Andrew R. Pines, MA*

Mohammed S. Alghoul, MD[‡]

Youssef J. Hamade, MD, MSCI*

Mithun G. Sattur, MD*

Rami James N. Aoun, MD,
MPH*

Tariq K. Halasa, MD*

Chandan Krishna, MD*

Samer G. Zammar, MD, MPH[§]

Najib E. El Tecle, MD, MS[¶]

Tarek Y. El Ahmadi, MD^{||}

Salah G. Aoun, MD^{||}

Richard W. Byrne, MD[#]

James S. Harrop, MD**

Brian T. Ragel, MD^{††}

Daniel K. Resnick, MD^{§§}

Russell R. Lonser, MD^{¶¶}

Nathan R. Selden, MD, PhD^{††}

Bernard R. Bendok, MD, MSCI*

*Department of Neurological Surgery, Mayo Clinic, Phoenix, Arizona; †Department of Plastic and Reconstructive Surgery, Northwestern University Feinberg School of Medicine, Chicago, Illinois; ‡Department of Surgery, Mayo Clinic, Phoenix, Arizona; §Department of Neurological Surgery, St. Louis University Hospital, St. Louis, Missouri; ||Department of Neurological Surgery, University of Texas Southwestern, Dallas, Texas; #Department of Neurological Surgery, Rush University Medical Center, Chicago, Illinois; **Department of Neurological Surgery, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania; ††Department of Neurological Surgery, Oregon Health & Science University, Portland, Oregon; §§Department of Neurological Surgery, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; ¶¶Department of Neurological Surgery, Ohio State University, Columbus, Ohio

Correspondence:

Bernard R. Bendok, MD, MSCI, FAANS, FACS, FAHA,
William and Charles Mayo Professor,
Chair,
Department of Neurological Surgery,
Mayo Clinic,
5777 E. Mayo Blvd,
Phoenix, AZ 85054.
E-mail: Bendok.Bernard@mayo.edu

Received, November 24, 2015.

Accepted, June 16, 2016.

Published Online, August 22, 2016.

Copyright © 2016 by the
Congress of Neurological Surgeons

Assessment of the Interrater Reliability of the Congress of Neurological Surgeons Microanastomosis Assessment Scale

BACKGROUND: The potential for simulation-based learning in neurosurgical training has led the Congress of Neurosurgical Surgeons to develop a series of simulation modules. The Northwestern Objective Microanastomosis Assessment Tool (NOMAT) was created as the corresponding assessment tool for the Congress of Neurosurgical Surgeons Microanastomosis Module. The face and construct validity of the NOMAT have been previously established.

OBJECTIVE: To further validate the NOMAT by determining its interrater reliability (IRR) between raters of varying levels of microsurgical expertise.

METHODS: The NOMAT was used to assess residents' performance in a microanastomosis simulation module in 2 settings: Northwestern University and the Society of Neurological Surgeons 2014 Boot Camp at the University of Indiana. At Northwestern University, participants were scored by 2 experienced microsurgeons. At the University of Indiana, participants were scored by 2 postdoctoral fellows and an experienced microsurgeon. The IRR of NOMAT was estimated by computing the intraclass correlation coefficient using SPSS v22.0 (IBM, Armonk, New York).

RESULTS: A total of 75 residents were assessed. At Northwestern University, 21 residents each performed microanastomosis on 2 model vessels of different sizes, one 3 mm and one 1 mm. At the University of Indiana, 54 residents performed a single microanastomosis procedure on 3-mm vessels. The intraclass correlation coefficient of the total NOMAT scores was 0.88 at Northwestern University and 0.78 at the University of Indiana.

CONCLUSION: This study indicates high IRR for the NOMAT. These results suggest that the use of raters with varying levels of expertise does not compromise the precision or validity of the scale. This allows for a wider adoption of the scale and, hence, a greater potential educational impact.

