The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE) – Development of a Competency Based Procedure Log

WT Gofton MD MEd FRCSC, Assistant Professor of Surgery University of Ottawa, Study Design, Study introduction to Staff and Trainees, Oversight of evaluation collection within Division, Data Review, Writing and Presentation of Results

NL Dudek MD MEd FRCPC, Associate Professor of Medicine University of Ottawa, Outside lead for focus groups, Data Review and Writing

Steve R Papp MD Med FRCSC, Assistant Professor of Surgery University of Ottawa, Study Design, Assistance in evaluation collection within Division, Data Review and Writing

TJ Wood PhD, Medical Council of Canada, Data Analysis and writing

SJHamstra PhD, Acting Assistant Dean, Academy of Innovation in Medical Education University of Ottawa, Data Analysis and writing

Budgetary Request – $25,540
Abstract

*Design*—Surgical training programs are responsible for ensuring the competency of their graduates. This includes their ability to perform the various surgical procedures safely and successfully. In most training programs, a resident’s ability to do these procedures is evaluated by general observations documented under the single item of “technical skills” on an In-training Evaluation Report (ITER). We have developed and novel competency based surgical evaluation tool (O-SCORE) that demonstrated a significant effect of post-graduate year (PGY) level with higher scores for increases in training level except for PGY 4 and 5 regardless of procedure type. Qualitative analysis indicated that the staff surgeons found it easy to use and trainees found it improved feedback.

In this study the O-SCORE will be converted to an electronic format and assessed for continued validity and feasibility. Thirty volunteer surgical trainees from the Division of Orthopedic Surgery at the University of Ottawa will collect electronic O-SCORE evaluations when any of six pre-selected procedures are performed during a 4-month period. The psychometric characteristics of the electronic scale will be examined to determine reliability and validity. Following the quantitative data collection period, two focus groups will be conducted to gather information on the utility of the scale, and to provide valuable input, which can be used as part of the ongoing effort to improve content validity. The feasibility of using the O-SCORE to develop a ‘competency based’ procedure log will also be evaluated.
Background

Clinical competency has been defined as “the degree to which the individual uses the knowledge, skills and judgment associated with the profession to perform effectively in the domain of possible encounters defining the scope of professional practice.”(1) While many of the aspects of clinical competence have been addressed by the core competencies as defined by CanMEDS and the Accreditation Council for Graduate Medical Education (ACGME), surgical and technical competence has “historically been assessed poorly and continues to receive little attention among the core competencies defined by CanMEDS and ACGME”.(3) However, with the rise of professional accountability, training programs and licensing bodies will be increasingly responsible for ensuring the clinical and surgical competence in their graduates.

At present national licensing boards mainly rely on written and oral examinations to assess competence. These forms of assessment primarily examine a trainee’s knowledge for the basic and clinical science relevant to their field of practice along with some diagnostic and knowledge based management skills. They do not assess a trainee’s ability to actually perform the required procedures. The assessment of surgical competence is limited to a candidate’s description of what they would do in a given operating room (OR) scenario, not their ability to perform the task. While in the future simulation may assist in the assessment of surgical competence, for now the training programs bear a significant responsibility in ensuring that its graduates can competently perform in the OR, as these skills will not be tested on their licensing exams.

Clearly the surgical programs’ ability to accurately evaluate the trainee’s ability to actually perform the surgical procedures relevant to their specialty field is of utmost importance. Ideally a surgical training program would utilize multiple forms of assessment to determine the competency of a trainee. At the current time, the majority of surgical training programs rely on faculty ratings for technical competence usually as part of the In-training Evaluation Report (ITER).

Technical Skills Assessment Tools

Several methods of evaluating surgical technical skills have emerged over the last decade. While not new, procedure or case logs are increasingly being used by programs and the ACGME to assess trainee exposure. However exposure does not necessarily translate to competence, and although a certain number of procedures performed are clearly necessary to provide opportunity to progress through stages of acquisition of technical skill, the actual number is highly variable between individuals.1 Procedure logs are often highly dependant on the trainee’s accurate completion, as well as deficient in timely feedback and evaluation. The result is a log of procedural performance lacking content validity rather than a reflection of trainee’s operative ability.(7,8) In essence, ‘just because you were there does not mean you are competent to do it yourself’. For these reasons procedure logs are most useful for assessing opportunities provided by a training program and as a guide for program directors rather than to assess individuals.2

