

Design and Validation of 2 Objective Structured Clinical Examination Stations to Assess Core Undergraduate Examination Skills of the Hand and Knee

NICHOLAS RAJ, LOUISA J. BADCOCK, GEORGE A. BROWN, CHRISTOPHER M. DEIGHTON, and SHEILA C. O'REILLY

ABSTRACT. *Objective.* To evaluate the development, validity, and reliability of 2 undergraduate Objective Structured Clinical Examination (OSCE) stations for core hand and knee examination skills.

Methods. Two OSCE stations for hand and knee based on core skills were developed, and qualitatively assessed for face and content validity by an expert consensus panel. Construct validity was evaluated by comparing the performance of third- (n = 21) and fifth-year (n = 50) medical students with 6 specialist registrars (SpR) in rheumatology. Concurrent validity was evaluated by correlating the scores of the fifth-year students with their eventual final examination scores. The fifth-year data were used to calculate the interrater and intrarater reliabilities of 2 examiners. Intrarater reliability analyzed repeat scores using videotapes of the examinations.

Results. Both stations were deemed to fulfil face and content validity criteria by the expert consensus panel. There was no significant difference in the mean scores of the third- and fifth-years. There were significant differences in the mean scores between both student groups and the SpR in both stations consistent with a valid construct theory. The fifth-year hand OSCE results correlated moderately with other indices of clinical skills, but not knowledge, and satisfied concurrent validity. Inter- and intrarater reliability for both stations was high.

Conclusion. These OSCE stations are valid and reliable tools for testing competency in core hand and knee examination skills. They can be used in educational research as outcome measures of specific teaching interventions and can also be used as an early feedback tool when teaching joint examination. (J Rheumatol 2007;34:421–4)

Key Indexing Terms:

OBJECTIVE STRUCTURED CLINICAL EXAMINATION
MUSCULOSKELETAL EXAMINATION
RELIABILITY
VALIDITY
EDUCATION

Musculoskeletal (MSK) disorders account for a significant proportion of acute and chronic illness in the western world¹. Although 20% of primary care consultations involve MSK disease, the teaching of this topic to undergraduates lacks the

emphasis that it deserves². The ever-increasing expansion of medical student numbers will put strain on existing teaching resources. Therefore current resources need to be used as effectively as possible, and new avenues to deliver both teaching and assessment should be developed and validated to augment these. Core MSK examination skills are an essential part of undergraduate medical training. As part of a project to train patients to teach joint examination to students³ we developed 2 Objective Structured Clinical Examination (OSCE) stations. These stations were predominantly designed as outcome measures to assess the effectiveness of the core skills teaching of the hand and knee; however, it was envisaged that they could also be used as formative assessment and feedback tools in third- and fifth-year clinical attachments. Our objectives were to evaluate the development, validity, and reliability of these 2 undergraduate OSCE stations for core hand and knee examination skills.

MATERIALS AND METHODS

Subjects. With ethical approval from the Southern Derbyshire Local Research

From the Department of Rheumatology, Derbyshire Royal Infirmary, Derby; and Postgraduate Dental and Medical Education School of Community and Health Sciences, University of Nottingham, Nottingham, United Kingdom.

Supported by the Department of Rheumatology, Derbyshire Royal Infirmary.

N. Raj, BSc, MBBS, MRCP, M Med Sci, Consultant Rheumatologist; L.J. Badcock, BSc, MBBS, MRCP, MSc, Consultant Rheumatologist; C.M. Deighton, B Med Sci, MBBS, FRCP, DM, Consultant Rheumatologist; S.C. O'Reilly, B Med Sci, MBBS, MRCP, DM, Consultant Rheumatologist, Department of Rheumatology, Derbyshire Royal Infirmary; G.A. Brown, BSc, DPhil, Professor of Education, Postgraduate Dental and Medical Education School of Community and Health Sciences, University of Nottingham.

