Student assessment by objective structured examination in a neurology clerkship

Rimas V. Lukas, MD
Taiwo Adesoye, MD
Sandy Smith, PhD
Angela Blood
James R. Brorson, MD

Correspondence & reprint
requests to Dr. Lukas:
rlukas@neurology.bsd.uchicago.edu

ABSTRACT

Objectives: We evaluated the reliability and predictive ability of an objective structured clinical examination (OSCE) in the assessment of medical students at the completion of a neurology clerkship.

Methods: We analyzed data from 195 third-year medical students who took the OSCE. For each student, the OSCE consisted of 2 standardized patient encounters. The scores obtained from each encounter were compared. Faculty clinical evaluations of each student for 2 clinical inpatient rotations were also compared. Hierarchical regression analysis was applied to test the ability of the averaged OSCE scores to predict standardized written examination scores and composite clinical scores.

Results: Students' OSCE scores from the 2 standardized patient encounters were significantly correlated with each other ($r = 0.347$, $p < 0.001$), and the scores for all students were normally distributed. In contrast, students' faculty clinical evaluation scores from 2 different clinical inpatient rotations were uncorrelated, and scores were skewed toward the highest ratings. After accounting for clerkship order, better OSCE scores were predictive of better National Board of Medical Examiners standardized examination scores ($R^2\Delta = 0.131$, $p < 0.001$) and of better faculty clinical scores ($R^2\Delta = 0.078$, $p < 0.001$).

Conclusions: Student assessment by an OSCE provides a reliable and predictive objective assessment of clinical performance in a neurology clerkship. *Neurology*® 2012;79:681-685

GLOSSARY

NBME exam = National Board of Medical Examiners Clinical Neurology Subject Examination; **NICU** = neurointensive care unit; **OSCE** = objective structured clinical examination; **SP** = standard patient.

There is a developing body of research on clinical neurology education.¹ One goal in neurology education is improved student assessment using valid objective methods demonstrating consistency and also providing well-distributed scores, avoiding the grade inflation prevalent with faculty evaluations. Semiobjective systems using nonstandardized patients to evaluate performance on neurology clerkships have been reported. These have shown, however, limited ability to predict subjective clinical scores in neurology.² An alternative approach is use of an objective structured clinical examination (OSCE) in neurology.³⁻⁶ The OSCE involves a series of encounters objectively testing clinical competencies including history taking, physical examination, communication, and data interpretation. It can be both a formative and summative evaluation.⁷ Feasibility of use of an OSCE in the examination of medical students at the completion of a neurology clerkship has been reported,^{3,4,6,8-10} but assessment of its reliability as an evaluation tool in a neurology clerkship has not been described.

We recently implemented an OSCE as a supplementary tool for the objective assessment of clinical skills of medical students completing the neurology clerkship. We assessed the parallel forms reliability of this OSCE by comparing student performance across 2 separate standardized patient (SP) experiences. We also tested the OSCE's ability to predict outcomes on other

Supplemental data at
www.neurology.org

Supplemental Data

From the Department of Neurology (R.V.L., J.R.B.), Pritzker School of Medicine (A.B.), Pritzker School of Medicine (retired) (S.S.); Simulation Center (A.B.), University of Chicago, IL; and Department of Surgery (T.A.), University of Wisconsin, Madison.

Go to Neurology.org for full disclosures. Disclosures deemed relevant by the authors, if any, are provided at the end of this article.

measures of knowledge and performance. We report evidence that the neurologic OSCE provided an assessment of student skills demonstrating parallel forms reliability as well as predictive ability for performance on faculty evaluations and the National Board of Medical Examiners Clinical Neurology Subject Examination (NBME exam).

METHODS **Standard protocol approvals, registrations, and patient consents.** This project was submitted to the institutional internal review board, and it was deemed to be exempt from review because it fit into the minimal risk category of research conducted in established or commonly accepted educational settings, involving normal or special educational practices, strategies, or comparison of techniques.