Formal laboratory based technical skills exams,(9,10) videotaped review of operative cases(11) and the use of motion analysis devices(12,13) have also been introduced. While all these methods have been demonstrated to objectively assess technical skills, they can be expensive and time-consuming to implement limiting their value at the level of the program.
Although many of these tools are of value in the assessment of ‘technical skill’, the overall evaluation of ‘surgical competence’ remains elusive. This is because surgical competence encompasses more than just the ability to perform the technical steps or skill of the procedure. It requires appropriate judgment in case selection, pre-operative planning, understanding the steps and potential pitfalls of the procedure, the application of the required skills and the ability to communicate with and effectively utilize the team around you. Therefore an assessment tool that focuses on these aspects of surgical competence could provide valuable information to a training program in the determination of a trainee’s overall surgical competence.

Moving Towards Reliable and Valid Clinical Assessment

It is generally believed that experts know what competent clinical performance entails and can judge both the quality and appropriateness of the trainee’s practice.(15,16) Therefore, a strong case can be made that only someone with knowledge of a similar scope of practice is qualified to judge competence.(17) The mini-clinical examination (mini-CEX) developed by the medical American Board of Internal Medicine encourages direct observation with immediate evaluation during residency training(18). It has been found to have good construct validity and reliability(19), and has been listed by the ACGME as a “best practice” tool for several of the defined competencies. Its strength lies in its brevity, allowing for multiple prospective assessments by different assessors in a variety of clinical scenarios, along with its potential to evaluate multiple competencies.

Unlike non-surgical specialties, surgical residents and their preceptors spend a large amount of time demonstrating and observing technical skills in the operating room. A number of authors have tried to develop intra-operative assessment tools, however most are procedure specific. (21)(22) Unfortunately the development and validation of procedure specific assessment tools for all surgical specialties and required procedures is prohibitive. Many of the authors have noted a central scoring tendency, demonstrating an evaluator’s reluctance to assign low marks on a GRS.(14) It is not surprising that an evaluator would avoid assigning a low score and drift towards a ‘comparison of level’ when the GRS provides the opportunity for drift or there is the perception that this one evaluation represents a significant portion of a trainee’s overall evaluation.

Development of the O-SCORE

In an effort to capitalize on this ‘missed opportunity’ for surgical evaluation in the operating room we have developed a succinct competency based surgical assessment tool that provides both summative and formative feedback. Aware of the reluctance of evaluators to assign low scores, (14,23) we decided to rate the trainee with respect to readiness for independent performance of the particular procedure as opposed to comparing them to peer group (i.e. ability to do this procedure relative to other PGY2). This essentially established an external reference criterion on which to base ratings. We hoped this would encourage the expert evaluator to make full use of the rating scale. For example, an evaluating surgeon would not expect that a second year resident has the ability to perform an entire total knee arthroplasty independently. Therefore, the trainee should be rated at the bottom end of the scale for some of the items assessed by this tool. We also felt that raters would be more likely to use the whole scale if the high stakes of procedural evaluation was decreased by using multiple observers across multiple attempts of the same procedure, similar to the mini-CEX(18,19).
An expert group of medical educators adapted previously validated scales (9,14) to develop a unique 12-item tool (two yes/no questions & 9 items rated on 5 point scale) for the assessment of surgical competence. Items addressed included the trainee’s preparation for the case, surgical skills (technical, visuo-spatial, etc) and intraoperative communication, as well as an assessment of whether the trainee is competent to perform that particular procedure. The scale was piloted in the Division of Orthopedic Surgery at the University of Ottawa. Seventy-two assessments were completed by 11 surgeons on 20 residents with an average of 3.15 procedures per resident. A generalizability analysis demonstrated high reliability (0.82) but also high correlations between items. One of the most important findings from the pilot study was that there was a significant effect of PGY level with higher scores for each increase in training level except for PGY 4 and 5, regardless of procedure type. This was an indication that raters were more comfortable in using the entire scale to rate performance, showing the value of using an external reference criterion. Post-study focus groups indicated that staff surgeons found it easy to use. Trainees felt that it helped define important aspects of the case, as well as improving the amount and quality of feedback. (Unpublished work, presented at CCME)

This scale is one of the first technical performance assessment tools to differentiate between year of training. Almost all of the other tools in the literature have only shown sensitivity to gross differences in level of training. Based on the qualitative and quantitative feedback the scale was revised to a 9-item tool called the O-SCORE (Ottawa-Surgical Competency Operating Room Evaluation – Appendix B). Continued work evaluating this scale across specialties is currently underway (Department of surgery Grant – $25,000)

