Measuring competence: Validating a new Pelvic Assessment Tool (The PEAT Study)

Aisha Janjua[2], Justin Chu[2], Paul Smith[2], T Justin Clark[2]

Corresponding author: Aisha Janjua aishajanjua@doctors.net.uk
Institution: 1. Birmingham Woman's Hospital, 2. Birmingham Women's Hospital
Categories: Assessment, Research in Medical Education, Students/Trainees

Received: 07/11/2016
Published: 25/11/2016

Abstract

Introduction

Medical students are required by the General Medical Council to be proficient in performing a full female pelvic examination (FPE) prior to graduating as doctors. However, many UK medical schools do not conduct a formal summative assessment. Exploration of three electronic databases found no citations evaluating an assessment tool for pelvic examination.

Methods

The ‘PEAT’ was developed with an evidence-based literature search and consists of six domains, evaluating attitudes; inspection; bimanual palpation; adnexal examination; speculum; communication; and a global score using a 10cm visual analogue scale. 60 participants including final year medical students and junior doctors were divided into three groups of 20 according to their experience in performing gynaecological examination. Each participant performed a pelvic examination with a role player and pelvic model, which was video recorded. 20 Consultant assessors, who were blinded to the level of experience, used the PEAT for 15 video clips with a mix of participants from each group. Assessors also completed a questionnaire about the utility, relevance and face validity of the PEAT. The inter-assessor reliability of the PEAT was calculated using Cronbach's alpha. Statistical analysis was performed for continuous variables using the Student t-test.

Results

8/20 (40%) assessors returned their assessment packs. There was a significant increase in the mean score in the ‘inspection’ domain for novice compared with experienced students (p=0.02). A significant decrease was seen in the mean score for ‘adnexal examination’ for novice and experienced students compared with junior doctors (p=0.04;p=0.05 respectively). The mean assessor score for the ease of use and perceived effectiveness of PEAT was 7.4 (SD ±1.32) and 6.6 (SD ±1.19) on a 10cm VAS respectively. The PEAT had good or excellent internal
consistency (α≥ 0.7) in 4/8 assessors (50%).

Conclusions

The PEAT appears to be easy to use and have face validity. However, further refinements are necessary for this PEAT to establish construct validity and improve reliability.

Keywords: Medical education; Medical students; Female pelvic examination; Validated exercises; Reliability

Introduction

Medical students are required by the GMC to be proficient in performing a full female pelvic examination prior to graduating as doctors. Iyengar et al. found that some students were graduating without having performed a female pelvic examination on a conscious patient. At present, in many UK medical schools a formal summative assessment of gynaecological examination is not undertaken. At medical schools where it is, this is based upon the use of manikins using non-validated assessment tools. In some UK Medical Schools such as the University of Birmingham (UoB), evaluation of the medical student's competence at pelvic examination is done by subjective faculty assessment during their obstetrics and gynaecology (O&G) placement. This assessment is typically performed at the end of the teaching placement and is based on the assessors' recollection of the students' performance. This kind of assessment has been shown to have poor validity and reliability. This is in contrast to objective structured assessments of technical skill (OSATS), which use task specific and global rating scales and have been used to evaluate a wide range of clinical and surgical skills.

After searching three electronic databases (Medline, Embase and CINAHL), no citations were identified that evaluated an assessment tool for pelvic examination. Thus, there is a need to develop an assessment tool for female clinical pelvic examination. Van der Vleuten described five criteria that determine the usefulness of an assessment method: namely reliability, validity, acceptability to learners and faculty, impact on future learning and practice, as well as costs. We designed a pilot study with the objective of evaluating the first two stated criteria of reliability and validity of a pelvic examination assessment tool (PEAT).

