

Improving the content and face validity of OSCE assessment marking criteria on an undergraduate midwifery programme: A quality initiative

Objective Structured Clinical Examinations (OSCE's) have been adopted as a means of assessing midwifery students' clinical skills. The purpose of the OSCE is to provide a standardised method for the evaluation of clinical skill performance in a simulated environment. This paper describes how a quality improvement initiative using both internal and external expert review was utilised to improve OSCE assessment marking criteria. The purpose of the quality initiative was to review the content and face validity of the marking criteria for assessing performance. The design and choice of tools used to score students' performance is central to reliability and validity. 20 videos of students from year one of a midwifery preregistration programme undertaking an OSCE assessment on abdominal examination and 18 videos of students response to obstetric emergencies e.g. PPH, and shoulder dystocia were available for review. The quality initiative aimed to strengthen the reliability and validity of the OSCE in assessing student performance. Conclusion: the use of global rating scales allows for the capturing of elements of professional competency that do not appear on specific criteria for skills performance checklists.

Key words

OSCE; global rating scales; marking criteria; student midwives; content validity; face validity.

Background

In undergraduate midwifery programmes a variety of assessment strategies are used to contribute to the currency and relevancy of the programmes and to reflect the core concept of woman centred care. Race and Pickford (2007) suggest that educators need to make systematic and thoughtful use of assessment as the principle motivator of the learning for the majority of students. Race (2005) proposes that the values and principles for assessment design should be valid and so should assess what it is you really want to measure. Assessments should be authentic so that student achievement is measured close to the intended outcomes and that performance skills should be measured in performance not just writing about it in an exam setting. If criteria and marking schemes are right there should be good inter-assessor reliability and good intra assessor reliability (Race, 2005). Harden described the OSCE in 1979 and was the first to focus on performance as a means of assessing medical professional competence (Hodges 2003). Objective structured clinical examinations (OSCEs) have been adopted in many countries as a means of assessing undergraduate nursing and midwifery skills in on campus clinical skills laboratories (Nulty et al., 2011, Mitchell et al., 2009, Wallace et al., 2002). Selby (1995) defines OSCEs as a circuit of stations where clinical skills are assessed by an examiner using previously determined objective marking criteria. It is important that assessments are efficient and manageable so that the demands on students and lecturers are effective in terms of time and resources (Race et al. 2005). OSCEs are time consuming and the challenges and opportunities associated with their use in a university setting have been documented (Brosnan et al., 2006, Noonan et al., 2008). Whilst OSCEs are seen as an objective and reliable means of assessment (Harden and Gleeson, 1979) there are many variations in use across nursing and midwifery which means that the reliability and validity of individual assessments need to be maintained (Rushforth, 2007). As

Hodges (2003) highlights the main concern of all disciplines is a reliable and valid assessment of competence. According to Pender et al., (2004) assessment of skill competence under examination conditions may well have a predictive value for performance in actual practise. However Newble (2004) acknowledges that a below standard performance of a practitioner does not necessarily reflect a lack of professional competence and that skill performance and professional competence should be seen as two different constructs. As part of our commitment to on-going quality improvement process, we sought feedback from lecturers and students (Barry et al. 2012). In the review of our assessment strategy we identified strengthening of our marking criteria and the need for consistency between markers as areas for improvement. The purpose of the quality initiative was to review the content validity and face validity of the marking criteria with the aim of improving the reliability and validity of the OSCE in assessing student performance.

The OSCE assessment process prior to internal and external review

Clinical scenarios that require specific responses were developed. At each station the students' performance was rated by two examiners using a checklist. The checklists required the assessors to tick each element from a list as either 'done' or 'not done'. The examiners were trained in the use of the checklist. The responses of the student were measured against agreed criteria developed by the panel of midwifery lecturers, reflective of best evidence based practice. The marking criteria had been devised with reference to the Advanced Life Support in Obstetrics (ALSO) 2000 Provider Manual, standard Midwifery textbooks and international, national and local guidelines on best practice. The use of highlighted 'compulsory' items was adopted to emphasise the clinical importance of selected checklist items. Omission of these safety related items (e.g. auscultating the fetal heart rate for cord prolapse) would lead to an automatic fail irrespective of the overall percentage achieved.

