



AOA Critical Issues in Education

Resident Training and the Assessment of Orthopaedic Surgical Skills

Joshua J. Bagley, MD, Brian Piazza, MD, MSc, Michelle D. Lazarus, PhD, Edward J. Fox, MD, FAOA, and Xiang Zhan, PhD

Investigation performed at Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania

Background: Medical knowledge and technical skills are foundations of surgical competency. The American Board of Orthopaedic Surgery (ABOS) and the Resident Review Committee for Orthopaedic Surgery recently mandated simulation training to improve surgical skills, listing 17 surgical skills modules to improve residents' technical skills. However, there is no established tool to measure the effectiveness of these modules. The Global Index for Technical Skills (GRITS) tool has been previously validated for evaluating general surgery residents. The aim of this study was to determine whether the GRITS tool is valid, practical, and reliable in evaluating the skills of orthopaedic residents in a simulation setting, whether the outcomes correlate to performance in the operating room, and to what extent these simulation modules are valued by residents.

Methods: Simulation performance was assessed longitudinally on 5 residents using the GRITS assessment through postgraduate years (PGY) 1 to 5 (n = 25 evaluations) in a simulated volar forearm approach using cadaveric specimens. An additional 20 PGY-1 residents were evaluated cross-sectionally in this same time frame. Written, open-ended feedback on the simulation experience was sought and analyzed via a thematic analysis. For correlative data, evaluations (n = 65 evaluations) of a variety of authentic surgical procedures were compiled on PGY-2 through PGY-5 orthopaedic residents and compared with the simulated experiences.

Results: GRITS scores were averaged for each group of residents, and validity and reliability were assessed using R-software. PGY-1 residents' mean GRITS evaluation score (expressed as a value from 1 to 5) was 3.4. Longitudinally, this mean score increased over the PGY years 2-5 to 4.4, 4.7, 4.9, and 4.8, respectively. Of the parameters measured by GRITS, the lowest average scores were "flow of operation" and "time and motion" across all levels, although these did improve over PGY years 2 to 5. Findings were consistent between simulation and "real-world" procedures. Open-ended responses evaluating the module were positive.

Conclusions: Our study suggests that the GRITS tool shows promise as an effective and reliable method for assessing orthopaedic resident's technical skills based on an ABOS module system.

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (<http://links.lww.com/JBJSOA/A303>).

Copyright © 2021 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Medical knowledge and technical skills are the foundation of general and specialty surgical competency. Although there are well-established methods to evaluating medical knowledge via written and oral board examinations provided by the American Board of Orthopaedic Surgery (ABOS), the Board relies on residency programs to teach residents the technical skills necessary to become independent, skilled surgeons. Over the course of 5 years, orthopaedic surgery residents actively participate in operating rooms alongside their attending surgeon(s) in an effort to master the essential knowledge and technical skills. The relatively recent changes in the healthcare landscape, such as growing concerns about patient safety, cost burdens, time pressures in busy operating rooms, and work hour restrictions, have decreased these resident education opportunities. As a result, ensuring competency of these necessary healthcare discipline skills is increasingly challenging^{1,2}.

With the added pressure to produce well-trained surgeons despite increasingly limited practice opportunities, the ABOS (in association with the Resident Review Committee for Orthopaedic Surgery) provided a potential solution with the recently mandated introduction of simulation training into orthopaedic residency training programs in an effort to foster improved surgical skill competencies. The ABOS initiated an educational framework that includes 17 surgical skills modules “to improve surgical skills by establishing goals and objectives and assessment metrics, providing training in skills used in the initial management of injured patients and basic operative skills to prepare residents to participate in surgical procedures”³. Although the module topics are outlined, the interpretation for execution and educational approach for undertaking these simulation modules is entirely up to the individual residency programs. Furthermore, recommendations for validated competency assessment approaches of these orthopaedic skill-based modules are not provided. These gaps, including how to run and effectively assess these modules, remain unknown.