Methods

Study population

We designed a cross sectional study to obtain provisional data pertaining to the potential reliability, utility and validity of a new PEAT for use in undergraduate medical education. The 60 study participants were divided into three groups of 20 according to their experience in performing gynaecological examination: (i) 'novice student' (defined as a medical student in week one of their O&G clinical placements); (ii) 'experienced student' (defined as a medical student having completed their five week clinical placement in O&G with signed-off competency to perform pelvic examination) and (iii) ‘competent practitioner’. This latter group consisted of junior doctors in their 1st and 2nd year of speciality training in O&G. For the purposes of our study we restricted competent practitioners to relatively junior trainees because it was felt that they were closer in age and appearance to the target student population. Medical students and junior doctors were approached by the Director of the Clinical Teaching Academy at the Birmingham Women’s Hospital (BWH) during the beginning of their O&G placement (TJC). They were made
aware that outcome and assessment as part of the trial would not be incorporated into their final academic grade or postgraduate performance reviews.

Each participant was asked to fill in a pre-assessment questionnaire to assess the comparability of participant groups. The following baseline demographic data were collected: age; gender; ethnicity; confidence in examination assessed using a 10cm VAS score; number of previously performed examinations; and an interest in a future career in O&G.

**Clinical scenario and pelvic examination**

Each participant was presented with the same clinical scenario which required the performance of a pelvic gynaecological examination. The written information provided instructions of the tasks to be completed and to provide a summary of findings to the examiner at the completion of the examination (Figure 1). A role player was used to simulate the verbal responses of a real patient. The written instructions informed the participants that they were to perform the pelvic examination on the manikin provided after positioning the role player as if they were a real patient. The scenario (including examination of the manikin) was video recorded and each recording was sent to three assessors unfamiliar with the participants to ensure that they were blinded to the participants’ level of experience. Each assessor received 15 recordings each. Within these there was an equal mixture of recordings in a random order from all three participant groups.

| Figure (Fig.) 1: Instructions to Candidate during PEAT Study |
|------------------------------------------------------------|
| **Setting**                                               |
| You are a doctor in the Gynaecology Outpatient Department. Please treat this as a real life situation. |
| **Clinical Details**                                       |
| Mrs Helen Smith has presented with painful periods.       |
| **Task**                                                  |
| Perform a speculum and bimanual vaginal examination. Do not perform an abdominal examination. |
| You will be presented with a role player (Mrs Smith) and a pelvic module. You are to address and position the role player but perform the pelvic examination on the manikin. |
| When / If you visualise the cervix please inform the assessor. |
| **Questions the Assessor will ask you at the end of the examination** |
| Summarise / present your findings.                        |
Assessment

The PEAT (Figure 2) was formulated using points that were considered important by members of the clinical undergraduate O&G teaching faculty at the BWH and UoB after consultation with gynaecologists, patients and the available medical literature. The PEAT (Figure 2) consisted of six domains assessing: (1) attitudes; (2) inspection; (3) bimanual examination; (4) adnexal examination; (5) speculum examination; (6) post examination communication; and an additional overall global score. Each domain was assessed using a 10cm visual analogue score (VAS).

Figure 2: The pelvic examination assessment tool (PEAT)

The assessors were sent a pack with a cover letter explaining the purpose of the PEAT study including the instructions given to the candidate. They were advised to use PEAT to assess the examination skills of the 15 participants, with the videos provided on a memory stick. To ensure confidentiality, the memory sticks were locked, requiring a password that was allocated to them by the researcher sent in the post. They were also sent 15 assessment tool marking sheets, a feedback form and a stamped addressed envelope to return all study material. The feedback form used VAS scoring to investigate whether the PEAT was easy to use, and whether the assessor considered this an effective tool to assess competence. There was a space for other comments that allowed assessors to give any additional feedback on PEAT. Themes were considered to have emerged where more than one respondent provided
similar feedback in the open responses.

The validity of an assessment tool provides an indication as to whether the test is measuring what it is intended to measure. We hypothesised that higher PEAT scores would be obtained for the participant groups with greater experience, i.e. junior doctors would score highest in all facets of the pelvic examination compared with novice medical students and possibly compared to experienced medical students. The feedback form completed after the assessments by the examiners was intended to determine face validity. The reliability of an assessment tool refers to its ability to give reproducible results. We examined inter-assessor reliability by observing agreement of two or more assessors evaluating the same video assessment.

Statistics

The sample size was influenced by the constraints imposed by the number of the trainees in O&G at the Birmingham Women’s Hospital (BWH) which is the main academic teaching institution delivering undergraduate medical education in O&G on behalf of the UoB.

