How to assess communication, professionalism, collaboration and the other intrinsic CanMEDS roles in orthopedic residents: use of an objective structured clinical examination (OSCE)

Tim Dwyer, MBBS†
Susan Glover Takahashi, MA, PhD‡
Melissa Kennedy Hynes, MA³
Jodi Herold, PhD⁵
David Wasserstein, MD³
Markku Nousiainen, MD, MSc, MEd⁶
Peter Ferguson, MD³
Veronica Wadey, MD, MA⁷
M. Lucas Murnaghan, MD, MEd⁸
Tim Leroux, MD⁹
John Semple, MD, MSc⁹
Brian Hodges, MD, Med, PhD¹⁰
Darrell Ogilvie-Harris, MD, MSc*¹†

From the *Women’s College Hospital, †Mt. Sinai Hospital, ‡Postgraduate Medicine Office, University of Toronto, §University of Toronto, ¶Holland Orthopaedic and Arthritic Centre, **Sunnybrook Health Sciences Centre, ††University of Toronto Orthopaedics Sports Medicine, ‡‡The Hospital for Sick Children, §§Toronto Orthopaedics Sports Medicine, †††University of Toronto, ¶¶Holland Orthopaedic Office, University of Toronto, §§§Toronto Western Hospital, Toronto, Ont.

Presented at the International Conference Surgical Education and Training (ICOSET) 2013, Canadian Orthopaedic Association 2013, Canadian Conference Medical Education 2013.

Acceptance for publication Oct. 29, 2013

Correspondence to:
T. Dwyer
Women’s College Hospital
76 Grenville St.
Toronto ON M5S 1B1
dwyer.orthopedics@gmail.com

DOI: 10.1503/cjs.018813

Background: Assessing residents’ understanding and application of the 6 intrinsic CanMEDS roles (communicator, professional, manager, collaborator, health advocate, scholar) is challenging for postgraduate medical educators. We hypothesized that an objective structured clinical examination (OSCE) designed to assess multiple intrinsic CanMEDS roles would be sufficiently reliable and valid.

Methods: The OSCE comprised 6 10-minute stations, each testing 2 intrinsic roles using case-based scenarios (with or without the use of standardized patients). Residents were evaluated using 5-point scales and an overall performance rating at each station. Concurrent validity was sought by correlation with in-training evaluation reports (ITERs) from the last 12 months and an ordinal ranking created by program directors (PDs).

Results: Twenty-five residents from postgraduate years (PGY) 0, 3 and 5 participated. The interstation reliability for total test scores (percent) was 0.87, while reliability for each of the communicator, collaborator, manager and professional roles was greater than 0.8. Total test scores, individual station scores and individual CanMEDS role scores all showed a significant effect by PGY level. Analysis of the PD rankings of intrinsic roles demonstrated a high correlation with the OSCE role scores. A correlation was seen between ITER and OSCE for the communicator role, while the ITER medical expert and total scores highly correlated with the communicator, manager and professional OSCE scores.

Conclusion: An OSCE designed to assess the intrinsic CanMEDS roles was sufficiently valid and reliable for regular use in an orthopedic residency program.
The 7 CanMEDS competencies (medical expert, communicator, collaborator, manager, health advocate, scholar and professional) have been clearly outlined in the CanMEDS 2005 Physician Competency Framework, by the Royal College of Physicians and Surgeons of Canada. A similar framework has been described by the Accreditation Council for Graduate Medical Education (ACGME), defining 6 core competencies. Each of these frameworks describe the principal generic abilities of physicians in health care, and are an integral component of postgraduate education.

Assessment options for the intrinsic roles (those other than the medical expert role) include in-training evaluation reports (ITERs), structured oral examinations, 360° assessments and objective structured clinical examinations (OSCE). A survey of a wide variety of medical and surgical program directors in Canada identified that the ITER is the most commonly used method to evaluate the CanMEDS roles, despite its acknowledged subjective nature. Respondents reported dissatisfaction with current methods of evaluating the intrinsic roles, especially the manager and health advocate roles.

The “OSCE” is a term used to describe a variety of multi-station examinations and is a format currently favoured at orthopedic certification examinations in Canada and other countries. Studies using OSCEs to assess the role of medical expert have demonstrated reliability and validity in postgraduate physician training, with some literature demonstrating the ability of the OSCE to assess communication skills. In fact, there is evidence that improved interpersonal skills are linked to better overall OSCE performance.

There is also some evidence that an OSCE can be used to assess other CanMEDS roles, including the scholar role (application of evidence-based medicine or demonstration of teaching skills) and the professional role (cultural awareness and the application of ethical principles). An OSCE has also been used to assess multiple CanMEDS competencies in other fields of postgraduate training, such as radiology and neonatology.