Development of an Electronic O-SCORE and Competency Based Procedure Log

At present the Orthopedic Surgery Department requires trainees to document operative experiences using T-RES©, an electronic procedure log supporting a variety of mobile platforms. This software currently allows residents to self-evaluate and faculty to evaluate residence procedure performance through a separate ‘faculty’ log in. The current evaluation software is ‘performance’ based rather than ‘competency’ based limiting its potential to demonstrate a resident’s competence for independent practice with a particular procedure. We also consider the current requirement for the evaluator to record the evaluation under a separate log in as a potential barrier to timely feedback and completion of the evaluation. Now the evaluator who ‘scrubs out’ early may leave the OR to complete the evaluation on their own computer, or may only complete the evaluation on a delayed basis if prompted by an email. As a result the validity of the retrospective evaluation is reduced and the opportunity for face-to-face feedback is lost. It is for this reason that the Orthopedic program (and many others) use the software mainly as a procedure log and do not utilize the evaluation portion of the software.

We have initiated discussions with T-RES and a verbal agreement is in place to modify the current software to allow the staff to ‘securely’ complete the O-SCORE evaluation on the resident’s mobile device or operating room computer. This will ensure the evaluation is immediate and face-to-face in an effort to maximize feedback opportunities and ensure the integrity of the evaluation. This modification will require a significant amount of programming time and expense, however T-RES© has verbally agreed to move foreword with a small financial commitment to the process by our institution.

The current T-RES© software staff evaluation will be converted to the O-SCORE. The paper based O-SCORE has 2 prompts for feedback where the staff documents a positive aspect of the performance and something the trainee needs to work on for improvement. It is
anticipated evaluators may find this more challenging than with the written format. The scale may require alteration, utilizing prompts encouraging this feedback orally in an effort to ensure feasibility of use.

Development of an electronic O-SCORE within T-RES©, a proven procedure log, will create a competency based procedure log. This should be of value in demonstrating competency in specialty defined core procedures. The electronic compilation of data across a large group of residents could allow for the determination of the ‘ideal’ number of cases for the average resident to achieve competency in a specific procedure. This information would be of significant value in identifying residents ready for progression, those in need of extra assistance or even in residency program design.

Objectives

This study is designed to assess the validity and feasibility of competency based surgical evaluation tool to an electronic format. Secondary objectives will compare the frequency and quality of feedback to the paper-based tool and assess feasibility of combining the O-SCORE with an electronic procedure log in an effort to develop a valid competency based procedure log.

Problem Statement

Surgical training programs are responsible for ensuring the competency of their graduates. This includes their ability to perform the various surgical procedures safely and successfully. In most training programs, a resident’s ability to do these procedures (i.e., shows how) is evaluated primarily by general observations documented under the single item of “technical skills” on an In-training Evaluation Report (ITER) (1). While trainees are increasingly required to maintain procedure logs, unfortunately the result is often a log of procedural performance lacking content validity rather than a reflection of trainee’s operative ability.(7,8) By linking a valid competency based surgical evaluation with a procedure log, a log of procedural competency can be developed. This would be a powerful tool in proving surgical competency for index procedures, identifying residents in need of assistance or in efforts to streamline training programs.

Reference List/Literature Review – See Appendix A

Methodology

Development of Software – Work with T-RES will focus on the modification of the current T-RES software to incorporate the O-SCORE competency format and to allow the staff to securely submit an evaluation on the resident’s electronic collection device.

Participants – Trainees and Staff Surgeons in the Divisions of General and Orthopaedic Surgery at The Ottawa Hospital.

Consent - Consent will be obtained from the trainees and the evaluating staff surgeons. Trainees will be informed that these performance assessments will not be used as part of their regular rotation evaluations. They will be informed of their right to decline the offer of participating. To protect the confidentiality of the ratings, trainees and will be assigned a coding number by a neutral third party person. None of the principal investigators will have access to these identification codes. When reporting results, only aggregate ratings will be released.

Data Collection - Over a 4 month period, six common orthopedic surgical procedures that all levels of residents would routinely participate in will be assessed: Open Reduction Internal
Fixation (ORIF) of a wrist, ORIF of a hip (dynamic hip screw (DHS) or cephalomedullary nail), ORIF ankle, hip bipolar hemiarthroplasty, primary total hip arthroplasty, and knee arthroscopy. Each time a trainee participates in one of these procedures they will be assessed using the modified electronic procedure log by the supervising surgeon. It is expected that a total of 20 post-grad trainees in the division of orthopedic surgery will each have at least 6 procedures evaluated over that time period.