Urgency in identifying effective orthopaedic residency skill assessment approaches is amplified in modern residency training programs because of greater focus on achieving competencies over simply relying on exposure or “practice hours”^{4,5}. Various assessment tools have been used over the years to assess surgical residents’ technical skills^{2,6}. These evaluation methods range from formal technical skills examinations in surgical skills laboratories, videotaped case reviews, motion-analysis, questionnaires, and observational assessments by trained surgeons^{2,4,6-9}. These scales, however, are yet to be assessed in the modern simulation modules proposed by the ABOS.

One promising assessment tool, with potential for surgical skill assessment, is the Global Index for Technical Skills (GRITS) scale (Fig. 1). Doyle et al. found this particular tool to be valid and reliable in the assessment of general surgery residents’ technical skills, suggesting that this tool may be applicable to other surgical contexts². It remains, however, unknown whether this tool is applicable for the evaluation of other surgical subspecialties, such as orthopaedics. Thus, the aim of this

study is to determine whether the GRITS tool is a valid, practical, and reliable means of effectively evaluating the skills of orthopaedic surgical residents in an ABOS-recommended simulation setting and if those scores correlate to performance in the authentic operating room. An additional objective of this study was to gauge the usefulness of the simulation module from the residents’ perspectives. We hypothesized that the mean GRITS score would correlate with the postgraduate years (PGY) level and that the resident’s GRITS score would increase with further training. Furthermore, we asked, “what is the impact of these simulations on residents’ perspectives of usefulness and value?”

Methods

Participants

This study used the GRITS evaluation assessment to evaluate orthopaedic residents’ soft-tissue handling in a simulated volar forearm approach using cadaveric specimens. Five residents were evaluated in the Spring annually through postgraduate years (PGY) 1 through 5, providing longitudinal data. An additional 20 PGY-1 residents were evaluated over this time period, providing additional cross-sectional data. All residents were from the same residency program, which accepts 5 residents per year, and no subset was used. Before beginning the laboratory test, participants were provided relevant articles describing the forearm approaches on which they would be evaluated.

ABOS Soft-Tissue Module Simulation

A single orthopaedic attending (E.J.F.) facilitated each session, grading residents (n = 25 residents) on completion by applying the GRITS score assessment. Each resident was graded in 7 categories—respect for tissue, time and motion, instrument handling/knowledge, flow of operation, knowledge of specific procedure, use of assistants, and communication skills—on a scale from 1 to 5 with a maximum total score of 35 points. The final score was then divided by 7 to calculate a mean evaluation score for each resident. After session completion, each participant provided feedback regarding their confidence in their ability to perform the task and the overall impact of the simulation to their residency training.

Authentic Surgical Resident Training

In addition to the simulated cadaver surgery and over the same 5-year period as the simulation model, 65 evaluations were completed using the GRITS scales during a variety of surgical procedures such as lumbar laminectomy and fusion, total knee arthroplasty, closed reduction and percutaneous pinning of pediatric supracondylar humerus fractures, tibial intramedullary nailing, and many others. Unlike the simulated soft-tissue module with a single assessor, various orthopaedic attending surgeons (n = 23) assessed these “real-world” surgical procedures (n = 65 evaluations). Resident participants (n = 22) in this part of the study were PGY-2 through PGY-5, some of whom (n = 5) participated in the simulation module portion of the study.

Respect for Tissue				
1	2	3	4	5
Frequent unnecessary force on tissues or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage	Consistently handled tissue appropriately with minimal damage to tissues	
Time and Motion				
1	2	3	4	5
Many unnecessary moves		Efficient time/motion but some unnecessary moves	Clear economy of movement. Maximum efficiency	
Instrument Handling/Knowledge				
1	2	3	4	5
Tentative/awkward moves or inappropriate use		Competent use of instruments, occasionally awkward	Fluid moves with instruments. No awkwardness	
Flow of Operation				
1	2	3	4	5
Frequently stopped, seemed unsure of next move		Some forward planning, reasonable progression	Obviously planned course, effortless flow	
Knowledge of Specific Procedure				
1	2	3	4	5
Deficient knowledge. Required specific instruction at most steps		Knew all important steps of operation	Demonstrated familiarity with all steps of operation	
Use of Assistants (if applicable)				
1	2	3	4	5
Consistently placed assistants poorly or failed to use		Appropriate use of assistants most of the time	Strategically used assistants to best advantage at all times	
Communication Skills				
1	2	3	4	5
Frequent problems working with team or fails to communicate		Appropriate communication with team most of the time	Co-ordinates surgical team in a superior manner	
Depth Perception (Laparoscopic Procedures Only)				
1	2	3	4	5
Constantly overshoots, swings wide, slow correction		Some overshooting but quick to correct	Accurately directs instruments in correct plane	
Bimanual Dexterity (Laparoscopic Procedures Only)				
1	2	3	4	5
Uses only one hand, poor coordination between hands		Uses both hands but does not optimize their interaction	Expertly uses both hands to provide optimal exposure	