KEY WORDS: Microanastomosis, Northwestern Objective Microanastomosis Assessment Tool, Objective structured assessment of technical skill, Resident training, Simulation education, Simulation in neurosurgery, Validity

Operative Neurosurgery 13:108–112, 2017

DOI: 10.1227/NEU.0000000000001403

The small margin for error and technical precision required in neurosurgery presents distinct challenges to neuro-

surgical educators in the training of residents. Ericsson et al¹ point to “deliberate practice” as the defining method of skill acquisition for masters in a wide variety of domains. Broadly, deliberate practice in medical education is defined by the repeated performance of a challenging skill (from the student's perspective), a thoughtful assessment of the performance, and feedback on specific areas for improvement.² Technique simulation modules provide all the components needed to engage in deliberate practice and are therefore well suited for training residents in neurosurgery. Using simulators, students can practice until the

ABBREVIATIONS: ICC, intraclass correlation coefficient; IRR, interrater reliability; NOMAT, Northwestern Objective Microanastomosis Assessment Tool

Supplemental digital content: is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.operativeneurosurgery-online.com).

technique is mastered, as judged by an educator using objective metrics.

Simulation modules have been used for training in various neurosurgical procedures, including ventriculostomy catheter placement, pedicle screw fixation, and posterior cervical laminectomy.^{3,4} Several obstacles, however, have prevented the widespread adoption of simulation modules into neurosurgical curricula. These challenges include the complexity of neurosurgical procedures, financial constraints, limits on residents' time, and the lack of validated assessment tools.⁵⁻⁷ Assessment scales are an essential element of an effective simulation-based learning program. When they are rigorously validated, assessment scales can provide quantified feedback that can reliably facilitate an individual's improvement.

The Congress of Neurological Surgeons has created simulation modules, each incorporating a scale derived from the Objective Structured Assessment of Technical Skill.⁸⁻¹⁰ One of these modules is the "Congress of Neurological Surgeons Microanastomosis Module," which has a corresponding assessment scale called the Northwestern Objective Microanastomosis Assessment Tool (NOMAT). A previous study by our group established the face and construct validity of the NOMAT.¹¹ The objective of the current study was to continue the process of validation by assessing the interrater reliability (IRR) of the NOMAT. The IRR is an assessment of a scale's objectivity, or to what degree the results from a scale vary when used by different raters. The ability of nonsurgeons, such as students or technicians, to serve as accurate raters is an important attribute of an assessment scale. This has the potential to make simulation paradigms more accessible to a wider audience, because an experienced surgeon would not necessarily be needed to give feedback. This may allow for widespread adoption of simulation paradigms and likely reduce cost. In addition, there is a potential cost benefit to implementing a scale with high IRR. We therefore aimed to determine (1) the IRR among experienced surgeons and (2) the IRR between an experienced surgeon and preresidency fellows with significant prior training in the module.

METHODS

Materials

One 3-mm silicone-based artificial vessel was provided for each resident in both settings. At Northwestern University, residents were given a 1-mm silicone-based artificial vessel in addition to a 3-mm one. The vessels were cut in half and were held together by a clip (Figure 1). The vessels were then fixed under a surgical microscope (Figure 2). Each resident was provided with 1 pair of scissors, 1 jeweler forceps, 1 needle holder, and one 9-0 Nylon suture. The residents were asked to perform an end-to-end anastomosis of a cut vessel with interrupted sutures.

NOMAT Scale

The NOMAT scale is a Likert-type assessment tool derived from the Objective Structured Assessment of Technical Skill. The items on the NOMAT involve surgical metrics that were selected and defined by 2

FIGURE 1. Close-up picture of cut vessel. Color version available online only.

experienced vascular neurosurgeons and a postdoctoral research fellow.¹¹ The initial NOMAT scale contained 14 items and is provided in the **Supplemental Digital Content** (<https://academic.oup.com/ons/article/2608035/Assessment-of-the-Interrater-Reliability-of-the-supplementary-data>).

Between the 2 settings of the study, the scale was abbreviated with the aim of refining assessment to the most objective measures of performance. The abbreviated scale omits 3 items: "understanding surgical equipment (III)," "microsurgical efficiency with the needle (X)," and "microsurgical efficiency with knot tying (XI)." Table 1 shows an example of the grading sheet that was used.

Settings

Neurosurgery residents participated in the microanastomosis module in 2 educational settings: Northwestern University and the 2014 Society of Neurological Surgeons Boot Camp at the University of Indiana.

Northwestern University

Twenty-one neurosurgery residents participated in the microanastomosis module at Northwestern University. Each resident performed

FIGURE 2. Picture of resident in module. Color version available online only.