The process for the quality review

The video recordings of 38 students undertaking their OSCE assessment using the Digital Nursing Archive (DNA) system were available to internal and external experts for review. The DNA system contained within the clinical skills laboratories has multiple uses in relation to teaching and learning strategies and the students are familiar with its use. The video is automatically transferred to the Digital Nursing Archive where the students' performance can be reviewed. Elder (1999) suggests that because video tapes can be viewed repeatedly they provide opportunities for more accurate pictures of the actual target behaviours. 20 students from year one of a midwifery preregistration programme were recorded undertaking an OSCE assessment on abdominal examination and 18 students in year three of the programme recorded responding to obstetric emergencies e.g. PPH, and shoulder dystocia. Permission was sought from the students prior to the internal and external reviewers viewing the video recordings. It was emphasised that this was a quality initiative and was not part of the students' assessment process. The viewing took place some weeks after the students' summative assessments and twelve of the recordings were reviewed internally.

Internal Review:

The panel of midwifery lecturers were asked to evaluate whether the checklists included all pertinent information related to a given scenario and to review the pass and fail criteria for each proposed OSCE. Face validity addresses the question of whether the items (the marking criteria) in the assessment tool (the OSCE) actually measure what they are intended to measure. Content validity in relation to an OSCE refers to the judgements made by a panel of experts about the range to which the content of the examination appears to correctly examine and include the characteristics and domains that it was designed to appraise and assess (Bowling 2002). The video recordings were reviewed independently by the panel with reference to the marking criteria in use.

Outcome of Internal Review

The review of the student performance showed a good level of theory underpinning their performance for instance all students demonstrated a systematic approach to performing abdominal examination with application of the underlying theory. Yet it was noted that the words used by the students when providing an explanation ‘to the woman’ during the assessment were very technical. It was clear from the recordings that emphasis was needed throughout the assessment criteria on the demonstration of a woman centred approach by the students in their interactions with the “woman” It was suggested that as lecturers we needed to emphasise more women friendly language and incorporate strategies into our teaching and learning to increase awareness of this amongst students.

It was felt that the present marking criteria did not guide the students towards demonstrating a holistic approach to the performance of abdominal assessment. For instance all the students indicated that they would gain consent from the “woman” at the outset of the examination thereby meeting the first criteria on the checklist. It was recognised that this technical approach was done for exam purposes and would not be the approach taken in practice. Consent needed to be emphasised as a process and not just an indication by the student that consent had been obtained thus fulfilling a tick box on a checklist.

Having individually reviewed a selection of recordings across both groups the internal review panel concluded that: the pass and fail criteria for each OSCE are clear, comprehensive and unambiguous, thus ensuring the consistency and fairness of marking during an actual OSCE. The team identified that there was an over concentration by the student on meeting each identified criteria and more emphasis was needed on the students overall performance. For instance for the OSCE on PPH the students were focused on remembering the next step rather than the demonstration of performance akin to the clinical context. It was felt that the present criteria didn’t differentiate enough between students who carried out the OSCE in a holistic manner compared to a student who met the identified steps on the checklist. The internal review concluded that revised guidelines needed to emphasise a more women centred approach. Lecturers needed to incorporate strategies for developing the use of more

woman centred language with students. It was decided that a global rating scale and a checklist incorporated into a user friendly document would improve the quality of the assessment process.

External Review

Two external experts, one a senior lecturer in midwifery outside of Ireland and the other an external examiner to one of the midwifery programmes, reviewed six of video recordings independently of each other, to appraise the quality and relevance of the marking criteria. The remit was to review the marking criteria with reference to the student's performance, and while this provided a framework it allowed for discussion and dialogue with the internal panel of midwifery lecturers. Both reviewers identified the high level of knowledge and skills demonstrated by the students across both of the OSCE assessments.

In relation to abdominal examination both reviewers highlighted the technical nature of the language used, such as uterus, fatal parts, syphilis pubis, and vertex when explaining to 'the woman'. Opportunities to engage with 'the woman' as the procedure continued were not availed of by the students. Greater emphasis was needed on explanation and revisiting consent as part of the holistic care of the woman. The external reviewers commended the use of OSCEs as a means of assessment of abdominal examination and recommended that equal emphasis should be placed on the language the student uses in the simulated setting. The difficulty of gaining a balance between the demonstration of technical skills and an emphasis on woman centred care was discussed with both reviewers. Following discussion it was decided to adopt a consistent approach to providing guidance for the OSCE on abdominal examination whereby first year students would go through the abdominal examination first 'with the woman' and then provide a synopsis at the end to the examiners. Thus the 1st year student can demonstrate technical skills yet more emphasis is placed on woman centred care without the distraction of communicating with the assessors during the demonstration. Practical suggestions were also given to reduce the number of OSCEs the students had to perform for the obstetric complications from three to two so that the student had the opportunity to provide a more in depth performance. The opportunity to critique our assessment marking criteria with external experts helped to provide new insight and to make simple changes that help guide the students towards providing a woman centred approach to care.