Address reprint requests to Dr. N. Raj, Department of Rheumatology, Derbyshire Royal Infirmary, London Road, Derby DE1 2QY, United Kingdom. E-mail: Nicholas.raj@nhs.net

Accepted for publication September 29, 2006.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2007. All rights reserved.

Ethics Committee, 21 third- and 50 fifth-year medical students from Nottingham Medical School and 6 rheumatology specialist registrars (SpR) were recruited. All gave consent to take part in the OSCE and the students gave consent to have their performance videotaped for reliability scoring and for their final examination results to be released. Participants received verbal and written feedback on their performance if they wished.

Face and content validity. Published MSK examination core skills⁴ were used as a basis to establish a pilot OSCE for the hand and knee, respectively. This was further refined using a Delphi process in a series of meetings with an expert consensus panel of rheumatology specialists. This group consisted of 6 consultant rheumatologists, 2 SpR in rheumatology, and a clinical educator in rheumatology. The panel focused on qualitative issues of face validity (e.g., is what is being examined in the OSCE worth teaching?) and content validity (e.g., do the stations seem to address that which they are purported to?).

Construct validity. Construct validity (do the OSCE stations accurately measure core examination skills?) was evaluated by piloting each station with 2 groups that would be expected to perform to different levels. The underlying construct theory is that if the OSCE is an accurate measure of examination skill then there should be significant differences in the performance between novices and experts at joint examination. Two groups of students, the third- and fifth-year medical students, who had no previous experience in joint examination, were given 2 hours of standardized small group teaching in hand and knee examination. These students were then tested using the 2 OSCE stations 4 days later. The 4-day delay was to avoid an instant recall bias in the student performance. The performance of these 2 groups was compared with that of 6 SpR in rheumatology. Throughout the validation process the teaching was performed by CD, SR, and NR, and the testing by LB, SR, and NR.

Concurrent validity. Concurrent or criterion validity (the extent to which the OSCE stations correlate with other measures of clinical skills) was evaluated by correlating the OSCE scores of the 50 fifth-year students with their eventual finals scores. The analysis included all of the composite parts of the finals examinations: a 14 station OSCE, a combined skills score made up of a structured long case examination (OSLER) and OSCE score, a combined knowledge score made up of 2 written papers. The final OSCE also included specific knee and hand OSCE stations, and these were correlated with our OSCE stations for the hand and knee, respectively.

Inter- and intrarater reliability. The reliabilities of the OSCE stations when used by different examiners were measured using the data from the fifth-years. LJB and NR were both present at 31 of the 50 hand and knee OSCE, and the interrater reliability was calculated from these scores. The intrarater reliability was evaluated using repeated scores performed by each of the 2 examiners using videotapes of the OSCE. Paired scores were performed after at least a 2 month period for both the hand (NR: n = 50; LB: n = 20) and knee OSCE (NR: n = 43; LB: n = 21).

Statistics. All data were tested for normality using P-P plots. OSCE score means from different groups were compared using an independent samples t test. All correlations were performed using a 2-tailed Pearson test. The examiner interrater and intrarater reliabilities were analyzed using the intraclass correlation coefficient (ICC). Statistical analysis was performed using SPSS v.13.

RESULTS

Face and content validity. A 28-point hand station and a 25-point knee examination (Table 1) station were developed using published core skills and refined via expert consensus. Both were deemed to fulfil face and content validity criteria by the expert consensus panel.

Construct validity. OSCE scores at both stations for the third-years, fifth-years, and SpR were normally distributed. Mean hand OSCE scores were 18.2 (range 13–24) for the third-years, 17.6 (11–23) for the fifth-years, and 24 (20–28) for the

SpR. Mean knee scores were 16.9 (13–20) for the third-years, 15.5 (10–20.5) for the fifth-years, and 24.2 (23–25) for the SpR. There was no significant difference in the mean scores of the third- and fifth-years. There were significant differences in the mean scores between both student groups and the SpR in the hand station [third-year: $p < 0.001$ (95% CI 3.0–9.1); fifth-year: $p < 0.001$ (3.9–8.9)] and the knee station [third-year: $p < 0.001$ (5.4–9.0); fifth-year: $p < 0.001$ (4.5–10.8)].