The neurology OSCE at the University of Chicago Pritzker School of Medicine, administered on the final day of a 4-week required clinical clerkship, consists of 2 encounters with trained SPs following scripts depicting common neurologic presentations. The specific scripts and scoring checklists were developed with the assistance of the University of Chicago Clinical Performance Center. They are in submission to MedEd Portal (<https://www.mededportal.org/>). During a 30-minute SP encounter, students must complete a focused history and the neurologic examination. They then complete a written report online describing the history, physical examination, assessment, and plan. The neurologic examination includes performance of a defined standard screening neurologic examination, taught during the clerkship, as well as the potential to expand the examination in a problem-focused manner. The assessment calls for both lesion localization and a prioritized differential diagnosis. The management plan should include both diagnostic and therapeutic components as well as considerations of longer term care. Students are not informed of the clinical scenarios before the OSCE.

Students' scores from each SP experience within the OSCE are averaged from 3 separate components. Two scores come from checklists completed by the SP: one for the student's taking of the history and one for performance of the neurologic examination. The third component is a score of the written formulation, assigned by faculty (J.R.B., R.V.L.) based on defined checklists of elements that should be included in each of the 4 sections (subjective, objective, assessment, and plan). The OSCE scores from the 2 separate SP encounters are then averaged.

Each student's clerkship experience consisted of a combination of clinical rotations depending on his or her site assignment and included 1 to 2 weeks of service on a general adult neurology service (either primary or consultation) and on either a neurointensive care unit (NICU) or a stroke consultation service. The first year's cohort of students ($n = 92$) also completed 1 week in the pediatric neurology inpatient service in place of an additional week on the adult neurology ward. The composite clinical scores and the average OSCE scores of the 2 groups of students (2008–2009 and 2009–2010) with slightly different clinical experiences were compared using t tests. We found no significant differences between the composite clinical scores ($t = 1.58$, $p = 0.116$) or average OSCE scores ($t = -0.27$, $p = 0.979$) and therefore treated the students as one group for further analyses. Specific learning objectives for the neurology clerkship are available in appendix e-1 on the *Neurology*® Web site at www.neurology.org. On each clinical rotation, students were evaluated by experienced clinical faculty using written forms that

called for categorical ratings in 5–8 areas and for written comments. The categories (and the corresponding assigned numerical scores) were Not observed (no score), Below expectations (2), Meets expectations (3), Exceeds expectations (4), or Top 10% (5). The assigned ratings were averaged across clinical areas to give the clinical attending score. When Not observed was indicated, it was not included in the average score. If 2 different clinical faculty members provided evaluations of a student during a clinical rotation, the scores were averaged to provide the clinical attending score for the rotation. A composite clinical score was then calculated as follows: 60% consisted of the faculty scores from individual rotations, weighted in proportion to the time spent on each rotation. The remainder was made up of 20% derived from a monthly small group teaching attending, 10% from a score based on activity in the outpatient clinic, and 10% from a score of a written patient case description.

All students also took the NBME exam on the last day of the clerkship. The OSCE score and the scores on the NBME exam were used in combination with the composite clinical score to determine final grades for the clerkship, with the OSCE and the NBME exam each receiving 25% weighting and the clinical score 50% weighting. All scores were compiled in a spreadsheet, and student identifiers were removed. Statistical analyses were performed using IBM SPSS version 18.

Interrater reliability of the written formulation component of the OSCE was evaluated by rescoring of 50 randomly selected SP write-ups by a second faculty member, blinded to the initial score given, with comparison to the original assigned score. To evaluate the parallel forms reliability of the OSCE, the 2 individual SP encounter scores that comprise the OSCE score were compared. We performed a similar comparison of the faculty scores from the 2 clinical rotations for each individual student. These correlations were assessed by calculation of Pearson correlation coefficient. Intraclass coefficients were also calculated and gave essentially identical results. We also tested how well the OSCE scores could predict NBME exam scores as well as composite clinical scores using hierarchical regression analysis. Hierarchical regression allows assessment of the independent contributions of several predictors, incorporated sequentially into the statistical model, to the variance in the dependent variable. In our analysis, it allowed us to control for the variable of clerkship order, which might influence scores through the increasing experience gained by medical students during the third year of medical school. Two hierarchical regressions were performed. In the first, clerkship order and OSCE scores were evaluated as predictors of the NBME exam score. In the second, clerkship order and OSCE scores were evaluated as predictors of the composite clinical score. The change in the predictive strength of the model conferred by inclusion of the OSCE scores was expressed as $R^2\Delta$. Original data are available in appendix e-2.