The validity and reliability of the PEAT were considered to be of equal importance and were the primary outcomes assessed. Validity was tested using the proxy measure of construct validity; this was demonstrated with scores derived from the PEAT reflecting the participant experience. Reliability was measured using inter-assessor reliability between assessors who were blinded to the other assessors’ marks. Statistical analysis was performed for continuous variables using mean, standard deviation and standard error of mean. Cronbach’s alpha was used to test reliability (inter-assessor variability) using a two way mixed, consistency, average-measures intra-class correlation for each element of PEAT (attitudes; inspection; bimanual examination; adnexal examination; speculum examination; post examination communication; and an additional overall global score) with p>0.7 considered good. We also tested whether the baseline characteristics of the medical students (novice and experienced) were predictive of student global performance as scored by the independent assessors. The independent variables: gender, age, ethnicity, confidence in examination and expressed interest in a future career in O&G, were tested in univariable and multivariable analyses.

Results

The background demographics are shown in Table 1. The distribution of gender and ethnicity were comparable across the three participant groups. The level of interest in a career in O&G was comparable between the two student groups as was the age distribution whereas these demographics were higher in specialist trainees in O&G. Confidence in pelvic examination was highest in junior doctors in O&G, followed by experienced students and lowest in inexperienced students. No student had performed more than 10 gynaecological examinations prior to the study in contrast to specialist trainees in O&G who had all done so (Table 1).
10 out of the 20 assessors returned their feedback forms but only 8/20 (40%) returned their completed student assessment packs although 1/8 assessors did not use the VAS to evaluate the participants performance and so could not be included in analysis.

**Construct Validity:**

There were two significant results supporting construct validity. There was a significant decrease in the mean score in the adnexal examination domain for the novice student versus the junior doctor (mean difference -8.0, 95% CI [-18.8, 2.8], p=0.04] and for experienced students versus junior doctors (mean difference -9.4, 95 CI [-18.8, -0.1],
p=0.05). However, there was a significant increase in the mean score in the inspection domain for the novice versus the experienced students (mean difference 13.2, 95% CI [2.4, 24.1], p=0.02) opposing the construct that more experienced students and junior doctors should score higher than novice students. No other significant differences were noted between participant groups for all other domains (Table 2).

### Table 2: Construct Validity: discriminative capacity of the pelvic examination assessment tool (PEAT) for participants with different levels of experience

| Grade          | n  | Mean (SD)    |
|----------------|----|--------------|
| **Global Assessment** |    |              |
| Novice student  | 29 | 59.7 (19.3)  |
| Experienced student | 37 | 55.5 (17.9)  |
| Junior doctors  | 34 | 59.4 (15.7)  |
| **Attitudes**   |    |              |
| Novice student  | 31 | 62.9 (23.8)  |
| Experienced student | 37 | 62.6 (19.6)  |
| Junior doctors  | 34 | 63.8 (20.4)  |
| **Inspection**  |    |              |
| Novice student  | 30 | 56.5 (21.1)* |
| Experienced student | 37 | 43.3 (22.8)**|
| Junior doctors  | 32 | 43.5 (26.3)  |
| **Bimanual Examination** |    |              |
| Novice student  | 29 | 58.3 (19.1)  |
| Experienced student | 36 | 55.7 (17.6)  |
| Junior doctors  | 33 | 61.7 (14.0)  |
| **Adnexal examination** |    |              |
| Novice student  | 29 | 51.4 (24.6)**|
| Experienced student | 36 | 50.0 (20.8)**|
| Junior doctors  | 33 | 59.4 (17.8)*****|
| **Speculum**    |    |              |
| Novice student  | 29 | 62.4 (17.8)  |

**Face validity:**

The mean score for ease of use of the PEAT was 7.4 (range 5.2 – 9.7) on a 10cm VAS derived from the 10 assessors completing the feedback forms. The assessors mean perception of the effectiveness of the PEAT in assessing
competence in gynaecological examination was 6.6 (range 4.7 – 8.5). Themes from examiner open feedback included the view that the order of the examination was not accurately reflected in the PEAT because the domain evaluating "speculum examination" came after the domains evaluating bimanual examination (Figure 2). Another theme was the ease of use with one examiner commenting that the PEAT was "a nice simple tool".