**Quantitative Analysis.** The quantitative analysis will focus on reliability and validity. The psychometric characteristics of the scale will be determined assessed using descriptive statistics for each item and an analysis of missing data and the internal consistency. In addition, evidence for scale validation will be examined by comparing performance of different groups of trainees and by seeking post hoc feedback from all participants. Results will be compared with data from a pilot studies run in an identical fashion using a similar paper based scale.

**Qualitative Analysis.** All participants will be given the opportunity to provide written feedback on the clarity and utility of the scale (Appendix C). As well, two focus groups will be held; one for the staff and one for the trainees. The results of the quantitative analysis of the scale will be presented, in aggregate format, to the focus groups. The facilitator of the focus group will explore the participants thoughts on what worked and did not work with regards to the use of the electronic scale, and the frequency and quality of feedback in the new format. The transcripts of the focus groups will be analyzed for recurrent themes. Information from both the psychometric analysis, the participant feedback form and the focus groups will be used to assess validity of the electronic version of the O-SCORE, as well as issues with feasibility related to this electronic competency based procedure log.

**Ethics Approval** - pending

**Project Time Line** – 12 Months- See Appendix D

**Anticipated Outcomes**

We anticipate that the O-SCORE will continue to reliably discriminate between junior and senior surgical residents, and ideally between specific years. We also anticipate that the O-SCORE will remain valid and reliable in the electronic format.

We anticipate that the feasibility of using the scale will be improved using an electronic format, and that it will continue to encourage timely feedback for the resident. These results will be used to inform further development of the electronic scale if necessary. The development of a competency based procedure log will allow for us to assess the feasibility of using this tool to determine the “average” number of procedures required to achieve competency in specific procedures.

**Relevance to Postgraduate Medical Training**

With the rise of professional accountability, surgical training programs and licensing bodies will be increasingly responsible for ensuring the clinical and surgical competence of their graduates. Our surgical assessment tool, anchored by competency for independent practice as a maximum score was able to differentiate between PGY levels. Trainees found the tool useful for providing both direction to successful completion of a case and the provision of formative feedback. The development of a feasible and valid competency based electronic surgical procedure log could be of value in demonstrating competency in specialty defined core
procedures. The electronic compilation of data across a large group of residents could allow for the determination of the ‘ideal’ number of cases for the average resident to achieve competency in a specific procedure. This information would be of significant value in identifying residents ready for progression, those in need of extra assistance or even in residency program design.

**Plan to disseminate Research Findings**

Outcomes of both the quantitative and qualitative analysis will be submitted for presentation at a national (Canadian Conference on Medical Education) and/or international medical/surgical education (Association for Surgical Education) conferences as well as submitted for publication in a peer-reviewed journal.

**Budget- see Appendix E**
Appendix A – Reference List


Appendix B – The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE)

<table>
<thead>
<tr>
<th>Trainee:</th>
<th>Level:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Staff:</th>
<th>Date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Relative complexity of this procedure to average of same procedure

<table>
<thead>
<tr>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
</table>

The purpose of this scale is to evaluate the trainee’s ability to perform this procedure safely and independently. With that in mind please use the scale below to evaluate each item, irrespective of the resident’s level of training in regards to this case.

Scale
1 – “I had to do” – *i.e. Requires complete hands on guidance, did not do, or not given the opportunity to do*
2 – “I had to talk them through” – *i.e. Able to perform tasks but requires constant direction*
3 – “I had to prompt them from time to time” – *i.e. Demonstrates some independence, but requires intermittent direction*
4 – “I needed to be in the room just in case” – *i.e. Independence but unaware of risks and still requires supervision for safe practice*
5 – “I did not need to be there” – *i.e. Complete independence, understands risks and performs safely, practice ready*

1. **Pre-operative plan**
   Gathers/assesses required information to reach diagnosis and determine correct procedure required
   1 2 3 4 5

2. **Case Preparation**
   Patient correctly prepared and positioned, understands approach and required instruments, prepared to deal with probable complications
   1 2 3 4 5

3. **Knowledge of specific procedural steps**
   Understands steps of procedure, potential risks and means to avoid/overcome them
   1 2 3 4 5

