Formal faculty observation and assessment of bedside skills for 3rd-year neurology clerks

Robert Thompson Stone, MD
Christopher Mooney, PhD, MPH
Erika Wexler, MD
Jonathan Mink, MD, PhD
Jennifer Post, MD
Ralph F. Jozefowicz, MD

Correspondence to
Dr. Thompson Stone:
Robert_stone@urmc.rochester.
edu

ABSTRACT

Objective: To evaluate the feasibility and utility of instituting a formalized bedside skills evaluation (BSE) for 3rd-year medical students on the neurology clerkship.

Methods: A neurologic BSE was developed for 3rd-year neurology clerks at the University of Rochester for the 2012–2014 academic years. Faculty directly observed 189 students completing a full history and neurologic examination on real inpatients. Mock grades were calculated utilizing the BSE in the final grade, and number of students with a grade difference was determined when compared to true grade. Correlation was explored between the BSE and clinical scores, National Board of Medical Examiners (NBME) scores, case complexity, and true final grades. A survey was administered to students to assess their clinical skills exposure and the usefulness of the BSE.

Results: Faculty completed and submitted a BSE form for 88.3% of students. There was a mock final grade change for 13.2% of students. Correlation coefficients between BSE score and clinical score/NBME score were 0.36 and 0.35, respectively. A statistically significant effect of BSE was found on final clerkship grade ($F_{2,186} = 31.9, p < 0.0001$). There was no statistical difference between BSE score and differing case complexities.

Conclusions: Incorporating a formal faculty-observed BSE into the 3rd year neurology clerkship was feasible. Low correlation between BSE score and other evaluations indicated a unique measurement to contribute to student grade. Using real patients with differing case complexity did not alter the grade. **Neurology® 2016;87:2266–2270**

GLOSSARY

BSE = bedside skills evaluation; **OSCE** = objective structured clinical examination; **URSMD** = University of Rochester School of Medicine and Dentistry.

In 1976, Engel wrote, in an editorial in *JAMA*, that “relatively little time is devoted to supervised instruction in interviewing and physical examination,”¹ lamenting the academic model in which faculty made clinical decisions based on trainee evaluations without specific knowledge of the reporter’s clinical competence. Since then, the model for trainee assessment has not changed dramatically.² In a United States Medical Licensing Examination report, 40% of students stated that they were observed ≤ 4 times by faculty while performing history and examinations during medical school.³ An Association of American Medical Colleges survey found that 27% of students reported never being observed by a faculty member while performing history and physicals.⁴ The current medical education system generally utilizes narrative evaluations and written examinations to inform student grades, although research suggests that multiple assessment methods are better.⁵ In neurology, there has been an increase in objective structured clinical examinations (OSCEs) in medical students’ education, but feedback is primarily provided by standardized patients rather than faculty.^{6,7}

Few studies have examined the feasibility and utility of medical student bedside clinical evaluations during the neurology clerkship. Schmahmann et al.⁸ reported on an experience with a bedside examination exercise and Lukas et al.⁶ described utilizing an OSCE, but neither approach included direct faculty observation of students conducting a full history and

Supplemental data
at Neurology.org

From the University of Rochester, Rochester, NY.

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

examination. This study set out to evaluate whether a faculty-observed encounter of all medical students would be feasible. Secondly, we aimed to determine whether this evaluation would contribute meaningfully to the student's summative evaluation and overall clerkship experience.

METHODS This study was conducted using data collected from July 2012 through June 2014. It included 214 3rd-year medical students on their neurology clerkship at the University of Rochester School of Medicine and Dentistry (URSMD) who were evaluated using a newly designed bedside skills evaluation (BSE) administered by approximately 40 different neurology faculty. The students were in their 3rd year of 4 years of undergraduate medical education training. At URSMD, all students complete a 4-week neurology clerkship during their 3rd year. During this rotation, students are responsible for working as part of a team to evaluate, work up, and manage patients with supervision by neurology faculty and residents.