RESULTS Over the 2-year period of July 2008 through June 2010, 204 students completed the neurology clerkship at the University of Chicago. Of these, 9 were students in their fourth year of medical school, and these were excluded from analysis, leaving a cohort of 195 third-year students completing the clerkship. Of these students, 162 were assigned to the University of Chicago Medical Center site and 33 to the NorthShore Medical University site.

Each student completed clinical rotations including general adult inpatient neurology or neurology inpatient consult service, a NICU or stroke consult service, and an outpatient specialty clinic rotation, as well as didactic sessions including lectures, workshops, and weekly case discussion sessions with an assigned teaching attending. Students were assessed by their supervising faculty using standard written evaluation forms calling for both categorical ratings and qualitative comments and by a written examination and an OSCE on the last day of the clerkship.

We assessed the distribution of faculty scores from both the general inpatient rotation and NICU/stroke service rotation (figure 1A). Scores were non-normally distributed ($p < 0.001$, Kolmogorov-Smirnov test) with a distinct rightward skew toward higher scores (-0.618), the most frequently assigned score being the highest rating (Top 10%). Although each score represented an average across 5–8 clinical categories, the most common scores were 4.0 and 5.0, indicating a faculty preference for selection of a single category across all the multiple clinical areas queried. In contrast to the attending scores, OSCE scores from individual standardized patients, as depicted in figure 1B, were more normally distributed (mean 81, SD 8) and were not significantly skewed (calculated skew of -0.472 ; kurtosis 0.492).

As stated earlier, student grades were determined from subjective (composite clinical score) and objective (OSCE score and NBME exam) components. The NBME exam scores for this cohort of students showed a median value of 78, and, when compared with national norms, a median percentile value of 76. The composite clinical score distribution, primarily derived from the faculty scores, again showed a substantial shift toward higher scores compared with either OSCE scores or NBME exam scores (figure 1C).

We examined the interrater reliability of 2 faculty members scoring the same OSCE write-ups ($n = 50$) and found that faculty scores were closely correlated ($r = 0.87$, $p < 0.001$). We next examined the parallel forms reliability of scores within the OSCE and attending scores between the adult and NICU/stroke rotations. OSCE scores for individual students were significantly correlated ($r = 0.347$, $p < 0.001$), whereas there was no correlation between faculty assessments of student performance ($r = -0.055$, $p = 0.447$) (figure 2). Thus, although the OSCE scores showed evidence of moderate reliability, the faculty evaluations did not demonstrate this.

We also examined the effects of increasing clinical experience during the clinical year of training on student performance, by examining the effect of clerkship order on performance. Clerkship order as the academic year progressed was shown to be correlated

(A) Student scoring by faculty in individual clinical inpatient rotations. Attending scores were skewed toward the maximum possible score of 5 and were not smoothly or normally distributed. (B) Distribution of objective standardized clinical examination (OSCE) scores in individual standardized patient encounters. Scores were normally and broadly distributed. (C) Comparative distributions of composite clinical scores, OSCE scores, and scores on the National Board of Medical Examiners Clinical Neurology Subject Examination (NBME exam). NICU, neurointensive care unit.

to higher OSCE scores ($r = 0.199$, $p = 0.005$) and higher NBME exam scores ($r = 0.340$, $p < 0.001$). In contrast, there was no significant correlation between clerkship order and composite clinical score ($r = 0.109$, $p = 0.129$).