Concurrent validity. The fifth-year hand station results correlated moderately with the overall OSCE score in the finals ($r = 0.44$, $p < 0.01$) and the overall skills score ($r = 0.51$, $p < 0.01$). Low correlation was found with the overall finals score ($r = 0.31$, $p < 0.05$), but no correlation was found with the paper section of the finals or the final hand OSCE. The fifth-year knee OSCE correlated moderately with the overall OSCE score in the finals ($r = 0.46$, $p < 0.01$), the overall skills score ($r = 0.46$, $p < 0.01$), and the final knee OSCE ($r = 0.40$, $p < 0.01$), but no correlation was found with the paper or overall finals score. This moderate correlation with other indices of clinical skills, but not of knowledge, is in keeping with the levels of internal correlates of the final examinations⁵ and satisfies concurrent validity.

Inter- and intrarater reliability. Interrater reliability for both OSCE was high [hand: ICC = 0.86, $p < 0.001$ (95% CI 0.63–0.94); knee: ICC = 0.90, $p < 0.001$ (0.79–0.95)]. Intrarater reliability was high for both examiners for the hand station [examiner 1: ICC = 0.94, $p < 0.001$ (0.89–0.97); examiner 2: ICC = 0.73, $p < 0.01$, (0.31–0.89)], and for the knee [examiner 1: ICC = 0.91, $p < 0.001$ (0.83–0.96); examiner 2: ICC = 0.69, $p < 0.03$ (0.19–0.91)].

DISCUSSION

Main findings. Both the hand and the knee OSCE stations were deemed by the expert panel to be addressing a subject that is worth teaching to undergraduates, in a manner that appears to adequately cover this, thus fulfilling both face and content validity. The significant difference in the performance between novices (third- and fifth-year medical students) and experts (rheumatology SpR) in MSK examination as well as the lack of difference between the 2 similar novice groups is consistent with a valid construct theory for both stations. The lack of difference between the third- and fifth-year students enhances construct validity, as it shows that they performed the same, despite being different ages and at different stages of medical training, as both groups received identical teaching and had not had any formal joint examination teaching before this project. Groups that were expected to achieve high scores did so, while groups that were expected to be competent but not exceptional achieved moderate scores. Construct validity is imperative in any assessment that is expected to differentiate performance, especially if used in a high-stakes examination.

The correlations of our individual hand and knee stations with the final fifth-year scores showed moderate correlation

Table 1. Summary of knee and hand OSCE score system.

	Knee Examination	Hand Examination
Observation	Standing, walking, lying (0 = not done/2 omissions, 1 = one omission, 2 = all performed, total = 2)	Palms and dorsum (1 mark, 0 if incorrect, total = 1)
Description	Skin changes; muscle wasting; swollen joint; bursa; popliteal cyst/swelling; deformity (1 mark for each, 0 if incorrect, maximum total = 4)	Skin changes; muscle wasting; nodules or nodes; nail changes; joint swelling; tendon swelling or deformity; symmetry and distribution** (1 mark for each, 0 if incorrect, maximum total = 6)
Functional tests	—	Power grip; pincer grip; assess strength of both power and pincer grips (1 mark for each, 0 if incorrect, total = 3)
Palpation	Joint temperature; tibiofemoral joint line tenderness; popliteal fossa; collateral ligaments (1 mark for each, 0 if incorrect, total = 4) Knee effusion (bulge and balloon)* (2 mark total = 2)	Warmth; nodules (elbow or hand) Wrist joint; MCPJ; PIPJ; DIPJ; 1st CMC*** (1 mark for each, 0 if incorrect, total = 7)
Movement	Assess range of movement; quadriceps lag (1 mark for each, 0 if incorrect, total = 2) Anterior and posterior cruciate ligament stress test; collateral ligaments stress test* (2 marks maximum for each, total = 4)	Range of movement at wrist (1 mark, 0 if incorrect, total = 1)
Nerves	—	Phalens test; sensation; power (e.g., thumb abduction/finger spread) (1 mark for each, 0 if incorrect, total = 3)
Summary	Gait assessment; functional assessment; diagnosis (1 mark for each, 0 if incorrect, total = 3)	Function; diagnosis; assessment of whether inflammatory and if so a comment on activity (1 mark for each, 0 if incorrect, total = 3)
Overall	Rapport; approach (logical, professional, systematic) (maximum 2 marks for each: 0 = poor, 1 = average, 2 = good, total = 4)	Rapport; approach (logical, professional, systematic) (maximum 2 marks for each: 0 = poor, 1 = average, 2 = good, total = 4)
Total score	/25 * (-1 per mistake)	/28 ** (compulsory) *** (swelling and tenderness)

