queen Mary, University of London's ethics committee (reference number: QMREC2012/67) and all students provided informed written consent. The trial was prospectively registered with the Australian New Zealand Clinical Trial Registry (reference number: 363283).

### Participants

Medical students scheduled to undertake the standard female pelvic examination training before commencing a four week obstetrics and gynaecology rotation were recruited from seven hospitals during the 2012-13 academic year. Students who had previously undertaken female pelvic examination training were excluded. Enrolled participants completed a questionnaire recording demographic information including age, gender, ethnicity, and their additional academic achievements.

### Interventions

All participants received the standard (control) training consisting of lectures, demonstration of the pelvic examination on a manikin, and the opportunity to practice on it. Each teaching session lasted three hours and was facilitated by an experienced gynaecologist. Computer-generated randomisation (1.4 experiment to control allocation ratio), with concealment using consecutively numbered, opaque sealed envelopes allocated enrolled students to receive additional GTA delivered training (experiment). Sixty GTA training opportunities were available. The control to experimental ratio ensured these opportunities were maximally utilised. Randomisation and allocation concealment was performed by a third party.

GTAs delivering the experimental intervention had undertaken 28 hours of structured training and were certified competent by the medical school faculty before delivering student training. The participant training sessions lasted two and a half hours and were conducted by two experienced GTAs who taught a group of four participants. Participants observed an associate undertaking a gynaecological consultation, requesting informed verbal consent, and pelvic examination on another associate. The associates then guided each participant through a gynaecological consultation and examination, giving each participant the opportunity to practice and receive individualised feedback. All participants subsequently attended a four-week obstetrics and gynaecology rotation.

**Outcomes**

At recruitment, participants were asked to complete baseline measurements including knowledge of the pelvic examination components (yes/no) and self-rated comfort at the prospect of performing a pelvic examination on a conscious patient, using a response to four items on a Likert scale anchored between 1 [very uncomfortable] and 4 [very comfortable] (score range: 4-16). At the end of their clinical rotation participants were asked to re-score these measures and their confidence in performing a female pelvic examination, using a response to six items on Likert scale anchored between 1 [No] and 3 [Yes] (score range: 6-18). The comfort and confidence measures were adapted from existing validated tools [6,7]. At the end of the academic year the participants undertook a summative objective structured clinical examination (OSCE), which included a female pelvic examination station. This station involved a simulated patient (an associate not involved in the trial) lying on a couch with a manikin placed strategically [8]. The participant was asked to interact with the patient and examine the manikin. Technical and

interpersonal skills were assessed using a 54 item standard assessment tool scored by a trained gynaecologist and the simulated patient, blinded to the student's allocation. Twenty-eight items contributed to technical skills score and the remaining 26 items contributed to the interpersonal skills score. Quality assurance included outcome assessor training, an independent invigilator observing, and formal assessment conditions. The OSCE score served as the primary outcome measure.

**Statistical Analysis**

The sample size calculation employed the assumption that there would be a 15% improvement, equating to a moderate effect on Cohen's scale, in technical skill scores in the experimental intervention compared to the control (score 23 vs 20 with standard deviation estimated to be 5.2 in the 2012 student cohort) [7]. The power was set at 80% and significance level at 5%. We used a 1.4 experiment to control allocation ratio in the randomisation process to optimise the use of the available GTA training slots. We planned to recruit 101 participants (59 and 42 in experimental and control groups respectively) with complete data. To allow for a 10% drop out or loss to follow-up, 112 participants were sought.

Descriptive statistics (frequencies, means and standard deviations, or medians and $25^{th}$ and $75^{th}$ percentiles) were used to describe the participant demographics. Technical and communication skills were assessed during the summative OSCE and compared by means of non-parametric Mann-Whitney test in light of non-normal distribution. In order to estimate the effect of the intervention for self-reported knowledge and student comfort, we fitted two generalised estimating equations models, with the overall score as dependent variable and time of observation (baseline or after intervention), group (control or experimental) and the product of

time x group as independent variables. We defined an independent covariance structure. For self-reported knowledge, binomial family was used with the logit link function. For self-reported student comfort, Gaussian family was used with an identity link function. GEE models use all information available consistently with the intention to treat principle making imputation strategies unnecessary. Self-reported student confidence scores were compared by means of non-parametric Mann-Whitney test. We determined the importance of the size of educational effect using Cohen's standardised effect size for measures on continuous scales and for proportions [9]. An effect of 0.2 is considered small, 0.5 moderate, and 0.8 large. All analyses were performed using Stata v 13.0 (StataCorp, College Station, Texas) and $p<0.05$ was considered statistically significant.