Discussion

This quality initiative process identified that the check list needed to include direction for a more holistic approach to the performance assessment. Cox (1990) argues that checklists are insufficiently sensitive to detect higher clinical components such as empathy, rapport and ethics. All of these along with qualities such as respect for the patient and humanism are important components of Professionalism (Mazor, et al. 2007). Some of the aspects of professionalism can and should be observed during an OSCE assessment e.g. how the student communicates with the woman while responding to different situations such as postpartum haemorrhage and shoulder dystocia. Nulty et al. (2011) have developed best practice guidelines for the use of

OSCES in nursing education with the use of a holistic marking guide allowing student performance to be related to clinical practice rather than be judged on a set of discrete independent actions. Fink (2003) proposes a taxonomy of significant learning which is relational rather than hierarchical. As well as core concepts of foundational knowledge, application, integration, others such as, the human dimension, caring, and learning how to learn, are included. Finks believes that a synergy can be created, by achieving one dimension it enhances the possibility of the other dimensions being achieved. According to Richter and Sinclair (2005) the use of a global rating scale is beneficial as it includes the concepts of application, integration and the human dimension and it helps to anchor abstract skills to behavioural objectives (Richter and Sinclair 2005).

Nulty et al (2011) in their review highlight the usefulness of checklists in providing feedback for formative assessment of OSCEs with use of global ratings for summative assessments as reliable and providing a more holistic view. Major (2005) in an evaluation of a seven year OSCE programme suggests that it is important to have a simple skills breakdown criterion referenced evaluation within the holistic patient encounter. A combination of a check list and a global rating scale is seen as having greatest value in contributing to consistency in grading students in an OSCE assessment (Rushforth, 2007). Marjan et al. (2002) in their study showed that there was low reproducibility of checklist scores across tasks. Jones et al. (2010) make the point that it is important for the content validity of the OSCE station that the marking criteria relate only to the skill that is being assessed, to identify those students who can/cannot perform a skill such as blood pressure. They suggest that as the student progresses methods of identifying discrete skills must be included within the overall care of the patient. In developing their guidelines they also support the use of global ratings as it provides markers with the opportunity to comment on individual performance. Examiners in the Marjan et al. study (2002) stated that global rating scales allowed for capturing elements of professional competency that did not appear on the checklist. These authors suggest that while checklists are highly content specific global ratings allow for a broader range of skills necessary for good professional practice. In Rushforth's (2007) review a key aspect of reliability explored in studies, is the accuracy of judgements made by examiners which is frequently reliant on single examiners. Jones et al. (2010) suggest that it is important to establish whether there is a correlation between global ratings and the mark achieved. In our review the issue of having two markers at each station was raised by both the internal and external midwife experts as a positive and was perceived to help address the inter assessor reliability. This is not always achievable and may be a challenge for resource allocation with OSCEs.

Implementation

Following the internal and external review and a review of the literature the following changes were implemented into our OSCE assessment process and documentation. The core concepts of foundational knowledge, application, integration, and the human dimension (Fink, 2003) were incorporated into our assessment framework see Table 1.

Table 1 here

A global rating scale (Table2) was linked to grading bands and was incorporated into the marking criteria to capture comprehensively the students overall performance (Table 3). If some aspects of the student's performance are at different levels then the assessors judge the overall level based on the student's performance during the totality of the assessment with reference to the core concepts. Key criteria must be met (all associated with safety in practice), if not this results in a fail of the assessment. While the bands appear broad it is possible to fail on a single criterion and it is possible for the assessors to fail a student based on their overall judgement.

Table 2 here

Table 3 here

Conclusion and recommendations

The use of global rating scales allows for the capturing of elements of professional competency that do not appear on a checklist (Marjan et al., 2002). Tanner (2006) states that in the development of clinical judgement skills, educational practices must ensure that students engage fully with patients in the provision of care. Aspects of professionalism can and should be observed during an OSCE assessment along with meeting specific criteria for skills performance.

Considering the many variations of OSCES used in Nursing and Midwifery (Rushforth, 2007) it is important that the OSCE process should be open to debate and empirical. A quality initiative using internal and external expert review informed by the literature can improve OSCE assessment marking criteria and encourage lecturers to scrutinise assessment strategies. This paper has described the process of reviewing the content validity and face validity of the marking criteria for using the examples of abdominal examination and obstetric emergencies. The purpose of the OSCE is to provide a standardised method for the evaluation of clinical skill performance in a simulated environment. This paper provides an example of user friendly specific marking criteria and global rating scales for use by students and lecturers on the undergraduate midwifery programme. There is a growing body of knowledge on the use of OSCES for the assessment of student midwives. The opportunity exists now for collaboration to provide a standardised approach to the assessment of core competencies for midwifery education. The standardised assessment of obstetric emergencies using agreed marking criteria and global rating scales would be an efficient and effective means of educating all student midwives. The use of specific criteria and global rating scales help to ensure that OSCES provide a robust assessment of performance and are therefore valuable in maintaining safety of practice and standards of public protection. The challenge now is one of collaboration on the development of best practice guidelines for OSCE assessment on the undergraduate midwifery programmes.