Standard protocol approvals, registrations, and patient consents. The evaluation was created and data collected as part of an educational curriculum change. Retrospectively, the data were anonymized to both student and faculty prior to analysis, and the study was approved and deemed exempt by the University of Rochester Internal Review Board.

Third-year neurology clerk evaluation. Prior to institution of the BSE, 3rd-year neurology clerks at URSMD were evaluated using 2 primary measures: (1) numeric and narrative evaluations of clinical performance (i.e., Clinical Evaluation Form) submitted by faculty and residents who rotated with the student on the wards, and (2) the National Board of Medical Examiners subject test (Shelf). A Neurology Grading Committee consisting of 5 faculty and 1 chief resident reviewed the Clinical Evaluation Forms and gave each student a consensus Clinical Score. Z scores were then used to calculate a composite score, with the Clinical Score contributing 80% and the Shelf examination contributing 20% to the final score. The final scores were rank-ordered and cutpoints were inserted such that approximately 20% of the students received an honors grade, 30% received a high pass grade, and 50% received a pass grade (as dictated by URSMD policy).

Bedside skills evaluation. A BSE was developed to assess student clinical skills via direct observation by faculty in 4 major categories: (1) history gathering, (2) neurologic examination, (3) humanistic and professional qualities, and (4) presentation and clinical reasoning skills. The BSE form utilized an item checklist in association with a global assessment score for each category (figure e-1 at Neurology.org). The evaluation was scored out of 40 points (10 points for each major category) using the global assessment scores. In addition, during the last 3 blocks of the second year of the study, a case complexity score from 1 to 3 was added to evaluate the patient complexity with regard to neurologic chief complaint, medical history, and the challenging nature of patient as informant. The case complexity score was determined by the faculty evaluator using precise complexity definitions (figure e-1). This complexity score was used to assess statistical correlation between BSE score and case complexity, but was not incorporated into the student's actual BSE score. Both true and mock final grades were calculated for all students completing the BSE using Z scores as follows for the mock score: Clinical Score = 65%, Shelf = 20%, and BSE score = 15%.

A mock final grade (e.g., honors, high pass, pass) was determined for each student using identical cutpoints as used for the true final grades. The true grades were assigned as described above.

For training purposes, faculty received a comprehensive preceptor guide that consisted of the following elements: (1) a general description of the purpose of the evaluation, (2) detailed descriptions of the procedures to follow, (3) a BSE scoring guide, (4) guidelines to giving effective feedback, and (5) troubleshooting. In addition, training was augmented during an annual faculty development grand rounds presentation on education that included follow-up on BSE data and a presentation on effective techniques for evaluation and feedback.

In terms of conducting the evaluation, students would have minimal or no prior knowledge of the patient, and would not have any specific opportunity to review the chart prior to the start of the evaluation. Students were given 45 minutes to conduct a thorough history and neurologic examination. They were given 10 minutes to briefly present and discuss the case with the faculty evaluator. There was 5 minutes at the end for direct feedback.

Implementation strategy and curricular change. During the first year of the study, the medical students were assigned to complete a BSE during the 4th week of their rotation with their on-service faculty member as evaluator. The evaluation was semi-scheduled as it was purported to be completed during that last week, but no specific time or patient was assigned. As this resulted in a suboptimal completion rate, a curricular change was made during the second year of the study to improve project feasibility. We created an electronic master schedule for the full academic year, pairing faculty members, when they were not on-service, with students with whom they had not worked. Faculty were recruited for these evaluations through e-mail at the start of the academic year, and each faculty was asked to voluntarily sign up for one 2.5-hour slot/year to evaluate 2 students. They were not specifically compensated for this activity.