## Results

We approached 130 eligible medical students, of whom 94 (72%) were randomised (Figure 1). At baseline the characteristics of the randomised participants, including age, gender, ethnicity, knowledge and comfort were similar between groups (Table 1).

At the end of the clinical rotation, when compared to the control intervention, the experimental intervention had a moderate effect on knowledge (21.1% in the control group vs 50.9% in the experimental group; difference 29.9% [95% CI 11.2 to 48.6%]; p=0.002; effect size=0.63) and participant confidence (median 17 in the control group vs 18 in the experimental group; difference 1 [95% CI 0 to 2]; p=<0.001; effect size =0.51), and a large effect on participant comfort (12.7 in the control group vs 14.6 in the experimental group; difference 1.8 [95% CI 0.6 to 3.0]; p=0.004; effect size = 1.2) (Table 2 & 3).

At the end of the academic year, after an average follow up of 5.3 months in the experimental group and 5.6 months in the control group, the experimental intervention had a small effect on technical and interpersonal skills when compared to the control intervention (effect size = 0.30 and 0.25 respectively). Median values were 24 (IQR 21 -27) and 20 (IQR 17-24) in the experimental group compared with 24 (IQR 20-26) and 19 (IQR 17-22) in the control group respectively (Table 3). Overall, the experimental intervention had no impact on skills compared to the control (median 43 in the control group vs 44 in the experimental group; difference 2 [95% CI -1 to 4]; p=0.26; effect size 0.3).

## Discussion

### Main Findings

Among medical students taught the female pelvic examination by low fidelity simulation, additional training by GTAs improved knowledge and student comfort and confidence at the end of the clinical rotation, but it did not improve examination skills at end of the academic year.

### Strengths and Limitations

The strengths of this prospectively registered study include its robust methodological design with rigorous random sequence generation and allocation concealment methods. Previous RCTs were associated with several limitations outlined in the introduction. This is, to our knowledge, the first multi-centre RCT evaluating the effectiveness of GTA delivered training, enhancing the generalisability of its findings. The validity of the study was also enhanced by robust measurement of technical and interpersonal skills. Unlike previous studies measurement occurred five months following the intervention, and deployed a 54 item standard assessment tool scored by a trained outcome assessors blinded to the student's allocation. Further quality assurance included formal assessment conditions supervised by an external invigilator. The use of a range of outcomes including knowledge, skills, and student reported confidence and comfort measures informed a more complete evaluation of the experimental intervention.

Multi-centre RCTs are not without limitations. We approached 130 eligible medical students, of whom 94 (72%) were randomised. This student non-participation rate could introduce non-response bias. The 28% non-participation rate is not

uncommon in educational research where participation is entirely voluntary. Students were reluctant to explain their justification for non-participation. Several students considered the GTA training sessions, which were scheduled during the evening, to be inconvenient. It would have been interesting to explore if the decision not to participate within the trial was influenced by academic performance or perceived psychosocial difficulties with the female pelvic examination. Furthermore, although several outcome measures have been reported in other trials, some skills learned may not have been assessed in sufficient detail, especially in the areas of professionalism and patient satisfaction.

Interpretation

Our primary outcome measure was assessed at the end of the academic year, approximately five months following the intervention. The experimental intervention had a small effect on skills when compared to the control intervention. We can speculate students trained by low fidelity methods acquired additional skills during the subsequent obstetrics and gynaecology rotation. We are aware that formal summative examinations are strong motivators for learning. Students may have equipped themselves with the skills needed regardless of prior training and skills gained during their clinical rotations [10].

## Conclusion

Medical schools considering new or continuing investment in GTA delivered training should carefully consider its cost effectiveness as it did not appear to produce any gains in summative assessments.