Fig. 1

Global Rating Index for Technical Skills. (Reproduced, with permission, from: Doyle JD, Webber EM, Siddu RS. A universal global rating skill for the evaluation of technical skills in the operating room. *Am J Surg.* 2007;193[5]:551-5.)

Thematic Analysis

Open-ended feedback was analyzed using applied thematic analysis where text was coded into a codebook to identify key themes¹⁰. Coding in NVivo (version 12 plus) was undertaken on simulation participants' responses to the following post-module questions: "Please share your opinion of how this session helped and/or hindered your residency education" and "Please share any input regarding ideas to improve this session and why these changes would help you." Over 1982 words from 25 participants were analyzed using open coding by M.D.L. Themes were reviewed by all coauthors, and discrepancies in data were rectified before the final analysis.

GRITS Statistical Analysis

The GRITS scores were averaged for each group of residents in each of the 7 categories for comparison of different PGY levels. Raw p values were calculated from 2-sample comparisons using the Wilcoxon signed-rank test. To account for multiple testing comparing

all years together, we used the Bonferroni correction to obtain adjusted p values. Validity was assessed using the Pearson correlation coefficient for both the simulated and "real-world" experiences. Reliability was determined using Cronbach's alpha for each setting as well. All statistical analyses were conducted in the R-software¹¹.

Results

Residents were evaluated using the GRITS assessment tool in 2 educational contexts, within a cadaveric volar forearm surgical simulation and within the authentic surgical environment. In the simulation model, most participants were evaluated at a single timepoint (n = 20), with an additional 5 residents evaluated longitudinally across multiple timepoints during the 5-year residency training. In the authentic surgical environment, 65 total evaluations were collected on residents (n = 22 residents) of differing levels of training.

During the PGY-1 year in the simulation model, the residents' mean evaluation score (expressed as a value from 1 to 5)

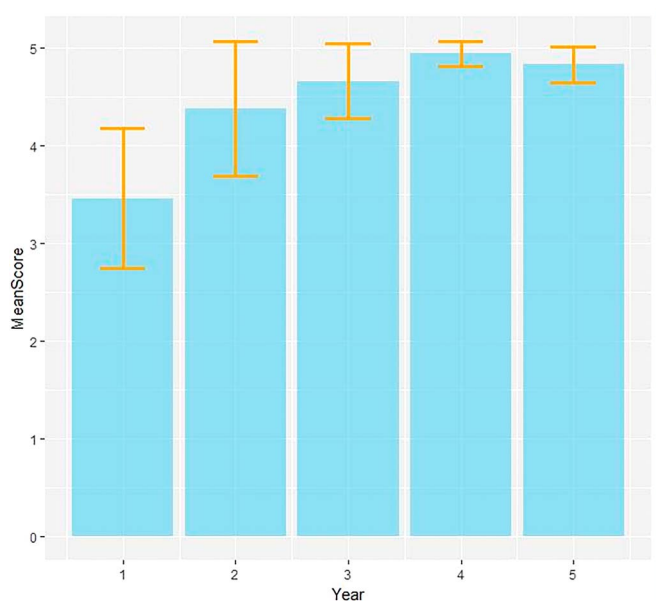


Fig. 2
Mean scores of evaluations by the level of training (postgraduate year) of simulation surgical procedures. Possible scores range from 1 to 5. Error bars represent SD.

was 3.4. Longitudinally, this mean score increased over the PGY-2, PGY-3, PGY-4, and PGY-5 years to 4.4, 4.7, 4.9, and 4.8, respectively (Fig. 2). Raw *p* values comparing each year against one another showed statistical differences only when comparing PGY-1 against all other years. When comparing scores across all years together, scores varied significantly by PGY level (adjusted *p*-value = $4.77e^{-07}$). A Pearson correlation also demonstrated the mean evaluation score and PGY level to be highly correlated (Pearson coefficient = 0.879, *p*-value = 0.0495). Cronbach's alpha (a measure of interitem consistency) was calculated for the 5 residents with 5 simulation procedures (1 each year from 2015 to 2019) and was 0.94.