Data collection and statistical analysis. The raw BSE score as well as true and mock final grades were collected for each student. BSE raw score mean, median, and SD were calculated as well as absolute percentage with a final grade change when comparing mock to true grade. Bivariate correlations were then used to examine the relationship of the BSE score with students' Clinical Score and Shelf. Several one-way analyses of variance were used to test the study hypotheses regarding differences in BSE score across the 3 final grade categories (i.e., honors, high pass, and pass) as well as differences in raw BSE score across the 3 case complexity categories. All analyses were performed using SPSS, version 21 (SPSS Inc., Chicago, IL).

Surveys were given to all students at the end of their Neurology Clerkship except for the first block as this block was dedicated to piloting the most effective survey design. The survey inquired about whether students were observed by a faculty member or resident during a focused or comprehensive neurologic encounter other than during the BSE. A focused encounter was defined as direct acquisition of the history of present illness, a brief review of history, obtaining 1–2 review of systems items, and testing of several components of the neurologic examination with verbal review of the remaining components. A comprehensive encounter was defined as direct acquisition of the history of present illness, full history, and direct testing of a full neurologic examination. Students were also asked whether the BSE was a useful component to their neurologic education and whether it

would be reasonable to count the BSE as a percentage of the overall grade in the future.

RESULTS BSE scores and grade change. A BSE was successfully completed for 84/109 (77.1%) students during the first year of the study and 105/105 (100%) students during the second year of the study (after the curricular change). Overall, 189 of 214 (88.3%) medical students who rotated through the Neurology Clerkship during the 2-year study period completed a BSE. Using a *Z* test to compare the proportion of BSE compliance across the 2 years, we observed a statistically significant difference ($Z = 5.2$, $p < 0.0001$). During the first year of the study, the mean raw BSE score was 78.2 (13.8 SD), the mean Clinical Score was 85.5 (4.7 SD), and the mean Shelf score was 79.3 (7.7 SD). During the second year of the study, the mean raw BSE score was 71.9 (14.1 SD), the mean Clinical Score was 85.3 (4.8 SD), and the mean Shelf score was 78.2 (7.8 SD). Over 2 years, 25 of 189 students (13.2%) had a difference between mock and true grade when accounting for the BSE (e.g., honors vs high pass) with approximately half experiencing a grade elevation and half experiencing a decline.

Correlation of BSE with clinical score, shelf, final grade, and case complexity. The overall correlation of BSE scores with the Clinical Score was 0.36 ($p < 0.001$) and correlation with the Shelf was 0.35 ($p < 0.001$). During the first year of the study, when the same faculty who completed the BSE also contributed to the student's Clinical Evaluation Form, the correlations with Clinical Score and Shelf were 0.55 ($p < 0.001$) and 0.41 ($p < 0.001$), respectively. During the second year of the study, when an independent faculty member conducted the evaluation, the correlations with Clinical Score and Shelf were lower at 0.27 ($p = 0.02$) and 0.29 ($p = 0.003$), respectively.

In analyses examining differences in BSE across the 3 true final grade categories, the results showed that the BSE scores differed significantly across all 3 grades. The mean BSE score for students receiving a final grade of honors, high pass, and pass were 82.6 (13.0 SD), 76.3 (13.9 SD), and 69.5 (13.0 SD), respectively ($F_{2,186} = 14.3$, $p < 0.0001$) (figure 1). Similar observations were noted when stratifying by year (1 and 2).

There were no significant differences in BSE score across the 3 case complexity categories. For a complexity score of 1 (least complex), 2, and 3 (most complex), the mean BSE scores were 73.1 (15.2 SD), 72.3 (13.7 SD), and 71.9 (12.0 SD), respectively ($F_{2,76} = 0.05$, $p = 0.951$) (figure 2).