TABLE 1. Illustrative Example of a Resident's Grading Sheet (Indiana University)

Item	Postdoc	Postdoc	Faculty
Positioning	4	4	5
Microscope	5	3	5
Instruments	4	3	4
Vessel	3	3	3
Needle	3	3	3
Bite	4	3	4
Suture	4	4	4
Knot	3	4	3
Off	3	4	3
On	4	3	3
Lumen	3	3	2
Total	40	37	39

a microanastomosis procedure on both a 3-mm and a 1-mm model vessel. Individual grades on each vessel counted as distinct variables in statistical analysis. The performance of each resident was recorded by 2 cameras: 1 attached to the microscope and 1 mounted on the wall. Each resident wore a surgical gown for deidentification. Two senior experienced surgeons, a cerebrovascular neurosurgeon and a microvascular reconstructive surgeon, graded the performance of the participants by using the NOMAT scale. These surgeons were both considered "faculty."

Indiana University

Fifty-four first-year neurosurgery residents from 35 US residency programs participated in the microanastomosis simulation module at the Society of Neurological Surgeons Boot Camp in the University of Indiana in 2014. Residents' performance on a 3-mm vessel was graded by using the abbreviated NOMAT scale by the same cerebrovascular surgeon who served as a rater at Northwestern University and 2 postdoctoral research fellows.

Statistical Analysis

The intraclass correlation coefficients (ICCs) were determined in both settings for each item of the NOMAT as well as for the total NOMAT score (maximum 55). ICC values ≥ 0.7 were considered favorable, values 0.5 to 0.69 were considered acceptable, and values < 0.5 were considered poor. Statistical analysis was performed using SPSS (SPSS Statistics for Macintosh, Version 22.0, IBM, Armonk, New York).

RESULTS

Northwestern University

Results of IRR between the 2 faculty members for the Northwestern University setting are shown in [Table 2](#). There was a favorable ICC for the total NOMAT score (0.88; confidence interval = 0.61-0.89). The ICC values for the individual NOMAT items ranged from acceptable to favorable (0.54-0.83). Favorable values were reached for "evaluation of the completed anastomosis-on-pump" (0.83), "knot tying" (0.79), "handling of the surgical instruments" (0.76), "needle bite uniformity" (0.71), and "evaluation of the completed anastomosis-off-pump" (0.70).

TABLE 2. Interrater Reliability Between 2 Faculty Members (Northwestern University)

NOMAT Items	ICC	Lower CI	Upper CI
Total	0.88	0.61	0.89
On	0.83	0.68	0.91
Knot	0.79	0.4	0.83
Instruments	0.76	0.57	0.88
Bite	0.71	0.32	0.8
Off	0.7	0.45	0.84
Microscope	0.68	0.4	0.83
Lumen	0.67	-0.06	0.69
Vessel	0.66	-0.01	0.71
Suture	0.65	0.08	0.73
Needle	0.58	-2.46	0
Positioning	0.54	0.14	0.75

CI, confidence interval; ICC, intraclass correlation coefficient.

TABLE 3. Interrater Reliability Between 2 Postdoctoral Fellows and One Faculty Member (Indiana University)

NOMAT Items	ICC	Lower CI	Upper CI
Total	0.78	0.63	0.86
Knot	0.77	0.64	0.86
Suture	0.75	0.57	0.83
Needle	0.65	0.41	0.77
Bite	0.65	0.41	0.77
Vessel	0.58	0.34	0.74
Off	0.54	0.19	0.68
Lumen	0.51	0.17	0.68
On	0.5	0.11	0.65
Instruments	0.44	0	0.61
Microscope	0.25	-0.2	0.53
Positioning	0.17	-0.46	0.43

CI, confidence interval; ICC, intraclass correlation coefficient.

Acceptable values were reached for "operator positioning and posture" (0.54), "needle handling and care" (0.58), "spacing of the sutures" (0.65), "vessel handling and respect for tissue" (0.66), "evaluation of the completed anastomosis-examination of the lumen" (0.68), and "use of the surgical microscope" (0.68). No poor values were observed.

Indiana University

Results of interrater reliability between the 2 postdoctoral fellows and 1 faculty member are shown in [Table 3](#). There was a favorable ICC for the total NOMAT score (0.78, confidence interval = 0.63-0.86). The ICC values for the individual NOMAT items ranged from poor to favorable (0.17-0.77). Favorable values were reached for "knot tying" (0.77) and "spacing of the sutures" (0.75). Acceptable values were reached for "needle handling and care" (0.65), "needle bite uniformity" (0.65), "vessel handling and respect for tissue" (0.58), "evaluation

of the completed anastomosis–off-pump” (0.54), “evaluation of the completed anastomosis–examination of the lumen” (0.51), and “evaluation of the completed anastomosis–on-pump” (0.50).