Reliability:

The inter-rater reliability was assessed using a two-way mixed, consistency, average-measures intra-class correlation (McGraw and Wong 1996). For the global score the intra-class correlation was in the excellent range 0.83 indicating that the assessors had a high degree of agreement. For the domain of speculum examination the Cronbach's alpha score was questionable at 0.65 and for the other elements of PEAT the intra-class correlation was poor: attitudes = 0.15; inspection = 0.56; bimanual examination = 0.32; adnexal examination = 0.37; post examination communication = 0.15)

Discussion

Main findings

In our study we were unable to establish construct validity for the PEAT. There were no consistent differences between the three groups representing different experience levels in female pelvic examination. The PEAT demonstrated construct validity for adnexal examination, where the most experienced group, namely junior doctors, performed this aspect of examination better compared with novice and experienced medical students. However, no differences were observed between the levels of experience between student groups within this domain. Furthermore, in contrast to our expectation, novice students scored more highly in the inspection domain compared with their more experienced counterparts; experienced students and junior doctors.

Face validity was demonstrated for the PEAT with examiners considering the tool effective in measuring competence in gynaecological examination. Moreover, examiners reported the PEAT to be comprehensible, quick and easy to use. The reliability of the PEAT appeared to be excellent for the global score, but poor for the individual elements of the assessment. The individual elements of PEAT could be removed from the tool but the consideration of each element maybe important for informing the assessor when considering the global score of a participant.

This study demonstrates that when considering reliability, the global score should be used in preference to the other elements for hypothesis testing, although inferences should be made with caution due to the small number of respondents. The PEAT in its current form appears to have some strengths, especially when considering its utility and face validity. However, revision is needed to demonstrate construct validity and improve reliability before the PEAT can be routinely used in clinical education and formative assessment.

The useability of the PEAT expressed by the assessors may reflect its design. By incorporating a VAS for marking each domain we hoped to harness the sensitivity of a continuous scale and produce an easy, rapid way of scoring performance. We expected most clinicians to be familiar with VAS for measuring clinical outcomes and so anticipated that proficiency in using the PEAT would be quickly achieved. In addition, by explicitly defining the components to be considered when deciding where to score each domain within the PEAT, we hoped to optimise its
user-friendliness thereby aiding its validity and reliability. In a clinical examination setting, standardised assessment instruments need to be clear and easily completed to score candidates in real time reducing the likelihood of recall bias and also to enhance their practical use especially where a large number of candidates are to be assessed.

The failure of the PEAT to consistently discriminate between the relative experiences of candidates in gynaecological examination suggests that the tool requires considerable refinement before it can be widely adopted. However, it is possible that the design of our study was biased in favour of the medical students for several reasons. Firstly undergraduates in their final year have generally become well versed in OSCE style assessments because they are commonly used in assessments across different specialties. In contrast, junior doctors at ST1 and ST2 levels may have become less familiar with this type of assessment. Secondly, whilst our choice of using role players combined with manikins for examination is representative of assessment methods for intimate examination used by the majority of UK medical schools (Janjua and Clark, unpublished), it may have biased against more experienced junior doctors who undoubtedly would be more familiar with examining real patients. Thirdly, the pelvic manikin used in our chosen scenario had no genital tract pathology. It is possible that the presence of pathology may have yielded different results with more experienced participants displaying higher levels of competence. Fourthly, our population of students were in their fifth and final year and so many may have already gained some experience of pelvic examinations e.g. during their sexual health attachments in their fourth year, within general practice placements or during their student selected modules. Therefore, the baseline experience of students will have varied and the 'novice student' may potentially have more experience in performing gynaecological examinations than those deemed an 'experienced student', having completed their final year O&G placement.