To examine how student performance in the OSCE related to other measures of student competency, we first examined the correlations of OSCE scores with clinical scores and NBME exam scores. Modest correlations between higher OSCE scores and higher clinical scores and NBME exam scores were evident, with substantial scatter indicating that

Figure 2 Objective standardized clinical examination (OSCE) scoring shows greater consistency than does scoring by clinical attendings

(A) Comparison of average attending scores for individual students from the general ward rotation with those for individual students in the neurointensive care (NICU) or stroke service rotation showed no significant correlation between these 2 sources of student evaluation ($r = -0.055$; $p = 0.45$). (B) Comparison of student OSCE scores from one standardized patient encounter with those from the other showed a moderate correlation in scores ($r = 0.347$; $p < 0.001$).

OSCE scores provide a measure of competence with substantial independence from the other 2 scores.

To statistically evaluate the contribution of the OSCE score in predicting other measures of student performance, we performed 2 hierarchical regression analyses (figure 3). In the first, we used the OSCE score to predict the NBME exam score, taking into account clerkship order. Clerkship order was a significant predictor of NBME exam scores ($F_{1,193} = 25.17$, $p < 0.001$), accounting for 11.5% of the variability in NBME exam scores. The OSCE score significantly added to the initial model predicting NBME scores ($F_{1,192} = 33.34$, $p < 0.001$, $R^2\Delta = 0.131$), accounting for an additional 13% of variance

Figure 3 Objective standardized clinical examination (OSCE) performance predicts both composite clinical scores and written examination scores

(A) Student scores on the OSCE correlated with their composite clinical scores ($r = 0.296$; $p < 0.001$). (B) Student scores on the OSCE also strongly correlated with their scores on the National Board of Medical Examiners (NBME) written examination ($r = 0.422$; $p < 0.001$).

in NBME exam scores. Approximately 25% of the variability in NBME exam scores was attributable to rotation order and OSCE performance. In the second analysis, we used the OSCE score to predict the composite clinical score. Clerkship order was not a significant predictor of composite clinical score, accounting for only approximately 1% of the variability. The OSCE score significantly added to this predictive model ($F_{1,192} = 16.48$, $p < 0.001$, $R^2\Delta = 0.078$), accounting for an additional 7.8% of variance in composite clinical scores.

DISCUSSION Efforts are continually being made to improve the assessment of medical students during their medical education. Attempts to evaluate not only the breadth and depth of students' knowledge but also the ability of students to apply this knowledge clinically have long been present in the form of faculty evaluations. Such traditional evaluations, based on impressions gained during encounters at the bedside or conference table during clinical service rotations, can be subjective and substantially influenced by faculty experience and temperament as well as by variability in available clinical load.^{2,11} Systems that add greater objectivity in the evaluation of medical students' neurology skills have been reported.^{2,12,13} We implemented an OSCE in the required clinical clerkship in neurology and have examined the performance of the OSCE in the objective evaluation of students.

The OSCE provided a broad normal distribution of student scores with a less pronounced rightward skew toward grade inflation and without the binning of scores present in the distribution of faculty clinical scores, providing greater ability to distinguish students' performance along a continuum. The SP scores generated by individual students correlated with one another, whereas the 2 independent faculty evaluation scores did not, thus suggesting that the OSCE is a more reliable measure of student clinical performance than is the clinical rating by faculty. The lack of consistency between faculty scores may be in part due to individual faculty members assessing students approximately once or twice per year and having limited access to a comparison group. In contrast, SPs have more extensive and frequent exposure to students and may be able to perceive smaller variations between student performances. These findings support the parallel forms reliability of the alternate OSCE scenarios in evaluating students.

To further assess the validity of the OSCE as a measure of student abilities in neurology that are relevant to clinical success, we examined its relationship to other variables, in particular, other external performance measures available such as performance on the

NBME exam, composite clinical scores, and final grade. In hierarchical regression analyses, the OSCE was a significant predictor of performance on the NBME exam and the composite clinical score, the traditional measures for the evaluation of student performance, with clerkship order accounted for. Taken together, these features support the OSCE as a valid objective assessment measure for clinical ability in neurology.¹⁴ In addition, the OSCE measures skills distinct from those assessed by faculty clinical evaluations and standardized written examinations. Whereas the NBME exam evaluates a student's knowledge, the OSCE tests more hands-on basic clinical skills and application of neurologic knowledge to clinical situations. Faculty evaluations also assess clinical skills; however, they are limited by skewing of grades and lack of correlation between faculty members. The OSCE is thus an effective tool to use as a part of the comprehensive assessment of medical students at the completion of a neurology clerkship. It may serve as a tool by which to predict students' performance as future clinicians. This will require additional longitudinal study.