When applied to “real-world” procedures, the mean evaluation scores also varied significantly by PGY level (Fig. 3, *p*-value = 0.0017), with a Pearson correlation demonstrating a strong correlation between the mean evaluation score and PGY level (Pearson coefficient = 0.951, *p*-value = 0.0487). Cronbach's alpha for the 10 residents with 3 or 3+ real surgical procedures was 0.7. When calculated for the 6 residents with 4 or 4+ real surgical procedures, Cronbach's alpha was 0.76.

In the simulation setting, of the 7 individual categories evaluated by the GRITS tool, “flow of operation” (mean score = 2.9) and “time and motion” (mean score = 2.8) were consistently the lowest scoring categories during the PGY-1 year. Although these parameters continued to improve over PGY years 2 to 5, these parameters remained the lowest average score in the residents' final year (flow of operation = 4.4, time and motion = 4.6).

After completing the simulation, open-ended responses evaluating the module were overwhelmingly positive. Themes identified included the following: “Increased opportunities for practice and the ability to practice in a low-stress environment.”

The PGY-1s, in particular, specified that under the “increased opportunity for practice” theme, the simulations provided the novel opportunity for them to be “lead surgeons” despite being early in their residency training, although—also under this theme—senior residents reportedly valued the “repetition.” When responding to questions related to improving the fresh-tissue simulation module universally, the theme identified was to “increase frequency” of these. PGY-1s also wanted more “feedback checkpoints,” whereas more senior residents wanted the modules to “expand into fractures and fixation.”

Discussion

The need for knowledgeable, skilled, well-trained surgeons is apparent to all¹². Many educational recommendations are made to help residents achieve these competencies, but debates about reliably and validly assessing residents in this training period remain, with particular gaps in the identification of an effective objective evaluation tool of surgical residents' technical skills^{2,4}. Various assessment tools have been created in hopes of filling this gap of assessing surgical residents' technical skills^{2,8,9}. One such tool showing promise among general surgery residents is the GRITS assessment created by Doyle et al.². The purpose of this study was to determine whether the GRITS assessment tool is effective in evaluating orthopaedic residents in a surgical simulation setting and, second, whether the scores translated to the genuine operating room. Our study supports the GRITS as a valid and reliable tool for objectively evaluating an orthopaedic surgery resident's technical skills in both the operating room and a

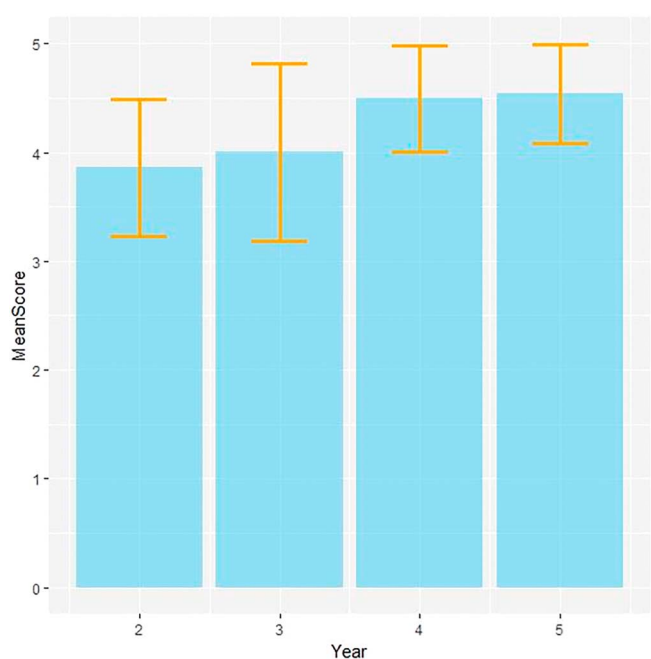


Fig. 3
Mean scores of evaluations by the level of training (postgraduate year) of real surgical procedures. Possible scores range from 1 to 5. Error bars represent SD.

simulation setting. This combination of cross-sectional and longitudinal data allowed for the evaluation of both the simulation module and the impact of these educational activities on residents' skill progression with experience. Equally important, our results illustrate that the assessment tool is effective regardless of the evaluating surgeon because the alignment of scores between the 2 settings was quite high.