Another explanation for the apparent lack of construct validity of the PEAT could relate to examiner training. A greater familiarity with the assessment tool would lead to improvements when assessing pelvic examinations. Face-to-face training to improve understanding of the tool would have been ideal. A final bias to consider is that of selection. Participation in this study was voluntary and so it is plausible that more self-confident students were recruited thereby overestimating average performance. However, to some degree these arguments apply to participating junior doctors and moreover confidence does not necessarily correlate with clinical skills. It is therefore possible that our results are not externally valid in students who possess a lower level of initiative, and further studies may need to consider using accounting for personality traits or designs whereby the full student cohort could be evaluated.

Strengths and limitations

Our study is novel because to our knowledge it is the first to attempt to validate a PEAT for use in evaluating competence in gynaecological examination in either undergraduate or postgraduates. The methodology to establish construct and face validity, as well as reliability was robust. A reasonable number of novice and experienced students, as well as junior doctors were recruited to this study to establish validity and reliability of the PEAT. The task required to be undertaken by candidates was standardised with identical instructions provided to all participants. We stipulated no time limit for the examination and inappropriate content that was unrelated to pelvic examination skills. We recruited all assessors from outside our University Medical School and post-graduate training region in order to ensure blinding of assessors to the level of experience of the participants in the study thereby enhancing objective impartial assessment.

Limitations of our approach include the fact that pelvic examination was performed on a manikin rather than a real patient. We tried to make the encounter as realistic as possible with the presence of a role player to provide verbal
responses during the examination of the manikin. Despite this, it is unlikely that this setting neither replicated the experience of examining a real patient nor induced similar levels of anxiety. To create a more realistic examination, future studies could consider the use of expert or simulated patients (gynaecology teaching associates or ‘GTAs’). However, given that most medical schools do not employ GTAs, any PEAT should be valid for use in simulated manikin-based scenarios and generalisable. Another limitation of our study is that we used a one-off assessment to establish competence. One could argue that using serial assessments looking at improvements in the VAS scores would give a superior indicator of competence as this would provide a formative method of assessment. However, summative assessments of core clinical examination skills are still common components of final medical school examinations and validated measuring instruments such as the PEAT should be developed for such purposes.

Our sample size was influenced by the constraints imposed by the number of the undergraduate and postgraduate trainees in O&G at the BWH. This limited the number of participant assessments, thus limiting the videos we could send out to test construct validity. Furthermore we had a low response rate, with only 40% of assessors returning their assessment sheets on the participants impacting adversely on the power of the study to establish validity.

Implications for practice and research

Assessment in postgraduate medical education has a better evidence base than in undergraduate education although we are aware of no validated tools for evaluating proficiency in female pelvic examination. In the postgraduate years, performance is assessed by not only by senior clinicians, but also by patient, peers and other members of the team e.g. secretaries in multi-source feedback (MSF). Mini-clinical evaluation exercise (mini-CEX) is commonly used in postgraduate education for observation of history taking, clinical skills as well as general attitude and professional behaviour. Similarly, directly observed procedural skills (DOPS) are used to assess and provide feedback on particular skills e.g. injections. Although these methods were initially created for assessment within a particular specialty, they are now more widely available, and applicable to other specialties such as O&G. Previous studies in O&G have used observed structured assessment tools (OSATS) for establishing competence in postgraduate surgical skills and have demonstrated construct validity. The PEAT we developed is a type of OSAT and with further revision we hope to show validity and reliability in both an undergraduate and postgraduate population. If developed it could be used for both formative assessment (where competency can be attained ideally over a period of time during the clinical placement) and summative assessment.

Many clinical competencies tested in undergraduate medical students are judged by methods which have not gone through rigorous psychometric testing. There is a need for valid, useable instruments for evaluating core clinical skills. In our study we attempted to develop a PEAT for assessing competence in gynaecological examination. However, qualitative work to obtain the views of students, patients, educationalists and clinicians may aid the development of a more valid and reliable PEAT. Pilot studies, such as the current one, can then be undertaken to identify the potential reliability and validity of a refined PEAT. In addition, the derived data can be used to help inform the design and size of future, larger-scale studies. Such studies should incorporate more assessors to better evaluate psychometric performance of the revised PEAT before it can be introduced into routine educational practice.
Conclusion

Training in obstetrics and gynaecology is integral to the undergraduate medical school curriculum. Achieving experience in gynaecological examination can be challenging given the time constraints of placements and the intimate nature of the clinical examination. Thus, innovations in training and assessment are urgently needed to ensure medical undergraduates qualify with the necessary competence in this core clinical skill. To our knowledge, this is the first study to try to develop and validate a pelvic examination assessment tool which ideally can be used in both undergraduate and postgraduate medical education both formatively and summatively. The PEAT we tested demonstrated utility and face validity but further refinements are needed to establish reliability and construct validity in larger scale studies inclusive of strategies to optimise response rates of assessors.