It has been demonstrated that the process of being tested enhances knowledge retention, particularly when testing occurs over time.¹⁵ It has also been shown that standardized patient encounters focused on the neurologic examination correlate with improved performance on neurology-specific OSCEs even 2 years after the initial encounter.⁸ Participation in an OSCE at the completion of a neurology clerkship may thus also be of formative value; in particular, it may be of benefit in preparation for the United States Medical Licensing Examination step 2 OSCE examination. This also requires further study.

AUTHOR CONTRIBUTIONS

R.V. Lukas, T. Adesoye, and J.R. Brorson contributed to the design and conceptualization of the study and drafting the manuscript. R.V. Lukas, T. Adesoye, A. Blood, S. Smith, and J.R. Brorson contributed to the analysis and interpretation of data as well as revising the manuscript. T. Adesoye, S. Smith, and J.R. Brorson performed the statistical analyses.

ACKNOWLEDGMENT

The authors thank Kris Slawinski at the Pritzker School of Medicine Clinical Performance Center for her contribution in implementation of the OSCE and the skilled standardized patients and students who participated in the OSCE.

DISCLOSURE

The authors report no disclosures relevant to the manuscript. **Go to Neurology.org for full disclosures.**

Received November 22, 2011. Accepted in final form March 21, 2012.

REFERENCES

1. Stern BJ, Lowenstein DH, Schuh LA. Invited article: neurology education research. *Neurology* 2008;70:876–883.
2. Schmähmann JD, Neal M, MacMore J. Evaluation of the assessment and grading of medical students on a neurology clerkship. *Neurology* 2008;70:706–712.
3. Heckmann JG, Knossalla F, Gollwitzer S, Lang C, Schwab S. OSCE in the neurology clerkship: experiences at the neurological department of the University Hospital Erlangen. *Fortschr Neurol Psychiatr* 2009;77:32–37.
4. Yoshii F. Advanced OSCE in neurology. *Rinsho Shinkeigaku* 2007;47:897–899.
5. Prislín MD, Fitzpatrick CF, Lie D, Giglio M, Radecki S, Lewis E. Use of an objective structured clinical examination in evaluating student performance. *Fam Med* 1998;30:338–344.
6. Gledhill RF, Capatos D. Factors affecting the reliability of an objective structured clinical examination (OSCE) testing neurology. *S Afr Med J* 1985;67:463–467.
7. Gupta P, Dewan P, Singh T. Objective Structured Clinical Examination (OSCE) revisited. *Indian Pediatr* 2010;47:911–920.
8. Safdieh JE, Lin AL, Aizer J, et al. Standardized patient outcomes trial (SPOT) in neurology. *Med Educ Online*. 2011;doi:10.3402/meo.v16i0.5634.
9. Schrauth M, Schmulius N, Zipfel S, Haarmeier T. Practical examinations for neurology: the Tuebingen model. *Nervenarzt* 2006;77:1464–1468.
10. Heckmann JG, Dutsch M, Rauch C, Lang C, Weih M, Schwab S. Effects of peer-assisted training during the neurology clerkship: a randomized controlled study. *Eur J Neurol* 2008;15:1365–1370.
11. Prislín MD, Fitzpatrick CF, Radecki S. A comparison of family medicine clerkship student performance across multiple teaching sites. *Fam Med* 1998;30:279–282.
12. Davis LE, King MK. Evaluating medical students' performance in a clinical setting. *Nat Clin Pract Neurol* 2007;3:702–703.
13. Davis LE, King MK. Assessment of medical student clinical competencies in the neurology clinic. *Neurology* 2007;68:597–599.
14. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006;119:166.e7–166.e16.
15. Larsen DP, Butler AC, Roediger HL 3rd. Test-enhanced learning in medical education. *Med Educ* 2008;42:959–966.