Effect of Ongoing Assessment of Resident Operative Autonomy on the Operating Room Environment

Jonathan P. Fryer, MD, MHPE,* Ezra N. Teitelbaum, MD,* Brian C. George, MD,† Mary C. Schuller, MS,* Shari L. Meyerson, MD,* Christina M. Theodorou, BA,* Joseph Kang, PhD,* Amy Yang, MS,‡ Lihui Zhao, PhD,‡ and Debra A. DaRosa, PhD*

*Department of Surgery, Northwestern University, Chicago, Illinois; †Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts; and ‡Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

OBJECTIVE: We have previously demonstrated the feasibility and validity of a smartphone-based system called Procedural Autonomy and Supervision System (PASS), which uses the Zwisch autonomy scale to facilitate assessment of the operative performances of surgical residents and promote progressive autonomy. To determine whether the use of PASS in a general surgery residency program is associated with any negative consequences, we tested the null hypothesis that PASS implementation at our institution would not negatively affect resident or faculty satisfaction in the operating room (OR) nor increase mean OR times for cases performed together by residents and faculty.

METHODS: Mean OR times were obtained from the electronic medical record at Northwestern Memorial Hospital for the 20 procedures most commonly performed by faculty members with residents before and after PASS implementation. OR times were compared via two-sample *t*-test. The OR Educational Environment Measure tool was used to assess OR satisfaction with all clinically active general surgery residents ($n = 31$) and full-time general surgery faculty members ($n = 27$) before and after PASS implementation. Results were compared using the Mann-Whitney rank sum test.

RESULTS: A significant prolongation in mean OR time between control and study period was found for only 1 of the 20 operative procedures performed at least 20 times by participating faculty members with residents. Based on the overall survey score, no significant differences were found between resident and faculty responses to the OR Educational Environment Measure survey before and after PASS implementation. When individual survey items were compared, while no differences were found with resident responses, differences were noted with faculty responses for 7 of the 35 items addressed although after Bonferroni correction none of these differences remained significant.

CONCLUSIONS: Our data suggest that PASS does not increase mean OR times for the most commonly performed procedures. Resident OR satisfaction did not significantly change during PASS implementation, whereas some changes in faculty satisfaction were noted suggesting that PASS implementation may have had some negative effect with them. Although the effect on faculty satisfaction clearly requires further investigation, our findings support that use of an autonomy-based OR performance assessment system such as PASS does not appear to have a major negative influence on OR times nor OR satisfaction. (J Surg Ed 75:333-343. © 2016 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: graduate medical education, surgery, evaluation, autonomy

COMPETENCIES: Interpersonal and Communication Skills, Practice-Based Learning and Improvement

INTRODUCTION

There has been concern that general surgical residency programs are not adequately preparing their residents to operate as independent surgeons on graduation.[1,2] Progression to operative autonomy with core procedures is arguably the single most important goal of surgical residency training. Factors that may impede the development of resident autonomy in the operating room (OR) include faculty

Correspondence: Inquiries to Jonathan P. Fryer, MD, MHPE, Department of Surgery, Northwestern University, Arkes Pavilion, Suite 1900, 676 N, St. Clair St, Chicago, IL 60611-2923; e-mail: jfryer@nmh.org

  1931-7204/$30.00 https://doi.org/10.1016/j.jsurg.2016.11.018

concerns about maintaining excellent surgical outcomes while addressing administrative pressures for high procedural volumes.[3-5] Currently, operative competence is determined by self-logged procedural numbers and aggregate faculty assessments based on recall. A real-time assessment system for resident operative performance would allow more reliable data to assess and guide the resident's progression toward operative competency. To be successful, this system must both provide valuable assessment data without requiring excessive time investment from the busy surgical faculty. We have previously demonstrated the validity of the Zwisch scale and the feasibility of using it as part of the Procedural Autonomy and Supervision System (PASS) for the assessment of intraoperative performance, deployed via smartphones directly after the finish of an operation.[6]

In the planning phases of PASS and before its implementation, feedback was sought from both faculty and residents at our institution and others. In this process, concerns were expressed by our faculty members and others that creating an operative performance assessment system that had resident autonomy as its primary focus may have unanticipated negative consequences including prolonging OR times and increasing tension in the OR between faculty and residents. Concerns were primarily based on the possibility that PASS would induce behavioral changes in the OR with both residents and faculty members that would lead to increased resident participation in more difficult parts of the operations that would increase faculty angst and prolong case duration. Therefore, this study sought to determine if, for cases performed involving both faculty and residents, implementation of PASS would significantly influence: (1) mean OR times or (2) faculty and resident satisfaction with the educational experience in the OR or both. We hypothesized that PASS implementation would not negatively affect OR times, nor faculty or resident OR satisfaction.