Take Home Messages

1. Medical students are required by the General Medical Council to be proficient in performing a full female pelvic examination prior to graduating as doctors.

2. Exploration of three electronic databases found no citations evaluating an assessment tool for female pelvic examination.

3. The new PEAT (PElvic Assessment Tool) consist of six domains, evaluating attitudes; inspection; bimanual palpation; adnexal examination; speculum; communication; and a global score.

4. The PEAT appears to be easy to use and have face validity.

5. The PEAT had good or excellent internal consistency (reliability).

Notes On Contributors

Aisha Janjua (MBBS BMedSci, DFSRH, Cert MedEd, MRCOG) is a Specialty Trainee Year 6 in Obstetrics and Gynaecology (Birmingham Heartlands Hospital, U.K.)

Justin Chu (MBChB, MRCOG, PG Cert HP) is an Academic Clinical Lecturer in Obstetrics and Gynaecology (Birmingham Women’s Hospital, U.K.)

Paul Smith (PhD MBChB (hons) BSc (hons)) is a NIHR Post Doctoral Research Fellow and Specialist Registrar in Obstetrics and Gynaecology (Birmingham Women’s Hospital, U.K.)

Professor T Justin Clark is a Professor of Obstetrics and Gynaecology (University of Birmingham), Head of Academy and Consultant Obstetrician and Gynaecologist (Birmingham Women’s Hospital NHS Foundation Trust, U.K.)

All authors meet the criteria for authorship, in detail: TJC, JC and PS designed the study. PS and JC conducted the study. AJ collected and analysed the data under the supervision of PS and TJC. AJ wrote and revised the manuscript. TJC, PS and JC revised the manuscript for important intellectual content.
Acknowledgements

We would like to thank all participants and the following assessors: Sadia Malick; Andrew Thompson; Alison Wright; Padma Pathi; Jonathan Chamberlain; Edward Morris; Kevin Cooper; Mary Connor.

Disclosure of interests

For all authors there are no financial or other conflicts of interest.

Funding

No funding was received.

Details of ethics approval

Not applicable.

Bibliography/References

1. General Medical Council (GMC), 2015. Outcomes for Graduates pages 4 – 5 [Online Article] Available at: http://www.gmc-uk.org/Outcomes_for_graduates_Jul_15.pdf_61408029.pdf [Accessed 31.3.2016]

2. Iyengar S., Raut N., Clark J. Gynaecology training associates (GTAS): the future in the training of final year medical students? International Journal of Gynecology and Obstetrics, 2012; 119:S377.

https://doi.org/10.1016/S0020-7292(12)60760-7

3. Reznick R.K. Teaching and testing technical skills. Am J Surg, 1993; 165(3), 358-361.

https://doi.org/10.1016/S0002-9610(05)80843-8

4. Cremers S.L., Lora A.N., Ferrufino-Ponce Z.K. Global Rating Assessment of Skills in Intraocular Surgery (GRASIS). Ophthalmology, 2005; 112, 1655-60.

https://doi.org/10.1016/j.ophtha.2005.05.010

5. Goff B.A., Lentz G.M., Lee D., Hounard B., Mandel L.S. Development of an objective structures assessment of technical skills for obstetric and gynecology residents. Obstet Gynecol, 2000; 96, 146-150.

https://doi.org/10.1016/S0029-7844(00)00829-2

6. Grober E.D., Hamstra S.J., Wanzel K.R., Reznick R.K., Matsumoto E.D., Sidhu R.S., Jarvi K.A. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. Ann Surg, 2004; 240, 374-381.

https://doi.org/10.1097/01.sla.0000133346.07434.30

7. Lentz G.M., Mandel L.S., Goff B.A. A six-year study of surgical teaching and skills evaluation for
obstetric/gynaecologic residents in porcine and inanimate surgical models. Am J Obstet Gynecol, 2005; 193, 2056-2061.