The data also revealed that the 2 GRITS categories with the lowest scores were "flow of operation" and "time and motion." This supports the data shown by D'Angelo et al.⁴ in which general surgery chief residents were evaluated while performing 3 simulated surgeries. Of the 6 categories evaluated, "time and motion" once again proved to be among the lowest scoring categories. This is particularly important given that most would agree that the cited skills are improved with time and experience. Our data support this notion and suggest that 1 important factor to improving the GRITS scores is to increase exposure to contexts for residential practice, such as the soft-tissue handling session used here because the 2 previously mentioned categories saw the greatest improvement in the residents evaluated over 5 years. Furthermore, the strong correlation between resident experience, inferred by the PGY level, and improved GRITS scores in both the simulation and real-world settings supports the GRITS as a reliable and valid tool for assessing orthopaedic residents' skills and progression. Given this, perhaps simulation, in association with the GRITS scale, could contribute to the overall evaluation of an orthopaedic resident's readiness to enter independent practice.

An important component of any educational activity is learner perceptions because this affects the uptake and learner engagement¹⁵. Not only was this simulation experience at least as effective for learners as the authentic surgical environment, residents of all year levels valued the increased opportunity for practice in a calm environment. Evidence suggests that these types of "repeated" practice opportunities in supportive environments decrease cognitive load, supporting enhanced learning opportunities¹⁴. The open-ended response feedback also suggests that scaffolding the module targeting the learners' experience level is recommended; for PGY-1s, providing early opportunities to "lead" a surgical approach seems important, but increasing feedback as they progress through this simulation may be important. In later years, as learners progress through competencies, residency programs are encouraged to build upon the module (in this case, adding in fractures and fixation to the fresh tissue module) to further support residents' development outside of the surgical operating rooms. A validated assessment tool such as the GRITS could certainly play a role in tracking progression and guiding both junior and senior residents in improving their technical skills in these repeated settings.

Although this study did show improvement in the overall surgical skills of participating residents as evidenced by the improvement in the mean GRITS score with increased training level, there was no measure of quality of the clinical outcome. Although we certainly hope that improved surgical skills directly translates to improved clinical outcomes, this remains unknown and is a potential weakness of this study. In fact,

Anderson et al. found that improvement in objective structured assessments of residents' surgical skills did not correlate to improvement in surgical results⁷. Certainly, further research directed at correlating a resident's technical skills and the quality of surgical outcome goal would be welcome, although this would be difficult to study outside of a simulation setting.

Another potential weakness of this study is that a single attending surgeon evaluated residents in the simulation setting, raising questions about the applicability of the GRITS and the intergrading reliability of the GRITS. However, when the GRITS was applied to real surgical procedures with numerous evaluating surgeons, the correlation between the mean GRITS score and PGY level remained, suggesting the GRITS as both applicable and reliable regardless of the evaluating surgeon. Still, in the real-world setting, we did not control for evaluating surgeon, so it remains possible that mean scores were overinflated or underinflated if a surgeon was prone to give lower or higher scores. Furthermore, this study was completed at a single institution; therefore, it should be recognized that generalization to other training sites has not yet been established.

Finally, one further weakness of this study is that in the simulation setting, there was very little difference in the mean scores for PGY 3-5 residents, suggesting residents may achieve competency by year 3, raising questions about the validity of the GRITS assessment. When viewing these results, 1 must remember that these scores reflected a volar forearm approach on a cadaveric specimen, a procedure that most PGY-3 and PGY-4 residents should feel comfortable completing. This was reflected by the GRITS score. Furthermore, when the GRITS was applied to the real-world setting with a myriad of differing procedures, the scores more closely reflected a linear progression from postgraduate years 2 to 5. We recommend future studies be focused at applying the GRITS to both simple and complex procedures to establish baseline scores for the procedures. Residents' scores could then be compared with these established norms to determine competency with respect to their level of training.