MATERIALS AND METHODS

Participants and Setting

All preliminary (PG1, $n = 6$) and categorical (PGY1, $n = 5$; PGY2, $n = 5$; PGY3, $n = 5$; PGY4, $n = 5$; PGY5, $n = 5$) general surgery residents who were clinically active at our primary teaching hospital, Northwestern Memorial Hospital, during the study period participated in this study ($n = 31$). All full-time general surgery faculty based at our primary teaching hospital who operate with our preliminary or categorical or both general surgery residents and agreed to use PASS participated ($n = 27$). This included faculty surgeons from several general surgery subspecialties (MIS, Trauma, Surg Onc, Endocrine, Colorectal, Vascular, Transplant, Thoracic, and private practice general surgeons) and with a broad range of experience. Overall, we were able to generate PASS assessments for 92% of the OR procedures performed between our general surgery residents (prelim or categorical) and our participating general surgery faculty members (as defined earlier) during the study time period. With this study, we did not breakdown participation by each individual resident or faculty member. Although, we fully intend to explore variability in assessment scores based on faculty variables (i.e., age, sex, subspecialty, years of experience, etc.) in future studies that was beyond the scope of the current study. All faculty members and residents participating in the study underwent frame-of-reference training to ensure the Zwisch scale was used consistently and the smartphone-based PASS system was used correctly. The study period was defined as December 1, 2012 through June 30, 2013. All data pertaining to individual resident performance were kept confidential and accessible only to the principal investigators and study coordinators.

Design

The Zwisch scale of progressive autonomy is a 4-level scale of operative autonomy.[7] The first level is "Show and Tell," in which the faculty general surgeon performs critical portions of the operation while explaining the operation and thought process to the surgical resident, who is essentially observing and assisting, but may participate on noncritical portions. The second level is "Active Help," in which the resident performs some portions of the operation, with active guidance from the faculty, who may take over as needed for difficult portions of the operation. The third level is "Passive Help," in which the resident is performing most critical portions of the operation with the faculty assisting but providing guidance only when requested or required for patient safety. The final level is "Supervision Only," in which the resident is the primary surgeon who safely performs most critical portions of the procedure with help from junior resident or OR staff as first assistant. The faculty silently supervises, but does not assist surgically nor provide significant guidance unless requested by the resident or required for patient safety (Fig. 1).

With PASS, on completion of the OR procedure, a text message is automatically sent to the faculty member's smartphone identifying the resident's name, procedure title, date, and time. After acknowledging their participation in the case, the faculty member is first prompted to rank the resident as 1 of the 4 levels discussed earlier on the Zwisch scale. Next, the faculty member must define the difficulty level of the procedure when compared with their previous experiences with the very same procedure or with other procedures of a similar nature. For this, they must define the case as ranking among the "Easiest 1/3," "Middle 1/3," or "Most difficult 1/3" of procedures of this type they have ever performed (Fig. 2).

Attending Name: Resident Name:
Date: Operation:

Procedural Difficulty: Compared to my entire previous experience with this procedure or, *if my experience with this procedure is limited*, compared to my overall procedural experience: In terms of difficulty/complexity, I would rank this procedure as among the:

Easiest 1/3 ☐, middle 1/3☐, most difficult 1/3 ☐

Based on his/her performance in this operation, I view the resident as being in the:

☐ **SHOW AND TELL STAGE**: Resident is in process of learning basics of operation (i.e. surgical anatomy, how to handle instruments and use equipment, steps of the operation, proper tissue planes, how to best position his/her body, etc.). Resident essentially observes and assists but may participate in some parts of the procedure such as opening and/or closing the incision. Attending "shows" resident how procedure is done and "tells" resident what needs to be known.

☐ **ACTIVE HELP STAGE**: Resident begins to assume the "surgeon" role in some parts of the operation with attending inserting him/herself as needed including swapping "surgeon" and "assistant" roles with the resident periodically. For parts of the case where resident assumes the surgeon role, the attending "actively" assists, essentially guiding the resident through the operation (i.e. smart help).

☐ **PASSIVE HELP STAGE**: Although some floundering, the resident is capable of safely doing *significant parts* of the procedure without active guidance/intervention from attending as long as the attending is providing passive assistance. However, the resident struggles with parts of the procedure when the attending (or an equally skilled assistant) is not the first assistant.