References:

Advanced Life Support in Obstetrics (ALSO) 2000 Provider Manual (4th Edn) ALSO, New Castle upon Tyne

Barry, M, Noonan, M., Murphy-Tighe, S., 2012. An exploration of student midwives experiences of the Objective Structured Clinical Examination assessment process, *Nurse Education Today*, 32, 6, 690-694.

Bowling, A, 2002. *Research methods in health: investigating health and health services*. Buckingham: Open University Press.

Brosnan, M, Evans W, Brosnan E, Brown, G, 2006. 'Implementing objective structured clinical skills evaluation (OSCE) in nurse registration programmes in a center in Ireland: A utilization focused evaluation'. *Nurse Education Today*, 26 (2), 115-122

Cox, K, 1990. No Oscar for OSCE. *Medical Education*, 24, 540-545.

Elder, H, 1999. Videotaped behavioural observations: Enhancing validity and reliability. *Applied Nursing Research* 12(4), 206-209.

Fink, L.D. 2003. *Creating significant learning experiences: An integrated approach to designing college courses*. San Francisco: Jossey-Bass.

Harden, R.M. and Gleeson, F.A. 1979. Assessment of clinical competence using an objective structured clinical examination (OSCE), *Medical Education*, 13: 41-54.

Hodges, B, 2003. Validity and the OSCE, *Medical Teacher*, 25(3), 250-254.

Jones, A., Pegram, A., Fordham-Clarke, C. 2010. Developing and examining an Objective Structured Clinical Examination. *Nurse Education Today*. 39(2) 137 -141

Mitchell, M., Henderson, A., Groves, M., Dalton, M., Nulty, D. 2009. The objective structured clinical examination (OSCE): optimising its value in the undergraduate nursing curriculum. *Nurse Education Today* 29(4):398-404

Major, D, 2005. OSCEs – seven years on the bandwagon: The progress of an objective structured clinical evaluation programme *Nurse Education Today* 442-454
Marjan J.B, Govaerts, Gees P.M., Van Der Vleuten and Schuwirth., 2002. *Advances in Health Sciences Education*, 7: 133-145.

Mazor, K.M., Zanetti, M.L., Alper, E.J., Hatem, D., Barrett, S.V., Meterko, V., Gammon, W. and Pugnaire, M.O., 2007. Assessing professionalism in the context of an objective structured clinical examination: an in-depth study of the rating process. *Medical Education*, 41(4), 331-340.

Newble, D, 2004. Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education* 38: 199-203

Noonan, M., Bradshaw, C. and Barry, M, 2008. Assessing student midwives' clinical skills using OSCEs in an academic setting. *The Practicing Midwife*, 12(5), 55-57.

Nulty, D.D., Mitchell, M.L., Jeffrey, C.A., Henderson, A., Groves, M., 2011. Best Practice Guidelines for use of OSCEs: maximizing value for student learning. *Nurse Education Today*, 31(2), 145-151.

Pender, F.T., de Looy, A.E. 2004. The testing of clinical skills in dietetic students prior to entering clinical placement. *The British Dietetic Association Ltd, Journal of Human Nutrition Dietetics*, 17, 17-24.

Race, P, 2005. *Making Learning Happen*. Sage, London.

Race, P. and Pickford, R, 2007. *Making Teaching Work. Teaching Smarter in Post-Compulsory Education*. Sage: Los Angeles

Richter, D., Sinclair, N, 2005. *The Bermuda Triangle in Education: navigating objectives content and assessment so that students and learning objectives don't disappear*. Teacher & Educational Development. University of New Mexico. [Http://hsc.unm.edu/som](http://hsc.unm.edu/som)

Rushforth, H.E, 2007. 'Objective structured clinical examination (OSCE): Review of literature and implications for nursing education'. *Nurse Education Today*, 27, 481-490.

Selby et al., 1995. Set up and run an objective structured clinical exam. *British Medical Journal*, 310, 1187-1190.

Tanner, C. A. 2006 thinking like a Nurse: a Research-Based Model of Clinical Judgement in Nursing. *Journal of Nursing Education*. 45, 6, 204-211

Wallace, J., Ranga, R., Haslam, R. 2002. Simulated patients and objective structured clinical examinations: review of their use in medical education. *Advances in Psychiatric treatment*, 8, 342-348