To our knowledge, no research exists regarding methods of assessing the intrinsic roles in orthopedic postgraduate training. We hypothesized that an OSCE designed to assess multiple intrinsic CanMEDS roles would have sufficient reliability and validity to distinguish between different years of postgraduate training in orthopedic residents.

Methods

Exam development

The orthopaedic residency program at the University of Toronto, in collaboration with the Postgraduate Medical Education Department, designed an orthopedic OSCE to test the 6 intrinsic CanMEDS roles. A focus group of academic orthopedic specialists was assembled with the goal of creating clinical scenarios evaluating selected CanMEDS principles for each of the roles. The focus group relied on the source document from The Royal College of Physician and Surgeons of Canada, in which each of the roles, and their key competencies, is clearly defined.

The OSCE was 1 hour long and comprised 6 10-minute stations. The roles of communicator, collaborator, professional, manager, health advocate and scholar were assessed; a deliberate attempt was made to avoid testing the medical expert role. Most of the 6 case-based scenarios were designed to assess a primary and secondary role. Stations 2–4 used standardized patients (SPs; station 2: relative concern regarding delay in surgery; station 3: grandmother of child with suspected nonaccidental injury; station 4: teenager being informed of osteosarcoma diagnosis), whereas station 5 used a standardized health professional (SHP; operating room manager). Two stations did not have an SP or SHP (station 1: ethical approach to needlestick injuries; station 6: evidence-based medicine in spine surgery). Table 1 lists the roles and associated key competencies tested in each station.

The CanMEDS OSCE development was facilitated by an exam blueprint and case development guides. A member of the focus group was assigned as the lead to design each station, which was then reviewed by the entire focus group. Any discrepancies or ambiguities were addressed or removed. A number of 5-point performance rating scales were developed for each of the intrinsic roles. The ratings were anchored by descriptions of performance to be demonstrated by the residents for each role. An overall 5-point global rating was also assigned for each resident at the end of the station. The SPs and SHPs were selected from an established standardized patient bank at the University of Toronto. For the OSCE, 2 SPs/SHPs were trained for each of the stations by an experienced SP trainer; each received a minimum of 3 hours’ training for each role. No SP or SHP assessment of performance was used.

Study design

Convenience sampling was used to recruit residents from specific postgraduate years (PGY) of training: PGY0 (incoming residents who had not yet begun training), PGY3 residents and PGY5 residents (volunteers who asked to sit the OSCE). The PGY5 residents had all recently passed their orthopedic certification examinations, and were included as the “gold standard.”

Members of the orthopedic faculty at each station evaluated residents independently. It was not possible to blind examiners from the year of training of the residents, as many of the residents were familiar to the staff surgeons. However, examiners were asked to disregard the year of training when making assessments. The OSCE was
conducted in 4 1-hour sessions over the course of a single day. Each candidate signed a consent form permitting the use of exam results for research purposes. On completion of the exam, residents were invited to provide feedback using a 5-point Likert scale. Summative and formative feedback was given to each resident at the end of the OSCE.

Concurrent validity was sought in 2 ways. First, we obtained the ITERs from the preceding 12 months for the PGY3 and PGY5 residents, and the results on the 6 intrinsic roles correlated with the OSCE total score and role scores. Second, the 2 program directors (PDs) formed an ordinal ranking of the residents in PGY3 and PGY5 and rated each resident’s ability in each of the CanMEDS roles using a 5-point scale (1 = needs significant improvement; 2 = below expectations; 3 = solid, competent performance; 4 = exceeds expectations; 5 = superb). The overall ranking and the rating for each role were also correlated to the total OSCE score and role scores.

Approval for this study was obtained from the Research Ethics Board, University of Toronto. Each resident signed a consent form to permit the use of the OSCE results and ITER results for research purposes.

Statistical analysis

All data were deidentified, and residents were assigned a study-specific number. Raw scores from the individual station scores and role scores were entered into a spreadsheet and analyzed using SPSS version 19. All scores were converted into a percentage, with results reported as mean ± standard deviation. Reliability was established using the interstation α coefficient of reliability (Cronbach α) for each of the scoring tools. We evaluated scores from the different rating scales using regression analysis. The effect of PGY on total test scores (%), overall ratings of performance, individual station scores and role scores were evaluated using 1-way analysis of variance (ANOVA). We considered results to be significant at \( p < 0.05 \). We used the Scheffe test for post hoc analysis to understand differences in scores between each possible pair of years of training. We determined the correlation between total OSCE scores and role scores with ITER role scores and PD rankings using Pearson correlation and Spearman Rho \( r \), and the Student \( t \) test was used to compare the PD ratings of resident performance between PGY levels.

