EDITORIAL

Measuring resident operative skills**John R. W. Kestle, MD**

Department of Neurosurgery, University of Utah, Salt Lake City, Utah

IN this issue, Hadley et al. describe the use of the Objective Structured Assessment of Technical Skills for Surgery (OSATS) global rating scale to evaluate trainees on their pediatric neurosurgery service.¹ Six faculty members scored 1 fellow and 8 residents on 299 procedures. In addition, the trainees assessed themselves using the scale. The authors found concordance between faculty scores and resident self-assessment scores. In addition, for more difficult procedures, the faculty members gave senior residents higher scores than juniors. On the self-assessment scores for difficult procedures, the senior residents gave themselves higher scores than the junior residents gave themselves. Based on these findings, the authors plan to use the OSATS to assess resident operative skill development and improve resident feedback.

The objective assessment of resident performance in the operating room has received little attention in neurosurgery. Our evaluation process varies from one program to another, and I suspect many use measurement scales that have not been validated. The work of Hadley et al. is therefore a welcome addition to the literature.

The OSATS was developed and found to be reliable and valid in surgical simulation scenarios.⁵ In addition, it has been used in the operating room in gynecological surgery and general surgery.^{2,3,4} So is further testing really necessary? Strictly speaking, it is. When a measurement scale is developed and tested in one population, it is not necessarily reliable and valid in a different one. Hadley et al. have therefore appropriately begun to address these issues in the pediatric neurosurgical operating room situation.

In order to test a measurement scale, the most important properties are reliability, validity, and responsiveness.

Reliability

Reliability means that the same scores are obtained on repeated measurements. Test-retest reliability means that an observer gets the same score when measuring on 2 separate occasions. Interobserver reliability means that 2 different observers get the same score when measuring on the same occasion.

The first step in testing a new scale is the assessment of

reliability. Although the authors stated this as a goal, it was not done. It may be difficult to have multiple people observing residents at the same time, but perhaps this could be done using operative videos. It is an important part of the process and should be the subject of future work.

Validity

Validity asks, “Does the scale measure what we think it is measuring?” There are a number of aspects to demonstrating validity. A common one is construct validity. This is demonstrated by making (and testing) predictions about how the new measurement scale will behave in certain situations.

The authors do have some data to support validity. The residents gave themselves scores similar to those assigned by faculty members, and for difficult procedures, senior residents scored better than junior residents. These findings fit with our expectations that performance improves with time and experience and therefore supports the construct validity of the OSATS in this scenario.

Responsiveness

Responsiveness means that the scale is able to detect the smallest change that is clinically important. Ideally, a scale that measures operative skill would detect the smallest improvement that results in better patient outcome, or reduces risk or cost. This would probably be difficult to demonstrate and would not usually be assessed until reliability and validity were known.

Objective accurate assessment of resident performance in and out of the operating room is important. Personally, I do not think that scales such as this should replace constructive comments at the time of the operation and providing residents with one-on-one feedback regarding operative skills. Both methods of feedback and evaluation are important and they complement each other. With increasing standardization of teaching and evaluation methods, validated scales will be needed. This one requires additional work, but appears to be a step in the right direction.

<http://thejns.org/doi/abs/10.3171/2015.2.PEDS1542>

References

1. Hadley C, Lam SK, Briceño V, Luerksen TG, Jea A: Use of a formal assessment instrument for evaluation of resident operative skills in pediatric neurosurgery. **J Neurosurg Pediatr** [epub ahead of print August 28, 2015. DOI: 10.3171/2015.1.PEDS14511]
2. Hiemstra E, Kolkman W, Wolterbeek R, Trimbos B, Jansen FW: Value of an objective assessment tool in the operating room. **Can J Surg** **54**:116–122, 2011
3. Hopmans CJ, den Hoet PT, van der Laan L, van der Harst E, van der Elst M, Mannaerts GH: Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. **Surgery** **156**:1078–1088, 2014
4. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al: Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. **Surg Today** **43**:271–275, 2013
5. Reznick RK, MacRae H: Teaching surgical skills—Changes in the wind. **N Engl J Med** **355**: 2664–2669, 2006

Disclosure

The author reports no conflict of interest.