MCPJ: metacarpophalangeal joint; PIPJ: proximal interphalangeal joint; DIPJ: distal interphalangeal joint; CMC: carpometacarpal joint.

Table 2. Correlations of hand and knee stations with overall finals and composite finals scores. All p Pearson 2-tailed.

	Final Hand OSCE	Final Knee OSCE	Final OSCE	Final OSLER	Overall Skills Score (OSCE + OSLER)	Overall Knowledge Score	Overall Finals Score
Hand station (n = 50)	r = 0.16 p = 0.27	r = 0.45 p = 0.001	r = 0.44 p = 0.001	r = 0.38 p = 0.007	r = 0.51 p = 0.001	r = 0.01 p = 0.93	r = 0.31 p = 0.03
Knee station (n = 50)	r = 0.09 p = 0.56	r = 0.39 p = 0.005	r = 0.46 p = 0.001	r = 0.27 p = 0.06	r = 0.46 p = 0.001	r = -0.04 p = 0.79	r = 0.24 p = 0.1

with both the 14-station OSCE score of the finals examination and the overall score of clinical skills (14-station OSCE plus an OSLER). These are perhaps the best correlates of examination skill available to measure in this group. The correlation is not higher because the overall scores of clinical skills in the finals examination contain a greater input from other skills, such as overall interpretive and management skills, which are also tested (e.g., “how would you treat RA?”). As we had focused more on the core skills with not as much emphasis on interpretation it is not unreasonable to expect that the correlation would not be as high. This probably accounts for the lack

of correlation of our hand station with the finals hand OSCE. The lack of correlation between our stations and the paper or “knowledge” section of the finals further enhances the concurrent validity. The hand station showed a low correlation with the overall finals score and the knee OSCE showed none. This was expected, as the overall score is a composite of knowledge, clinical skills, interpretive skills, and reasoning skills. The high inter- and intrarater reliability indicates that both stations were a reliable measure over time and between different examiners.

Limitations of this work. We used no comparable objective

measure of the third- and fifth-years' baseline competencies prior to taking part, such as structured questionnaires. As a result we assumed similar competency based on their reports of having had no formal joint examination training. However, the lack of difference in their scores does bear out the relative homogeneity of these 2 groups. The intrarater reliability was performed using a single "live" score and a subsequent video score. The use of 2 different methods to observe the same episode can be criticized, as the video had the potential to either miss the fine detail of the examination if in wide-angle view, or pertinent nonverbal communication if in close-up. We chose to use the videos in close-up so as to maximize the ability to accurately score the actual examination rather than make judgements on the rapport, which was less weighted in the scoring system. One method of concurrent validation would have been to use the finals marking system in parallel with our new marking system. This was not done due to resource implications, but may have shown a stronger correlation, as the finals marking scheme could then have been directly compared to the students' actual finals performance and the interpretive and reasoning weighting could have been quantified.