Results

Twenty-five residents from PGY0, 3 and 5 took part in the OSCE (Table 2). The roles of communicator, manager and professional were assessed in multiple stations; collaborator, health advocate and scholar were assessed in 1 station each. The total test scores (converted to a percentage) and the mean overall rating of performance are shown in Table 3; ANOVA testing demonstrated a significant
difference of the effect of PGY on both scores \( (p < 0.001) \). We found a significant difference between PGY0 and PGY3 \( (p = 0.039) \), PGY3 and PGY5 \( (p = 0.001) \), and PGY0 and PGY5 \( (p < 0.001) \).

The interstation reliability (percent) was 0.87 for total test scores and 0.83 for overall ratings of performance. The internal consistency for 4 of the 6 role scores is shown in Table 4; the consistency for each of these 4 roles was very high \( (> 0.80) \). We were not able to compute internal consistency coefficients for the scholar and advocate roles, as only 1 rating scale was used for each of these roles.

The total test scores for the individual stations by PGY are displayed in Figure 1. The effect of PGY on the individual station scores was significant (stations 1, 5 and 6, all \( p < 0.01 \); stations 2 and 4, both \( p < 0.05 \)) with the exception of station 3 \( (p = 0.07) \). Post hoc analysis demonstrated a significant difference in station scores between the PGY5 and PGY0 groups and between the PGY5 and PGY3 groups for all stations except station 3. No significant difference in scores was seen between the PGY0 and PGY3 groups, but there was a trend for increased scores in the PGY3 group.

The total test scores for each of the intrinsic roles by PGY are shown in Figure 2. The ANOVA testing for the effect of PGY on each of the role scores was significant (communicator, collaborator, manager and professional roles, all \( p < 0.001 \); health advocate and scholar roles, both \( p < 0.05 \)). For each of the role scores, the PGY0 and PGY3 groups differed significantly from the PGY5 group, but not from each other.

Analysis of the PD ratings of intrinsic roles demonstrated a good correlation between these and the corresponding OSCE role scores (Table 5). The ITERs from 12 months before the OSCE were available for the PGY3 and PGY5 residents. There was no correlation between ITERs and OSCE scores within any role except for communicator (0.64); however, the ITER overall scores correlated with the communicator (0.58), manager (0.51) and professional (0.56) OSCE role scores.

There was a 64% (16 of 25) response rate to the resident survey. Overall, 87.6% of respondents agreed or strongly agreed that the scenarios reflected encounters that an orthopedic surgeon would have to deal with in general practice, and 81.3% agreed or strongly agreed that participating in the OSCE would help prepare them for their final Royal College examination. However, only 56.3% agreed or strongly agreed that the OSCE was an effective way to assess their understanding of each of the CanMEDS roles.

<table>
<thead>
<tr>
<th>Role</th>
<th>( \alpha ) Coefficient</th>
<th>Item numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicator</td>
<td>0.91</td>
<td>17 items in 5 of 6 stations</td>
</tr>
<tr>
<td>Collaborator</td>
<td>0.96</td>
<td>3 items in 1 of 6 stations</td>
</tr>
<tr>
<td>Manager</td>
<td>0.83</td>
<td>5 items in 3 of 6 stations</td>
</tr>
<tr>
<td>Professional</td>
<td>0.84</td>
<td>3 items in 2 of 6 stations</td>
</tr>
</tbody>
</table>
**Discussion**

This orthopedic OSCE, designed to test the 6 intrinsic CanMEDS roles, has shown excellent overall reliability, as well as excellent reliability for the communicator, collaborator, professional and manager roles. Furthermore, using role-specific global ratings, we demonstrated an ability to distinguish between orthopedic residents at different levels of training. To our knowledge, this is the first time formal assessment of intrinsic roles has been studied in the field of postgraduate orthopedic surgical training.

With regards to total test scores, each of the stations were able to demonstrate a statistically significant difference by year of training, with the exception of station 3. Station 3 was a case-based scenario using a standardized patient — a grandmother who has brought in a child thought to have sustained a nonaccidental injury. Residents were asked to take a focused history regarding the home situation and explain to the grandmother the need to alert the appropriate authorities and admit the child. Despite the fact that the PGY effect on station 3 scores was not statistically significant ($p = 0.07$), the scores demonstrate the same general trend as all other stations: PGY5 (mean 90.5%) > PGY3 (mean 74.3%) > PGY1 (mean 69.5%). It may be that this station did not achieve significance because the PGY0 and PGY3 residents performed well, suggesting that these competencies may have been covered in undergraduate medical programs.

Careful blueprinting was used in this CanMEDS OSCE to avoid redundancy. Roles were spread among stations, and the stations that assessed the same roles focused on different competencies within that role, as outlined in the CanMEDS 2005 Framework.1 For example, in this OSCE, 2 stations (needlestick and trauma list) both examined the roles of communicator and manager, with the needlestick station additionally examining resident understanding of the professional role (bioethical principles and informed consent). However, the trauma list station focused on the competencies of priority setting and time management within the manager role, while the needlestick station sought to examine the competency of managing practice and career effectively.