Response

**Caroline Hadley, BS, Sandi K. Lam, MD, MBA,
Valentina Briceño, RN, Thomas G. Luerksen, MD, and
Andrew Jea, MD**

Division of Pediatric Neurosurgery, Texas Children's Hospital, and
Department of Neurosurgery, Baylor College of Medicine, Houston,
Texas

We thank Dr. Kestle for his editorial comments regarding use of the OSATS global rating scale for evaluation of resident operative performance in pediatric neurosurgery. His comments highlight important points: the need for a validated tool for evaluation of resident skill that can be used across residency training programs and the strengths and limitations of the tool under study.

Currently, objective evaluation of neurosurgery residents is highly variable between programs. Although milestones of proficiency have been established, the use of validated tools for assessment of technical skill has not become routine or standardized.⁶ Use of such a tool will be beneficial in providing feedback to trainees as well as in ensuring adequacy of training. As Dr. Kestle points out, although the OSATS has been validated previously in general surgery, a tool like this has not been used previously in neurosurgery.^{1–5}

We agree with Dr. Kestle that, in order to be useful, an objective assessment tool must demonstrate reliability, validity, and responsiveness. The current report reflects the first 6 months of using the OSATS. Because we chose to integrate the tool into the workflow, using it as a means to directly evaluate performance, we were unable to have multiple faculty evaluators in each surgery. Moreover, we were unable to blind faculty evaluators to the identity and level of training of the resident. This introduces bias, but it also means that we were unable to compare faculty evalua-

tions to assess the inter-rater reliability of the tool. We hope to address both of these limitations in the future by using video evaluation to permit both blinding and comparison of multiple faculty ratings of the same performance.

Our current results do support the validity of the tool, as resident and faculty scores were similar, and both scores reflect improvement, particularly in expert procedures, between levels of training. One limitation that we plan to address is the level of involvement in the procedure. Currently, there is no indicator for how independent the resident was during the procedure. Addition of this metric will further allow for comparison of technical skill between levels of training. Additionally, we hope to examine longitudinal data for each resident, to determine whether the OSATS reflects improvement across the duration of each rotation and between their junior and senior rotations.

We also agree that the responsiveness of this tool will be difficult to measure. The advantage of the OSATS is that it can be completed quickly, making it easy to integrate into the workflow. However, this means it is a less detailed assessment and may be less sensitive to minor changes.

Finally, we feel that Dr. Kestle is correct in his assertion that this tool should not be used in place of direct verbal feedback during and after operative cases. This tool is intended to help measure performance quantitatively over time and to show residents areas in which they are deficient and those in which they excel. This is complementary to verbal and written feedback. It can also help residency programs ensure residents are progressing appropriately in their operative skills. It is not intended to replace traditional feedback.

More work is required to fully evaluate this tool for use in neurosurgery. Our preliminary findings suggest that use of the OSATS in evaluation of neurosurgery resident operative performance is both feasible and valid and it is worthy of further investigation and use.

References

1. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB: Observational tools for assessment of procedural skills: a systematic review. **Am J Surg** **202**:469–480, 480.e1–480.e6, 2011
2. Doyle JD, Webber EM, Sidhu RS: A universal global rating scale for the evaluation of technical skills in the operating room. **Am J Surg** **193**:551–555, 2007
3. Faulkner H, Regehr G, Martin J, Reznick R: Validation of an objective structured assessment of technical skill for surgical residents. **Acad Med** **71**:1363–1365, 1996
4. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al: Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. **Surg Today** **43**:271–275, 2013
5. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W: Testing technical skill via an innovative “bench station” examination. **Am J Surg** **173**:226–230, 1997
6. Selden N (ed): **The Neurological Surgery Milestone Project**. Chicago: Accreditation Council for Graduate Medical Education; Woodbridge, CT: American Board of Neurological Surgery, 2013. (<http://www.acgme.org/acgmeweb/Portals/0/PDFs/Milestones/NeurologicalSurgeryMilestones.pdf>) [Accessed June 29, 2015]