With increasing responsibilities and limited work hours, ensuring residents are adequately trained and ready to proceed forward into independent practice is a recognized challenge. The ABOS recently mandated surgical simulation to improve surgical skills to meet this challenge; however, without any established tool to monitor progress, evaluating surgical skills objectively remains difficult. Based on the results of this study, our evidence supports that the GRITS tool shows promise as an effective and reliable method for assessing orthopaedic resident's technical skills based on an ABOS module system. ■

Joshua J. Bagley, MD¹
Brian Piazza, MD, MSc²
Michelle D. Lazarus, PhD^{3,4}
Edward J. Fox, MD, FAOA¹
Xiang Zhan, PhD⁵

¹Department of Orthopaedics and Rehabilitation, Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania

²Department of Orthopedics and Sports Medicine, Billings Clinic, Billings, Montana

³Department of Anatomy and Developmental Biology, Centre for Human Anatomy Education, Monash University, Clayton, Victoria, Australia

⁴Monash Centre for Scholarship in Health Education (MCSHE), Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Victoria, Australia

⁵Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania

Email for corresponding author: efox1@pennstatehealth.psu.edu

References

1. Atesok K, Satava RM, Van Heest A, Hogan MV, Pedowitz RA, Fu FH, Sitnikov I, Marsh JL, Hurwitz SR. Retention of skills after simulation-based training in orthopaedic surgery. *J Am Acad Orthop Surg*. 2016;24(8):505-14.
2. Doyle JD, Webber EM, Siddu RS. A universal global rating skill for the evaluation of technical skills in the operating room. *Am J Surg*. 2007;193(5):551-5.
3. Carpenter JE, Hurwitz SR, James MA, Jeffries JT, Marsh JL, Martin DF, Murray PM, Parsons BO, Pedowitz RA, Toolan BC, Van Heest AE, Wongworawat MD. ABOS Surgical Skills Modules for PGY-1 Residents. Available at: <https://www.abos.org/abos-surgical-skills-modules-for-pgy-1-residents.aspx>. Accessed January 21, 2018.
4. D'Angelo AL, Cohen ER, Kwan C, Laufer S, Greenberg C, Greenberg J, Wiegmann D, Pugh CM. Use of decision-based simulations to assess resident readiness for operative independence. *Am J Surg*. 2015;209(1):132-9.
5. Reznick RK. Teaching and testing technical skills. *Am J Surg*. 1993;165(3):358-61.
6. Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. *J Am Acad Orthop Surg*. 2017;25(10):665-72.
7. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. *Clin Orthop Relat Res*. 2016;474(4):874-81.
8. Glarner CE, McDonald RJ, Smith AB, Levenson GE, Peyre S, Pugh CM, Greenberg CC, Greenberg JA, Foley EF. Utilizing a novel tool for the comprehensive assessment of resident operative performance. *J Surg Educ*. 2013;70(6):813-20.
9. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa surgical competency operating room evaluation (O-SCORE): a tool to assess surgical competency. *Acad Med*. 2012;87(10):1401-7.
10. Abedi JH, Taslimi M, Faghihi A, Shekhzadeh M. Thematic analysis and thematic networks: a simple and efficient method for exploring patterns embedded in qualitative data municipalities. *Sterategic Manag Thought*. 2012;5:151-98.
11. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2016. Available at: <https://www.R-project.org/>. Accessed August 16, 2020.
12. U.S. Department of Health and Human Services, Health Resources and Services Administration, National Center for Health Workforce Analysis. National and Regional Projections of Supply and Demand for Surgical Specialty Practitioners: 2013-2025; 2016.
13. Perone JA, Fankhauser GT, Adhikari D, Mehta HB, Woods MB, Tyler DS, Brown KM. It depends on your perspective: resident satisfaction with operative experience. *Am J Surg*. 2017;213(2):253-9.
14. Labbé M, Young M, Mascarella M, Husein M, Doyle PC, Nguyen LHP. How consistent is competent? Examining variance in psychomotor skills assessment. *Acad Med*. 2020;95(5):771-6.