☐ **SUPERVISION ONLY STAGE**: The resident safely and fluidly performs the operation and can effectively incorporate OR staff or a junior resident as a first assistant. The attending does not need to scrub in and is present in the OR to provide supervision and consultation if needed

FIGURE 1. The Zwisch scale with descriptors for each level.

Data Collection

OR Times

We obtained mean OR times for procedures performed commonly on general surgical services participating in the study from the electronic medical record at Northwestern Memorial Hospital. We included all procedures performed at least 20 times by participating general surgery faculty with general surgery residents during both the study period (i.e., December 1, 2012 through June 30, 2013) and the control period, which was identical to the 7-month time period from the year preceding PASS implementation (i.e., December 1, 2011-June 30, 2012).

OR Satisfaction

We used the OR Educational Environment Measure (OREEM) scale[8] to determine OR satisfaction. The OREEM scale is based on series of questions regarding the OR environment that each require responses based on a graded Likert scale (i.e., strongly agree = 4, agree = 3, unsure = 2, disagree = 1, and strongly disagree = 0). To obtain baseline OR satisfaction data, we independently surveyed both faculty members and categorical general surgery residents before PASS implementation using modified versions of OREEM (faculty version 35 items, resident version 36 items). Modifications to the resident survey were minimal. For the faculty survey, the same questions were asked but, as the survey was originally developed for residents, the wording had to be slightly modified for several items to direct them to faculty members. At the completion of the study period, the OREEM survey was repeated with the same population of faculty members and residents.

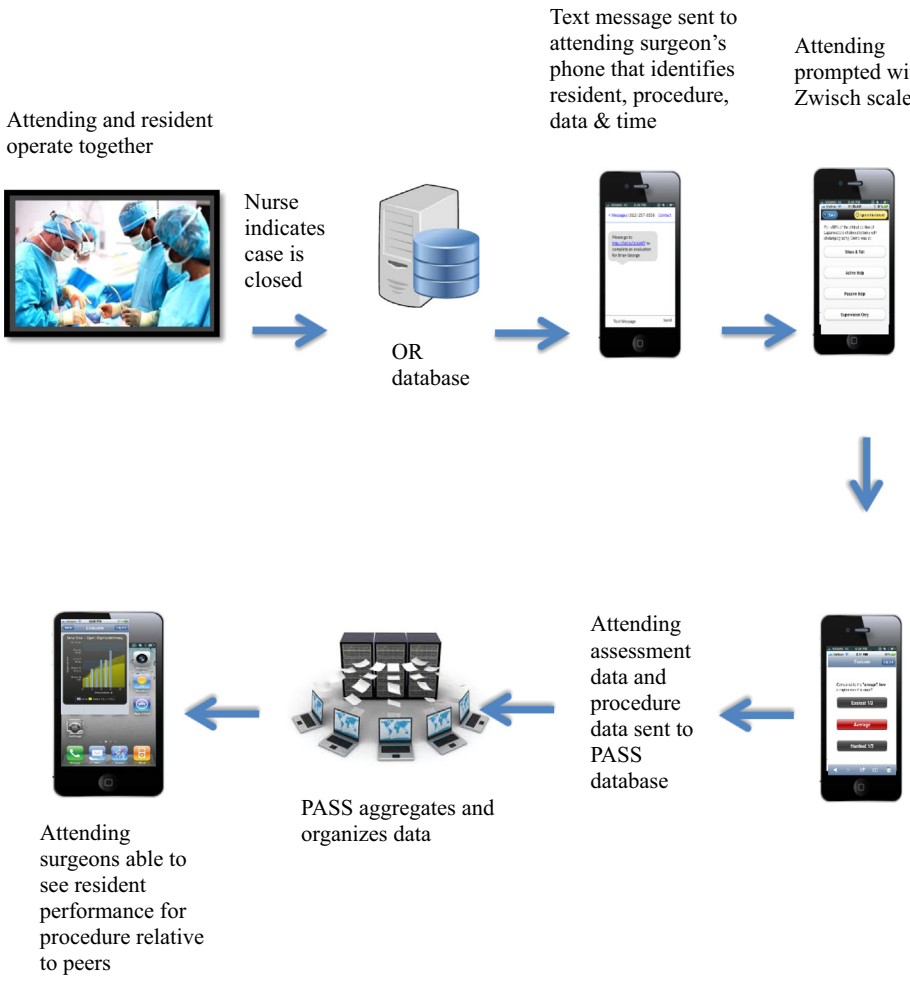

FIGURE 2. Workflow diagram for the procedural autonomy and supervision system (i.e., PASS).