8. Van der Vleuten C.P.M. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ, 1996; 1, 41-67.

9. Schuwirth L.W.T., van der Vleuten C.P.M. How to design a useful test: the principles of assessment. Association for the Study of Medical Education, 2006; Edinburgh.

10. Cicchetti D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 1994; 6(4), 284–290.

11. McGraw K.O., Wong S.P. Forming inferences about some intraclass correlation coefficients. Psychological Methods, 1996; 1(1), 30–46.

12. Fraser S.W., Greenhalgh, T. Coping with complexity: educating for capability. BMJ, 2001; 323, 799-803.

13. Epstein R.M. Assessment in Medical Education. The New England Journal of Medicine, 2007; 356, 387-396.

14. Academy of Medical Royal Colleges: Improving Assessment, 2009. Section 4.2.5 Step 5: Train Assessors. [Online Article] http://www.aomrc.org.uk/doc_view/49-improving-assessment. [Accessed 25 September 2014].

15. Mavis B. Self-efficacy and OSCE performance among second year medical students. Advances in health sciences education, 2001; 6(2), 93-102.

16. Pieters M.S., Jennekens-Schinkel A., Schoemaker H.C., Cohen A.F. Self-selection for personality variables among healthy volunteers. British Journal of clinical pharmacology, 1992; 33(1), 101-106.

17. McAleer S. Choosing Assessment Instruments. In: Dent, J.A. and Harden, R.M. eds. A practical guide for medical teachers, Edinburgh, London, New York Philadelphia, St. Louis, Sydney, Toronto: Churchill Livingstone; 2005, 302-310.

18. Leach D.C. Competence is a habit. JAMA, 2002; 287, 243-244.
19. Boursicot K.A.M., Roberts T.E., Burdick, W.P. Structured assessments of clinical competence. In: Swanwick, T. ed. Understanding Medical Education: Evidence, Theory and Practice. Oxford, UK: Wiley-Blackwell, 2010; 246-258.

https://doi.org/10.1002/9781444320282.ch17

20. Davis M.H., Ponnamperuma G.G. Work-based assessment. In: Dent JA, Harden RM. eds. A Practical Guide for Medical Teachers. 2nd edn. London: Elsevier Churchill Livingstone, 2005; 336-345.

21. Norcini J.J. The Mini Clinical Evaluation Exercise (mini-CEX). The Clinical Teacher, 2005; 2(1), 25-30.

https://doi.org/10.1111/j.1743-498X.2005.00060.x

22. Norcini J.J., McKinley, D.W. Assessment methods in medical education. Teaching and teacher education, 2007; 23(3), 239-250.

https://doi.org/10.1016/j.tate.2006.12.021

23. Setna Z., Jha V., Boursicot K.A.M., Roberts T.E. Evaluating the utility of workplace-based assessment tools for speciality training. Best practice & research Clinical obstetrics & gynaecology, 2010; 24(6), 767-782.

https://doi.org/10.1016/j.bpobgyn.2010.04.003

24. Goff B.A., Nielsen P.E., Lentz G.M., Chow G.E., Chalmers R.W., Fenner D., Mandel L.S. Surgical skills assessment: a blinded examination of obstetrics and gynecology residents. American journal of obstetrics and gynecology, 2002; 186(4), 613-617.

https://doi.org/10.1067/mob.2002.122145

25. Norcini J.J. Workplace-based assessment in clinical training Edinburgh: ASME, 2007.

26. University of Birmingham, 2015. CANVAS. [Online Article] https://canvas.bham.ac.uk/login [Accessed 6 February 2015].

Appendices

Declarations

The author has declared that there are no conflicts of interest.

This has been published under Creative Commons “CC BY 4.0” (https://creativecommons.org/licenses/by-sa/4.0/)

AMEE MedEdPublish: rapid, post-publication, peer-reviewed papers on healthcare professions’ education. For more information please visit www.mededpublish.org or contact mededpublish@dundee.ac.uk.