## Acknowledgements

We thank Mr David J. Mills for administrative and logistical support during the randomisation process.

## Disclosure of Interests

All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

## Contribution to Authorship

Prof Westwood and Dr Zamora had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Miss Chenoy, Prof Cushing, Dr Duffy, Prof Khan, Prof Kinnersley, and Mrs Showell.

Acquisition of data: Dr Braddy, Mr Chequer, Mrs Hayden, Dr Ip, and Dr Mylan.

Analysis and interpretation of data: Dr Duffy, Prof Khan, Dr Royuela, Prof Westwood, and Dr Zamora.

Drafting of the manuscript:  Dr Duffy, Prof Khan, and Dr Zamora.

Critical revision of the manuscript for important intellectual content: Dr Braddy, Mr Chequer, Prof Cushing, Mrs Hayden, Dr Ip, Dr Mylan, Dr Royuela, and Prof Westwood.

Statistical analysis: Dr Royuela and Dr Zamora.

Obtained funding: Dr Chenoy, Prof Cushing, Dr Duffy, Mrs Hayden, Prof Khan, and Prof Westwood.

Administrative, technical, or material support: Mrs Showell and Dr Mylan.

Study supervision: Prof Cushing, Dr Duffy, and Prof Khan.

## Details of Ethics Approval

Approval for the study was obtained from the Queen Mary, University of London's ethics committee (reference number: QMREC2012/67) and all students provided informed written consent.


## Funding

This trial was funded by The Medical College St Bartholomew's Hospital Trust (reference number: 8368137). The funding sources had no role in the design and conduct of the study; the collection, management, analysis, or interpretation of the data; or the preparation, review, or approval of the manuscript.


## References


[1] Centers for Disease Control and Prevention. National Ambulatory Care Survey (2013). http://www.cdc.gov/nchs/ahcd.htm. Published January, 2014. Accessed May 3, 2014.

[2] Herbers JE, Wessel L, El-Bayoumi J, et al. Pelvic examination training for interns: a randomised controlled trial. *Acad Med*. 2003;78(11):1164-1169.

[3] Livingstone RA, Ostrow DN. Professional patient-instructors in the teaching of the pelvic examination. *Am J Obstet Gynecol*. 1978;132(1):64-67.

[4] Nelson LH. Use of professional patients in teaching pelvic examinations. *Obstet Gynecol*. 1978;52(5):630-633.

[5] Vontver L, Irby D, Rakestraw P, et al. The effects of two models of pelvic examination instruction on student performance and anxiety. *J Med Ed*. 1980;55:778-785.

[6] Evans DE, Roberts CM. The effect of an extended hospital induction on perceived confidence and assessed clinical skills of newly qualified pre-registration house officers. Med Educ. 2004;38(9):998-1001.

[7] Siwe K, Wijma K, Stjernquist M, Wijma B. Medical students learning the pelvic examination: comparison of outcome in terms of skills between a professional patient and a clinical patient model. *Patient Educ Couns*. 2007;68(3):211-217.

[8] Kneebone R, Nestel D, Wetzel C, et al. The human face of simulation: patient-focused simulation training. *Acad Med*. 2006;81:919-924.

[9] Cohen J. *Statistical Power Analysis for the Behavioural Sciences.* 2nd ed. New York, NY: Lawrence Erlbaum Associates; 1988.

[10] Boud D. Assessment and the promotion of academic values. *Stud High Educ*. 1990;15:101-111.