Student survey data. Of the 186 students who rotated through the Neurology clerkship during the surveyed

Figure 1 Correlation of mean bedside skills evaluation (BSE) score with true final grade

A one-way analysis of variance was used to test for differences in bedside evaluation scores across the 3 final grade categories (1 = honors, 2 = high pass, 3 = pass). Result showed that BSE scores differed significantly across the 3 grades ($F_{2,186} = 14.3$, $p < 0.0001$). Post hoc analyses using the Tukey test revealed significant differences among all 3 grade categories.

blocks, 152 completed the survey (response rate 81.7%). A total of 54.6% of students reported not being directly observed prior to the BSE completing a comprehensive encounter on the wards. Nearly one-quarter of students did not get observed prior to the BSE at all by either a faculty or resident gathering a history or performing a neurologic examination. A total of 93% of students felt that the BSE was useful to their neurologic education. A smaller subset of students were asked whether they felt it was reasonable to count the BSE as a percentage of the overall grade (15%) in future years (54 of 79 students returned surveys; response rate = 68%). A total of 81.5% of students agreed that it was reasonable for the evaluation to be incorporated into the overall grade. Based on survey narrative comments, the most common concern students had regarding the evaluation being formally graded was evaluator variability.

DISCUSSION The main findings of the present study are (1) the proven feasibility of a comprehensive bedside assessment, (2) the significant difference between the BSE score and other forms of student evaluation, (3) the finding that using real patients with differing case complexity did not affect the BSE score, and (4) the quantitative reporting of the percentage of students who did not get supervised at the bedside when not accounting for the BSE.

Figure 2 Correlation between case complexity and mean bedside skills evaluation (BSE) score

A one-way analysis of variance was used to test for differences in bedside evaluation scores across the 3 case complexity categories (1, 2, and 3). Results showed that BSE scores did not differ significantly across the categories ($F_{2,76} = 0.05, p = 0.951$).

We were able to achieve a 100% completion rate with the scheduling strategy utilized during the second year of the study. The curricular change had a major impact on the feasibility of the evaluation as the completion rate increased from 77.1%. This statistically significant improvement in compliance allowed for successful incorporation of the evaluation into the clerkship. After the first year, informal discussions with faculty revealed that completion rate was suboptimal due to the following major factors: (1) pager interruptions, (2) busyness of service, and (3) perceived lack of equity among the different faculty performing the evaluation. Therefore, it was decided to allow faculty to have the opportunity to voluntarily sign up for uninterrupted time of their choosing during the second year in the form of an electronic yearly master schedule. Using this strategy, at our institution we needed a ratio of approximately 1 faculty member for every 2.5 students in order to complete all the evaluations. In addition, faculty directly involved with the clerkship (R. Thompson Stone, R. Jozefowicz) completed a disproportionately greater number of evaluations over the course of the year (approximately 15 and 8, respectively).

The number of students with a difference of mock grade vs true grade (13%) and the relatively low correlations (especially during the second year of the study) indicated that the BSE measured unique

information about student performance. Thus, it contributed to triangulation of student performance, and given the inherent subjectivity of clerkship evaluations, this will serve to improve reliability of the overall assessment.⁷ The fact that the BSE scores differed significantly across the 3 true final grades, however, provides evidence of convergent validity.

We found no statistically significant difference between BSE score and case complexity. This suggests that using real, admitted patients rather than standardized patients did not affect student grade with regards to case complexity. The authors prefer real patients for this type of evaluation given the ability to show real neurologic deficits on examination, and because the experience is less contrived.⁹

During the typical course of their clerkship, and excluding the BSE, student survey data indicate that nearly one-quarter of students were not observed at all performing a neurologic history and examination by a faculty or resident neurologist and over 50% were not observed performing a comprehensive encounter. This greatly limits the grading committee's ability to determine a fair and accurate assessment of student clinical skills performance. In addition, it is a missed opportunity for the student to get important formative feedback. The majority of students felt that the evaluation was useful and agreed with the concept of counting the BSE as part of the overall grade.