Implications of this work. Using established core MSK examination skills it is possible to develop structured tools such as OSCE stations for the assessment of those examination skills. The objectivity of an OSCE relies on the standardization of the task and scoring system used⁶. It is essential to address each individual facet of validity and reliability when designing any instrument that may be used for assessment purposes. These facets are often not amenable to the same method of measurement or evaluation. In this example, face and content validity were ascertained using a qualitative approach via an expert panel. Construct validity, concurrent validity, and rater reliabilities were all measured using quantitative statistics. These 2 OSCE are valid and reliable tools for testing competency in performing core hand and knee examination skills.

Although OSCE are considered to be resource-intensive^{7,8}, they are accepted as a reliable and valid measure of objective assessment⁹⁻¹¹. They can be used in educational research as outcome measures of specific teaching interventions and can also be used as an early feedback tool when teaching joint examination. Our model for teaching undergraduate MSK medicine is to teach basic core examination skills and then to build upon this with individual disease knowledge and signs. The lack of weighting towards interpretation of signs in these OSCE stations allows for the initial stages of acquisition of core skills to be assessed. For a comprehensive assessment of overall clinical competence, such as those used in final examinations, many more methods of testing are required to complement the OSCE.

It is a well established concept that learners benefit from well structured, prompt, and relevant feedback on their performance, especially when giving insight into their relative performance compared to what is expected at their level^{12,13}. Anecdotally, our students reported benefiting from feedback on their clinical skills. These are the first validated OSCE stations for core hand and knee examination skills to be published in detail, and we propose that they could be used to provide such feedback early in undergraduate musculoskeletal training.

ACKNOWLEDGMENT

We gratefully acknowledge our colleagues in the Department of Rheumatology at the Derbyshire Royal Infirmary, and the local rheumatology registrars and Nottingham University medical students who took part.

REFERENCES

1. Badley EM, Crotty M. An international comparison of the estimated effect of aging of the population on the major cause of disablement, musculoskeletal disorders. *J Rheumatol* 1995;22:1934-40.
2. Dequeker J, Rasker JJ, Woolf AD. Educational issues in rheumatology. *Baillieres Best Pract Res Clin Rheumatol* 2000;14:715-29.
3. Raj N, Badcock LJ, Brown GA, Deighton CM, O'Reilly SC. Undergraduate musculoskeletal examination teaching by trained patient educators: A comparison with doctor led teaching. *Rheumatology Advance Access*, published April 13, 2006. doi:10.1093/rheumatology/ke1126.
4. Coady D, Walker D, Kay L. Regional Examination of the Musculoskeletal System (REMS): a core set of clinical skills for medical students. *Rheumatology Oxford* 2004;43:633-9.
5. Neame R 2004. Report to the Nottingham Medical school ACE committee. Appendix A, Sept ACE committee meeting. Nottingham University, Nottingham, UK.
6. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured clinical examinations. *BMJ* 1975;1:447-51.
7. Cusimano MD, Cohen R, Tucker W, Muurnagen J, Kordama R, Reznick R. A comparative analysis of the costs of administration of an OSCE. *Acad Med* 1994;69:571-6.
8. Cohen R, Reznick RK, Taylor BR, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. *Am J Surg* 1990;160:302-5.
9. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination. *Med Educ* 1979;13:41-54.
10. Barman A. Critiques on the Objective Structured Clinical Examination. *Ann Acad Med Singapore* 2005;34:478-82.
11. Newble DI. Assessing clinical competence at the undergraduate level. *Med Educ* 1992;26:504-11.
12. Kulik JA, Kulik CC. College teaching. In: Peterson PL, Walberg HJ, editors. *Research on teaching: Concepts, findings and implications*. Berkeley, CA: McCutchan Publishing; 1979.
13. Rushton A. Formative assessment: a key to deep learning? *Med Teacher* 2005;27:509-13.