Poor values were reached for “operator positioning and posture” (0.17), “use of the surgical microscope” (0.25), and “handling of the surgical instruments” (0.44).

DISCUSSION

The NOMAT scale is an assessment tool that aims to quantify the most important aspects of performing a microanastomosis procedure in a simulation setting. In the current study, we sought to determine the IRR by calculating the ICC in 2 settings, Northwestern University and the Society of Neurological Surgeons Boot Camp at the University of Indiana. We observed high IRR scores for the total NOMAT score between the 2 faculty members at Northwestern University (0.88) and between the 2 postdoctoral fellows and 1 faculty member at Indiana (0.78). Microsurgery is a specialized skill that requires a longstanding dedication in order to master. It is of little surprise, then, that the IRR of the scale was lower when preresidency fellows served as raters than when raters were experienced microsurgeons (0.78 vs 0.88). Nonetheless, the fellows achieved a favorable rating, perhaps owing to their familiarity with the module. This suggests that, given proper training, this scale can be effectively used by individuals who do not necessarily have extensive surgical experience performing microanastomosis. The current study supports claims that relate to the reliability of the assessment of a participant graded by the NOMAT.

Individual NOMAT items in both settings showed consistent ICC values (acceptable to favorable), with the exception of “operator positioning and posture,” “use of the surgical microscope,” and “handling of the surgical instruments.” We believe that these items scored lower because the criteria for their results were more subjective than other items on the scale, and, hence, would improve by using more quantifiable criteria. For example, posture could be more objectively judged if the scale included benchmarks related to the angle of the back, neutral neck position, wrists and elbows resting on the table, and positioning with respect to surgical target. Handling of surgical instruments can be assessed by tracking tool tip motion.¹² Moreover, further rater training on more subjective items might further improve the consistency of this scale. One such method to train raters would be the utilization of videos that provide raters with examples of properly graded vessel models. These multimodal approaches to surgical education would provide health care workers worldwide with an opportunity to efficiently perform a simulated microanastomosis procedure and precisely grade performance.

Limitations

There are some limitations to our study. First, our study used only experienced microsurgeons and postdoctoral fellows as raters. Further studies are needed to assess the legitimacy of employing individuals with a broader range of expertise as raters for the NOMAT. Second, “operator positioning and posture,”

“use of the surgical microscope,” and “handling of the surgical instruments,” the most subjective items on the NOMAT, had lower ICC scores. The NOMAT was abbreviated between the 2 settings, but we believe that this issue can be further improved by suggesting quantifiable markers for these items by which raters can gauge their scoring. Third, while the raters at Northwestern University were blinded to the participant’s identity and postgraduate year level, the raters at Indiana University were not similarly blinded. The lack of blinding of the raters at Indiana University had the potential of introducing bias into the grading process and, in turn, might have slightly affected the IRR. We do not believe this to be a significant limitation, however, because the raters at the University of Indiana did not already know any of the participants, and normally would not find out the participants’ postgraduate year level until after the assessment took place. It could also be argued that grading 2 performances by the same resident separately at Northwestern University (with vessels of 2 sizes) could have affected the results. We do not believe this to be the case, however, because the results were based on rater agreement, not resident performance. Our study was also limited in that it did not assess construct validity, a metric that should be evaluated more than once. This aspect could be evaluated in future courses involving the NOMAT. Finally, contemporary educational theories state that validation of an assessment tool only applies to the setting and population in which that tool was validated.^{13,14} According to this framework, the validity of the NOMAT may not extend beyond assessing simulated microanastomosis for residents.

Future Direction

Ultimately, it is our goal to integrate the microanastomosis training module into a curriculum based on mastery education. Mastery education has been elucidated in previously published articles.¹⁵ Mastery education in medicine is a program defined by 7 steps:

1. Baseline, or diagnostic, testing of skills.
2. Clear learning objectives, sequenced as units in increasing difficulty.
3. Engagement in deliberate practice.
4. A set minimum passing standard for each educational unit.
5. Formative testing with specific feedback to gauge unit completion.
6. Advancement to the next educational unit given measured achievement.
7. Continued practice or study on an educational unit until mastery is achieved.