There are several limitations of this study and the strategy for student evaluation. The study is a retrospective analysis of collected education data. Given the ethical considerations, it would be difficult to prospectively give students different experiences in an effort to create an intervention and control group. It may be difficult to generalize these results for clerkships at different institutions nationally. The institution would need a similar ratio of faculty to students, and there would have to be a high volume of admitted neurology patients to utilize for the evaluation. In addition, it takes significant time and effort on behalf of the clerkship administration to set up and maintain the evaluation. The use of similar evaluators for the Clinical Score determination and the BSE score during the first year of the study is a limitation with regards to evaluating correlations. However, the correlation data for the second year of the study eliminated that overlap.

Another significant limitation is that the psychometric properties of the BSE instrument, including interrater reliability, were not formally tested. In fact, the most common concern that students expressed was variability (subjectivity) in faculty evaluation. While these concerns are partly mitigated through faculty training to standardize faculty performance, using multiple evaluators for a single student would have increased the reliability of the BSE score. Instead, we utilized a single faculty evaluator in an

effort to improve feasibility and generalizability since multiple evaluators for each student is likely not feasible at the vast majority of teaching hospitals. We felt that although the individual evaluations had an element of subjectivity, the BSE increased the reliability of the overall grade as it helped triangulate student performance through multiple methods and evaluators. However, in a subsequent study it would be worth exploring the interrater reliability of the assessment instrument as well as the feasibility of implementing the BSE at other institutions. In addition, we will consider revising the BSE form to capture student performance with regards to patient and family-centered interviewing, and will add information on this style of interviewing as well as humanism and professionalism to the faculty preceptor guide.

This educational initiative was carried out with the ultimate goal of improving the process of neurologic bedside skills teaching and evaluation for medical students at URSMD through direct observation by faculty. The process was a success and has added much-needed triangulation to the students' overall assessment of competence. The BSE has been added as a formal part of the Neurology Clerkship at URSMD and now counts as 15% of the final score. As the BSE was formative during the study period and is now summative, we are continuing to monitor student feedback and collecting raw data on case complexity and BSE scores for further evaluation in this format. This process of assessment may inform future curricular change nationally that would help bring teaching and evaluation back to the bedside in an attempt to design the most robust educational experience for optimal training of our future physicians.

AUTHOR CONTRIBUTIONS

Robert Thompson Stone: design and conceptualization of the study, analysis and interpretation of the data, drafting and revising the manuscript. Christopher Mooney: analysis and interpretation of the data, aided in drafting the manuscript. Erika Wexler: conceptualization of the study, analysis of the data, aided in drafting the manuscript. Jonathan Mink:

conceptualization of the study. Jennifer Post: conceptualization of the study, analysis of the data. Ralph Jozefowicz: design and conceptualization of the study, aided in drafting the manuscript.

ACKNOWLEDGMENT

The authors thank the leaders of the fellowship, Dr. Ronald Epstein and Dr. Annette Medina-Walpole, for their support and guidance throughout this process.

STUDY FUNDING

The Dean's Teaching Fellowship at University of Rochester sponsored this education initiative. This study was not industry sponsored or funded.

DISCLOSURE

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

Received April 11, 2016. Accepted in final form August 15, 2016.

REFERENCES

1. Engel GL. Editorial: Are medical schools neglecting clinical skills? *JAMA* 1976;236:861–863.
2. Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. *Acad Med* 2004;79:276–280.
3. Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med* 2004;79:16–22.
4. Szenas P. The role of faculty observation in assessing students' clinical skills. *Contemp Issues Med Educ* 1997;1:1–2.
5. Epstein R. Assessment in medical education. *N Engl J Med* 2007;356:387–396.
6. Lukas RV, Adesoye T, Smith S, Blood A, Brorson J. Student assessment by objective structured examination in a neurology clerkship. *Neurology* 2012;79:681–685.
7. 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 [in German]. *Fortschr Neurol Psychiatr* 2009;77:32–37.
8. Schmähmann JD, Neal M, MacMore J. Evaluation of the assessment and grading of medical students on a neurology clerkship. *Neurology* 2008;70:706–712.
9. Hodges B. Validity and the OSCE. *Med Teach* 2003;25:250–254.