**Figure 1.** Study Flow

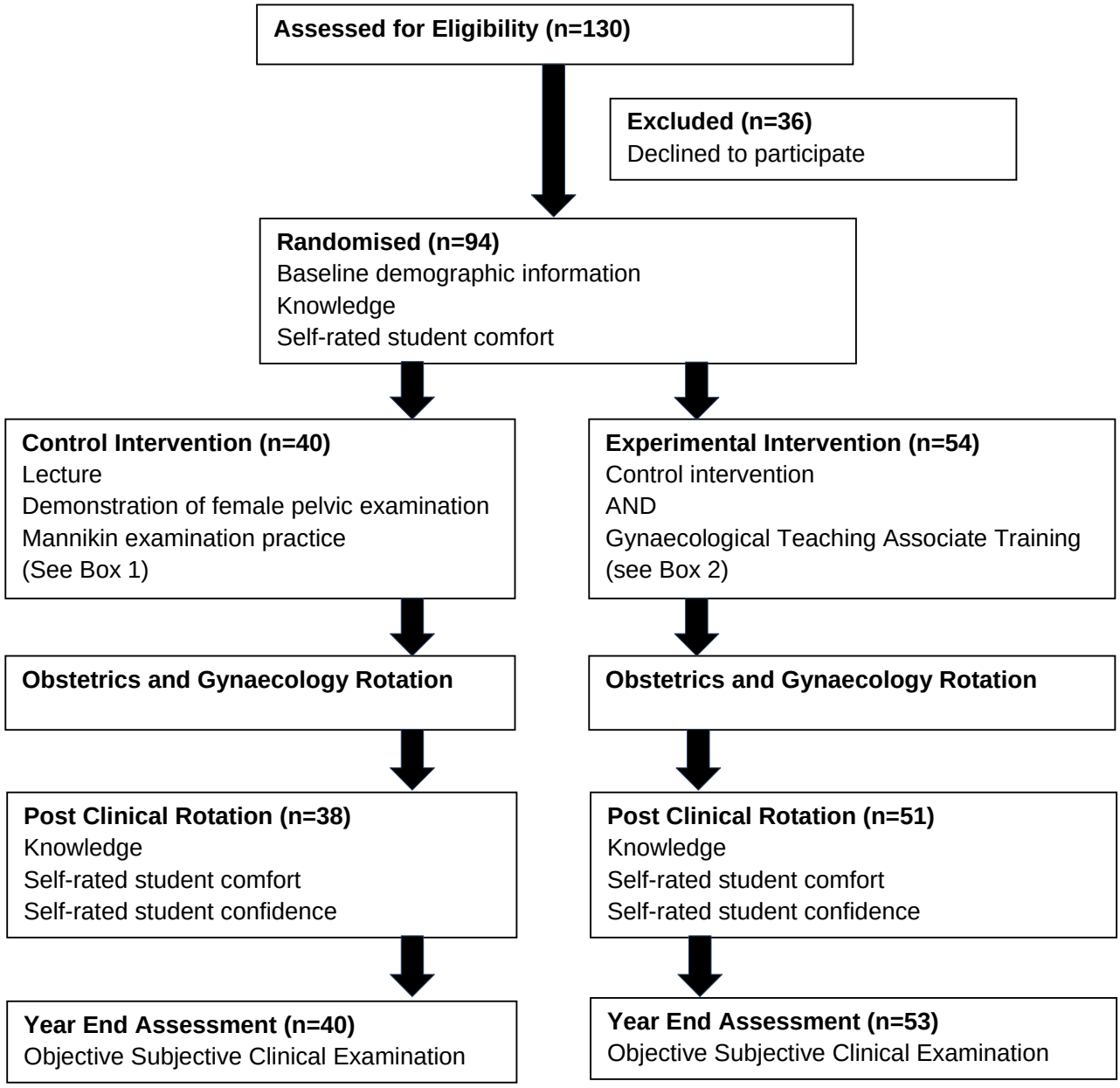

**Table 1**. Baseline Characteristics of Participants

| Characteristic | Control Intervention (n = 40) | Experimental intervention (n = 54) |
|---|---|---|
| **Age, median (IQR)** | 24 (22; 26) | 23 (22; 26) |
| **Women, n (%)** | 24 (60) | 29 (53.7) |
| **Ethnicity, n (%)** | | |
| **White** | 21 (52.5) | 27 (50.0) |
| **Asian** | 17 (42.5) | 27 (50.0) |
| **Black** | 2 (5.0) | 0 (0.0) |
| **Additional graduate degree (Yes), n (%)** | 15 (37.5) | 25 (46.9) |
| **Failed a Course Component (Yes), n (%)** | 4 (10.0) | 5 (9.3) |
| **International Student (Yes), n (%)** | 3 (7.5) | 4 (7.4) |
| **English First Language (Yes), n (%)** | 33 (82.5) | 42 (77.8) |
| **Time from intervention to primary outcome assessment (months), mean (SD)** | 5.6 (1.0) | 5.3 (1.3) |

Abbreviations: IQR: interquartile range; SD: standard deviation.