Compared with simulation-based medical education that lacks the master learning framework, mastery learning in simulation-based medical education has been shown to result in improved skill acquisition of participants and better patient outcomes.¹⁶ Aligning with this vision, the NOMAT can be used longitudinally to reliably assess resident improvement. This

quantification of the resident's skills gives the supervisor an objective perspective on the resident's capabilities. An important next step toward this curriculum is to conduct research on how accurately a participant's NOMAT score on a simulated model predicts performance on patients and clinical outcomes.

CONCLUSION

This study assesses the IRR of the NOMAT as used in the corresponding Microanastomosis Module. We analyzed the IRR of the scale in 2 scenarios, between 2 experienced raters in 1 setting and between research fellows and an experienced rater in the other setting, and found good ICC in both. The demonstration of high IRR implies that the NOMAT scale can be implemented effectively with proctors of varying skill levels. Future studies should focus on further assessing the IRR of the scale among a wider group of proctors and how the IRR can be optimized. In addition, future efforts should focus on how the scale can be used to assess and influence microsurgical skills in clinical practice and, ultimately, clinical outcomes.

Disclosures

Bernard R. Bendok has received research grants from Microvention, the NIH, and the Erika Keeney Foundation. The authors have no personal, financial, or institutional interest in any of the drugs, materials, or devices described in this article.

REFERENCES

1. Ericsson KA, Krampe RT, Tesch-Romer C. The role of deliberate practice in the acquisition of Expert performance. *Psychol Rev*. 1993;100(3):363-406.
2. Duvivier RJ, van Dalen J, Muijtjens AM, et al. The role of deliberate practice in the acquisition of clinical skills. *BMC Med Educ*. 2011;11:101.
3. Harrop J, Lobel DA, Bendok B, Sharan A, Rezai AR. Developing a neurosurgical simulation-based educational curriculum: an overview. *Neurosurgery*. 2013;73 (suppl 1):25-29.

4. Banerjee PP, Luciano CJ, Lemole GM Jr, Charbel FT, Oh MY. Accuracy of ventriculostomy catheter placement using a head- and hand-tracked high-resolution virtual reality simulator with haptic feedback. *J Neurosurg*. 2007;107(3):515-521.
5. El Ahmadi TY, El Tecle NE, Aoun SG, Yip BK, Ganju A, Bendok BR. How can simulation thrive as an educational tool? Just ask the residents. *Neurosurgery*. 2012;71(6):N18-N19.
6. Aoun SG, McClendon J Jr, Ganju A, Batjer HH, Bendok BR. The Association for Surgical Education's roadmap for research on surgical simulation. *World Neurosurg*. 2012;78(1-2):4-5.
7. Zammar SG, El Tecle NE, Bendok BR, et al. Impact of a vascular neurosurgery simulation-based course on cognitive Knowledge and technical skills in European neurosurgical trainees. *World Neurosurg*. 2015;84(2):197-201.
8. Zammar SG, Hamade YJ, Aoun RJ, et al. The cognitive and technical skills impact of the congress of neurological surgeons simulation curriculum on neurosurgical trainees at the 2013 neurological society of India meeting. *World Neurosurg*. 2015; 83(4):419-423.
9. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med*. 1996; 71(12):1363-1365.
10. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skills (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
11. Aoun SG, El Ahmadi TY, Bendok BR, et al. A pilot study to assess the construct and face validity of the Northwestern Objective Microanastomosis Assessment Tool. *J Neurosurg*. 2015;123(1):103-109.
12. McGoldrick RB, Davis CR, Paro J, Hui K, Nguyen D, Lee GK. Motion analysis for microsurgical training: objective measures of dexterity, economy of movement, and ability. *Plast Reconstr Surg*. 2015;136(2):231e-240e.
13. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ*. 2003;37(9):830-837.
14. Cook DA. When I say... validity. *Med Educ*. 2014;48(10):948-949.
15. Cohen ER, McGaghie WC, Wayne DB, Lineberry M, Yudkowsky R, Barsuk JH. Recommendations for reporting mastery education research in medicine (ReMERM). *Acad Med*. 2015;90(11):1509-1514.
16. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Acad Med*. 2013;88(8):1178-1186.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.operativeneurosurgery-online.com).