4. **Technical Performance**
   Efficiently performs steps avoiding pitfalls and respecting soft tissues
   1 2 3 4 5

5. **Visuospatial Skills**
   3D spatial orientation and able to position instruments/hardware where intended
   1 2 3 4 5

6. **Post-operative plan**
   Appropriate complete post-operative plan
   1 2 3 4 5

7. **Efficiency and Flow**
   Obvious planned course of procedure with economy of movement and flow
   1 2 3 4 5

8. **Communication**
   Professional and effective communication/utilization of staff
   1 2 3 4 5

9. Resident is able to safely perform this procedure independently (circle) Y N

10. Give at least 1 specific aspect of procedure done well
11. Give at least 1 specific suggestion for improvement

Signatures: Staff:

Trainee:

**Appendix C – Individual Participant Survey Tool**

What feature(s) of this electronic tool did you like?

What feature(s) of this electronic tool did you not like?

Do you have any suggestions for improving the electronic tool?

Do you feel that the tool improved the quality or quantity of feedback you would normally give or receive?

If you were involved in the original surgical evaluation pilot program, do you feel the amount or quality of your feedback was different with the electronic evaluation tool in comparison to the paper based tool.

Would you be supportive of this evaluation tool being used in this program?

If not, why not?
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Modification of software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial of software</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction and education session for evaluators and trainees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative Data Collection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative Data Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis of Qualitative Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manuscript preparation and Submission</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix E – Budget

<table>
<thead>
<tr>
<th>PERSONEL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Wade Gofton, Steve Papp (10 hrs/wk)</td>
<td>In Kind</td>
</tr>
<tr>
<td>Research Assistant – Data Collection at Civic Site (1/2 day per week)</td>
<td>2,520</td>
</tr>
<tr>
<td>Research Assistant – Data Collection at General/Riverside site (1/2 days per week x 4 months, 2 days per week x 4 months)</td>
<td>2,520</td>
</tr>
<tr>
<td>Nancy Dudek – Leading focus groups</td>
<td>In Kind</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MATERIALS AND SUPPLIES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Software Development</td>
<td></td>
</tr>
<tr>
<td>Company to contribute $30,000</td>
<td>10,000</td>
</tr>
<tr>
<td>U of O to contribute $10,000</td>
<td></td>
</tr>
<tr>
<td>Collection hardware</td>
<td></td>
</tr>
<tr>
<td>To be determined</td>
<td>6250 vs 2750</td>
</tr>
<tr>
<td>25 X 8GB itouch vs 5 fixed ipad input devices</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TRAVEL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Presentation of results by PI at Royal College or Surgical Education Conference</td>
<td>1,500.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTHER EXPENSES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Introductory training sessions – Food for participants ($20 x 40)</td>
<td>800.00</td>
</tr>
<tr>
<td>Focus group - Food for participants (5 participants per group + facilitator = 12 people x$20)</td>
<td>480.00</td>
</tr>
<tr>
<td>Transcription of focus group discussions (30 pages/hr x 4 hrs)</td>
<td>720.00</td>
</tr>
<tr>
<td>Qualitative Analysis of transcripts</td>
<td>In kind</td>
</tr>
<tr>
<td>Quantitative Analysis by Statistician (10 hrs x $75)</td>
<td>750.00</td>
</tr>
<tr>
<td>Grand Total</td>
<td>25,540</td>
</tr>
</tbody>
</table>

Research Assistant:
During the four months of data collection the research assistants will be responsible for the daily distribution and collection of research forms in an effort to maximize returns (Civic Site - This individual will be employed for a one day per week equivalent for a period of 4 months - $35/hr x 1/2 day/week x 4 months - $5040; General site – This individual will be employed for a ½ day per week equivalent to identify appropriate cases and remind OR team to complete an evaluation equivalent for a period of 4 months $35/hr x 1/2 day/week - 2050) {manages forms at the general and riverside campus sites and is involved in analysis}). While the research assistant expenses are significant, based on the pilot study we feel that the involvement of research personnel is essential in ensuring a high rate of evaluation form return, and success of the study.

Equipment
The equipment requirements are negligible. Rating scales will be paper based.
Materials and Supplies
We will work with the company to finance this move forward. The majority of the software development will be borne by the company.

To ensure adequate participation we will need to provide itouch mobile devices for residents, or we will utilize ipads fixed in the OR. The feasibility of each will be reassessed following final software development and trialing. This hardware can be used for future studies.