Reliability, Validity, and Feasibility of the Zwisch Scale for the Assessment of Intraoperative Performance

Brian C. George, MD,[*] Ezra N. Teitelbaum, MD,[†] Shari L. Meyerson, MD,[†] Mary C. Schuller, MSEd,[†] Debra A. DaRosa, PhD,[†] Emil R. Petrusa, PhD,[*] Lucia C. Petito, MA,[‡] and Jonathan P. Fryer, MD[†]

[*]Department of Surgery, Massachusetts General Hospital, Harvard University, Boston, Massachusetts; [†]Department of Surgery, Feinberg School of Medicine, Northwestern University, Chicago, Illinois; and [‡]School of Public Health, University of California, Berkeley, California

PURPOSE: The existing methods for evaluating resident operative performance interrupt the workflow of the attending physician, are resource intensive, and are often completed well after the end of the procedure in question. These limitations lead to low faculty compliance and potential significant recall bias. In this study, we deployed a smartphone-based system, the Procedural Autonomy and Supervisions System, to facilitate assessment of resident performance according to the Zwisch scale with minimal workflow disruption. We aimed to demonstrate that this is a reliable, valid, and feasible method of measuring resident operative autonomy.

METHODS: Before implementation, general surgery residents and faculty underwent frame-of-reference training to the Zwisch scale. Immediately after any operation in which a resident participated, the system automatically sent a text message prompting the attending physician to rate the resident's level of operative autonomy according to the 4-level Zwisch scale. Of these procedures, 8 were videotaped and independently rated by 2 additional surgeons. The Zwisch ratings of the 3 raters were compared using an intraclass correlation coefficient. Videotaped procedures were also scored using 2 alternative operating room (OR) performance assessment instruments (Operative Performance Rating System and Ottawa Surgical Competency OR Evaluation), against which the item correlations were calculated.

RESULTS: Between December 2012 and June 2013, 27 faculty used the smartphone system to complete 1490 operative performance assessments on 31 residents. During this period, faculty completed evaluations for 92% of all operations performed with general surgery residents. The Zwisch scores were shown to correlate with postgraduate year (PGY) levels based on sequential pairwise chi-squared tests: PGY 1 vs PGY 2 ($\chi^2 = 106.9$, df = 3, $p < 0.001$); PGY 2 vs PGY 3 ($\chi^2 = 22.2$, df = 3, $p < 0.001$); and PGY 3 vs PGY 4 ($\chi^2 = 56.4$, df = 3, $p < 0.001$). Comparison of PGY 4 to PGY 5 scores were not significantly different ($\chi^2 = 4.5$, df = 3, $p = 0.21$). For the 8 operations reviewed for interrater reliability, the intraclass correlation coefficient was 0.90 (95% CI: 0.72-0.98, $p < 0.01$). Correlation of Procedural Autonomy and Supervisions System ratings with both Operative Performance Rating System items (each $r > 0.90$, all p's < 0.01) and Ottawa Surgical Competency OR Evaluation items (each $r > 0.86$, all p's < 0.01) was high.

CONCLUSIONS: The Zwisch scale can be used to make reliable and valid measurements of faculty guidance and resident autonomy. Our data also suggest that Zwisch ratings may be used to infer resident operative performance. Deployed on an automated smartphone-based system, it can be used to feasibly record evaluations for most operations performed by residents. This information can be used to council individual residents, modify programmatic curricula, and potentially inform national training guidelines. (J Surg 71:e90-e96. © 2014 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: graduate medical education, surgery, educational measurement, evaluation, surgical education

COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement

This work has been supported by funding from the State of Illinois via the Excellence in Academic Medicine Program (Grant #238), the Department of Surgery at Northwestern University's Feinberg School of Medicine, and the Department of Surgery at Massachusetts General Hospital. We declare no financial relationships with any organizations that might have an interest in the submitted work and no other relationships or activities that could appear to have influenced the submitted work. Ethical and institutional board approval was granted before conducting this study. Information on the data used in this report is available on request.