**Table 2**. Effect of gynaecological teaching associate delivered training on knowledge and student comfort.

| | Control Intervention (n=38) | | Experimental intervention (n=51) | | Difference (95% CI) | p-value |
|---|---|---|---|---|---|---|
| | Baseline | Post-Placement | Baseline | Post-Placement | | |
| **Knowledge (Yes)** [a] | | | | | | |
| n (%) | 3 (7.5) | 8 (21.1) | 2 (3.7) | 27 (50.9) | 29.9 (11.2; 48.6) | 0.002 |
| **Student Comfort** [b] | | | | | | |
| Overall | 10.6 (2.5) | 12.7 (1.6) | 10.7 (2.4) | 14.6 (1.4) | 1.8 (0.6; 3.0) | 0.004 |
| Q1 | 3.5 (0.7) | 3.6 (0.5) | 3.6 (0.7) | 3.9 (0.3) | | |
| Q2 | 2.5 (0.9) | 3.2 (0.5) | 2.6 (0.6) | 3.7 (0.5) | | |
| Q3 | 2.1 (0.9) | 3.0 (0.7) | 2.1 (0.8) | 3.6 (0.5) | | |
| Q4 | 2.5 (0.7) | 2.8 (0.7) | 2.4 (0.9) | 3.4 (0.6) | | |

Abbreviations: CI, confidence intervals.

[a] Knowledge (see methods for details) was scored as yes if the student correctly ordered the components of the pelvic examination. It is summarised as n (%). Difference in knowledge is estimated as the between group absolute difference in these proportions.

[b] Student comfort (see methods for details): Q1: Palpating the abdomen; Q2: Inspecting the external female genitalia; Q3: Separating the labia majora and inserting fingers into the vagina; Q4: Talking to a patient while performing the examination. Student responded to these questions on a 4 point Likert scale from 1: very uncomfortable, 2: uncomfortable, 3: comfortable, and 4: very comfortable. Data expressed as means (standard deviation).

**Table 3**. Effect of gynaecological teaching associate delivered training on skills and student confidence

| Questionnaire | Control Intervention (n= 40) | Experimental Intervention (n=53) | Median difference (95% CI) | p-value* |
|---|---|---|---|---|
| **Skills** [a] | | | | |
| Overall | 43 (37; 46) | 44 (40; 48) | 2 ( -1; 4) | 0.260 |
| Technical | 22 (20; 26) | 24 (21; 27) | 1 (-1; 3) | 0.290 |
| Communication | 19 (17; 22) | 20 (17; 24) | 1 (-1; 3) | 0.353 |
| **Confidence**[b] | (n=38) | (n=51) | | |
| Overall | 17 (15;18) | 18 (18; 18) | 1 (0; 2) | <0.001 |
| Q1 | 3 (2; 3) | 3 (3; 3) | | |
| Q2 | 3 (2; 3) | 3 (3; 3) | | |
| Q3 | 3 (3; 3) | 3 (3; 3) | | |
| Q4 | 3 (2; 3) | 3 (3; 3) | | |
| Q5 | 3 (3; 3) | 3 (3; 3) | | |
| Q6 | 3 (3; 3) | 3 (3; 3) | | |

Abbreviations: CI, confidence intervals.
[a]Skills (see methods for details): measured by objective structured clinical examination scored by two trained blinded observers. Overall skill score (0-54), technical skills (0-28), and interpersonal skills (0-26). Median difference and 95% confidence intervals calculated and analysed by the Mann-Whitney test *.
[b]Student comfort (see methods for details):Q1: Were you adequately prepared to perform a pelvic examination?; Q2: Were you confident that you would not hurt the patient?; Q3: Were you confident explaining the pelvic examination?; Q4: Did you have the necessary communication skills for pelvic examination?; Q5: Were you confident that you could make her feel comfortable and at ease?; Q6: Were you confident in requesting consent from the patient?. Student responded to these questions on a 3 point Likert scale from1: No, 2: Unsure, and 3: Yes. Median difference and 95% confidence intervals calculated and analysed by the Mann-Whitney test *.