The OSCE has been previously been shown to be a valid and reliable tool for the assessment of the medical expert role, with some evidence for its use in assessing the communicator role. Improved communication skills have previously been linked to both advanced year of training and to increased clinical competence. The OSCE has also been adapted to assess competencies within the roles of professional and scholar, with varying amounts of success. For example, Singer and colleagues24 in an OSCE designed to assess clinical ethics, found a low reliability with only 4 stations; it was felt that increasing the number of stations would be required to obtain acceptable reliability.

Jefferies and colleagues recently demonstrated that the OSCE may be a valid and reliable method of simultaneously assessing multiple competencies in neonatal/perinatal medicine. Subspecialty trainees were assessed using a combination of binary checklists, 5-point CanMEDS ratings, as well as SPs’ and SHPs’ assessments of interpersonal and communication skills. Interstation reliability was acceptable for all 6 of the 7 roles, with the scholar role being the exception. Only the teaching component of the scholar role was assessed; the authors recommended creating a single station to assess the competencies inherent to the scholar role, including the ability to understand and evaluate research. We applied this technique with success; our scholar station, designed to assess application of evidence-based medicine, was able to distinguish between residents with different levels of training.

Jefferies and colleagues also studied the use of the structured oral examination in assessing the 7 CanMEDs roles, including the medical expert role. Interstation reliability was acceptable for the roles of medical expert, scholar and professional (0.6–0.8), but not for the roles of communicator, collaborator or health advocate (0.4–0.6) or for the role of manager (0.19). In comparison to their previous OSCE study, interstation reliability was lower for all roles except scholar. However, costs were reduced significantly by not using standardized patients. Although we felt that SPs were an important component of our OSCE, the costs (in the region of $3000) were not insignificant, equating to a cost of $250 per resident. It may be possible to substitute orthopedic fellows or staff in place of SPs in future iterations. However, given the importance of establishing competence in these areas by both the Royal College and the ACGME, this could be seen as a reasonable cost for training programs to bear on an annual basis.

The Royal College has a published handbook detailing assessment methods for the CanMEDS competencies. It states that oral examinations and OSCEs are not well suited to evaluate the roles of manager and scholar. Other documents attest to the perceived difficulty with assessing the intrinsic roles, especially the role of health advocate. However, the reliability of the manager role in our study was high enough to be used in a high-stakes examination. While we cannot attest to the reliability of the health advocate and scholar roles owing to insufficient items, ANOVA testing demonstrated a significant ability to distinguish between residents of different PGY levels for both of these roles. We believe that the OSCE is a very appropriate means of assessment, as clinical scenarios that mimic real life encounters can be used.

An advantage of this type of OSCE is that both teaching (formative evaluation) and assessment (summative evaluation) can be incorporated. As noted by Zuckerman and colleagues, assessment motivates residents to learn important skills and is therefore a form of learning in itself. We believe that by exposing very junior residents (PGY0) to
Importantly, the examiners will have known some residents and their PGY of training, raising the potential for bias. Examiners were asked to disregard the PGY level of the resident; however, it may be that the use of SPs’ ratings will help to offset this risk. In this OSCE, neither station nor role weighting was used, as it was felt that each of the CanMEDS roles was equally important. Finally, it is not possible to know how this CanMEDS OSCE compares to a more traditional OSCE with incorporated assessment of CanMEDS roles within those stations; however, we have demonstrated a high degree of reliability or internal consistency, a measure that indicates an exam is performing well. It may be that the high degree of reliability seen in this CanMEDS OSCE may be a result of its narrow focus.

CONCLUSION

An OSCE designed to assess the intrinsic CanMEDS roles proved to be sufficiently valid and reliable for regular use in an orthopedic residency program.

Competing interests: None declared.

Contributors: T. Dwyer, P. Ferguson, M.L. Murnaghan, B. Hodges and D. Ogilvie-Harris designed the study. T. Dwyer, D. Wasserstein, M. Nousiainen, P. Ferguson, V. Wadey, T. Leroux and D. Ogilvie-Harris acquired the data, which T. Dwyer, S. Glover-Takahashi, M. Hynes, J. Herold, J.L. Semple, B. Hodges and D. Ogilvie-Harris analyzed. T. Dwyer, S. Glover-Takahashi, M. Hynes, J. Herold, B. Hodges and D. Ogilvie-Harris wrote the article, which all authors reviewed and approved for publication.

REFERENCES


---

**Canadian Journal of Surgery**

**Online manuscript submission and peer review**

AVAILABLE at

[http://mc.manuscriptcentral.com/cjs](http://mc.manuscriptcentral.com/cjs)