Correspondence: Inquiries to Brian C. George MD, Department of Surgery, Massachusetts General Hospital, 55 Fruit Street GRB 425, Boston, MA 02114; fax: +(617) 724-3499; e-mail: bcgeorge@partners.org

  1931-7204/$30.00 http://dx.doi.org/10.1016/j.jsurg.2014.06.018

OBJECTIVES

The central goal of surgical training must be to graduate residents who are competent to operate independently. Unfortunately, surgical fellowship directors report that 66% of residents cannot operate unsupervised for 30 minutes of a major procedure.[1] This suggests that graduating residents are ill prepared for independent practice. This conclusion is supported by the opinion of many surgical residents themselves.[2] Many investigators are working to understand the factors contributing to this problem, but it is already clear that residents must be provided more opportunities to gain progressive autonomy in preparation for their first day as an unsupervised attending physician.[3,4]

Although progressive resident autonomy is an educational imperative, we must grant autonomy in a way that is safe. Indeed, 9% of surgical technical errors are linked to poorly supervised residents.[5] However, patient safety is but one factor that influences the amount of autonomy provided to residents in the operating room (OR). Increased productivity pressures,[6] evolving ethical considerations, and increased concern regarding malpractice liabilities have the potential to incentivize faculty to limit resident operative autonomy.

One solution is to transparently and deliberately grant progressive autonomy only to those residents who have demonstrated operative competence at a lower level of responsibility. In this type of system, it is imperative that resident autonomy and competence is accurately documented. Valid and reliable quantitative resident performance data permit faculty surgeons to individualize the amount of autonomy granted to a resident for any given procedure and patient. However, collecting these data requires continuous evaluation of resident intraoperative performance. Although widely desired,[4,7] surgical educators lack a universal method for feasibly collecting, analyzing, and distributing continuous resident performance data. Existing methods for evaluating resident operative performance interrupt the attending physician's workflow, are resource intensive, and are often completed well after the end of the procedure in question. These limitations lead to low faculty compliance and potential for significant recall bias.

In this study, we aim to demonstrate that the Zwisch scale can be used to make valid and reliable measurements of both faculty operative guidance as well as resident intraoperative performance. Furthermore, we describe the feasibility of continuous evaluation by deploying the Zwisch instrument on a novel smartphone-based evaluation platform.

PARTICIPANTS AND SETTING

All categorical and undesignated preliminary general surgery residents in a single department of surgery participated ($n = 31$) in this study. All attending faculty in the same department of surgery were eligible for this study and were invited to attend a rater training session. The training session used a frame-of-reference methodology and has been previously described.[8] Between October 2012 and January 2013, a total of 27 faculty raters completed the training and were subsequently enrolled in the study. This study has institutional review board approval from Northwestern University, and all resident and faculty data were deidentified before analysis. All participants signed an institutional review board–approved consent form.

DESIGN

Rating Scales

The raters were asked to evaluate the amount of guidance they provided to the resident using the "Zwisch" scale, previously described in detail.[9] Since the original theoretical description was published in 2013 the names attached to each level have been revised but it otherwise remains unchanged. Briefly, this 1-dimensional behaviorally anchored ordinal scale is used by raters to grade the degree of guidance the attending surgeon provides to the trainee during most of the critical portion of the procedure. At the lowest end of this 4-level scale, the attending physician performs the critical portion while explaining each step to the resident (termed "Show and Tell"). In the next level ("Active Help"), the attending physician actively guides the resident through the critical portion of the procedure. This is in contrast to the third level ("Passive Help"), where the resident performs critical portions of the operation independently while the attending physician passively provides skilled assistance and intervenes only when necessary to make an important teaching point or to optimize patient safety. At the most advanced level ("Supervision Only"), attending physician presence is necessary only to guarantee patient safety. At this level, the resident has enough proficiency to perform the procedure independently using a less skilled assistant, while the attending surgeon does not need to be directly involved in the procedure other than to provide close supervision. These levels were coded 1 to 4, with 4 representing the most advanced level.

In addition to the Zwisch scale, raters were also asked to rank the complexity of the procedure for which they were completing an evaluation. These ratings were made relative to the surgeon's overall experience with that same procedure. We used a 3-level scale, anchored with prompts of "Easiest 1/3," "Average," and "Hardest 1/3."

Lastly, for a subset of the data modified versions of the previously described OPRS[10] and O-SCORE[11] resident performance rating instruments were used to assess resident performance. The OPRS instrument was modified to exclude items which only pertain to the key steps of specific procedures. The O-SCORE instrument was modified to