Statistical Analysis

OR Times

Differences in OR times during the control time period and during the study period were compared and p-values were determined using the independent two-sample *t*-test. We performed a Bonferroni correction[9,10] to adjust for multiple comparisons. The new type I error rate is set to be at 0.05/20 = 0.0025. $p < 0.0025$ would be considered to be statistically significant.

OR Satisfaction

Results from the modified OREEM tool during the control time period and after the study period were compared for both faculty and residents using a nonparametric Wilcoxon rank sum test. Similarly, we performed a Bonferroni correction to adjust for multiple comparisons. The new type I error rate is set to be at 0.05/35 = 0.0014. $p < 0.0014$ would be considered to be statistically significant.

RESULTS

OR Times

OR times were compared for 20 procedures that were performed at least 20 times by participating general surgery faculty members with general surgical residents during both the control time period and the study time period (Table 1). Procedures were also separated based on whether they were performed in our main hospital ORs (Feinberg) or our ambulatory surgery facility ORs (Olson) because of previously demonstrated differences in OR times between these 2 facilities for similar procedures. Looking at specific individual procedures, only 2 procedures of 20 showed significantly different OR times between control and study periods at $p < 0.05$ level. Of these 2, only laparoscopic left hemicolectomies performed in our main hospital ORs (Feinberg) had longer OR times during the study period ($p = 0.034$), whereas open inguinal hernia repairs ($p = 0.023$) performed in our ambulatory surgery facility (Olson) had shorter OR times during the study period. With Bonferroni correction, no significant differences were

TABLE 1. Mean OR Times During Baseline Period and Study Period

Primary Procedure	Baseline Period 12/1/11–6/30/12 # of Cases	Baseline Period 12/1/11–6/30/12 OR Times (Min) Mean ± SD	Study Period 12/1/12–6/30/13 # of Cases	Study Period 12/1/12–6/30/13 OR Times (Min) Mean ± SD	p*
Appendectomy laparoscopic	135	48 ± 27	170	52 ± 32	0.24
Cholecystectomy laparoscopic (general)	176	85 ± 43	199	84 ± 43	0.82
Cholecystectomy laparoscopic (Olson)	111	65 ± 28	106	62 ± 23	0.39
Closure ileostomy loop	35	89 ± 63	32	100 ± 58	0.46
Excision chest melanoma	25	64 ± 42	28	72 ± 43	0.50
Gastrectomy sleeve laparoscopic	56	96 ± 22	70	98 ± 47	0.75
Hemicolectomy left laparoscopic	32	168 ± 70	40	205 ± 75	**0.03**
Hemicolectomy left open	48	193 ± 106	25	168 ± 90	0.29
Hemicolectomy right laparoscopic	43	139 ± 83	41	132 ± 42	0.63
Hemicolectomy right open	57	154 ± 113	22	133 ± 57	0.28
Hemorrhoidectomy (general)	42	28 ± 22	59	28 ± 23	1
Hemorrhoidectomy (Olson)	55	21 ± 14	92	24 ± 18	0.26
Herniorrhaphy inguinal open (general)	65	74 ± 34	96	78 ± 45	0.52
Herniorrhaphy inguinal open (Olson)	127	70 ± 22	104	63 ± 24	**0.02**
Herniorrhaphy umbilical (general)	21	51 ± 25	27	66 ± 40	0.12
Herniorrhaphy umbilical (Olson)	39	45 ± 22	40	48 ± 28	0.60
Nissen fundoplication laparoscopic	40	136 ± 66	37	132 ± 41	0.75
Parathyroidectomy	23	113 ± 48	31	107 ± 58	0.68
Thyroidectomy	58	198 ± 98	48	188 ± 89	0.58
Whipple open	35	252 ± 133	20	298 ± 126	0.21

*The reported p values were non-Bonferroni-corrected. The bolded p values reflect significant differences based on $p < 0.05$. With Bonferroni correction, the level of significance was adjusted from 0.05 to 0.05/20 = 0.0025, and no significant differences were found between OR times before and during PASS implementation.

noted in OR times between control and study periods for any of the 20 procedures evaluated (i.e., all $p > 0.0025$).

OR Satisfaction

Faculty members and resident compliance with completing the OREEM survey was high both before (faculty members = 81%; residents = 96%) and after (faculty members = 96%; resident = 80%) PASS implementation. Total scores were calculated by summing up all survey item responses obtained before and after PASS implementation, no significant differences in the total scores were noted with either residents ($p = 0.91$) or faculty members ($p = 0.69$) (Table 3). With comparison of individual item responses before and after PASS implementation, no significant differences were noted with resident responses to any individual item (Table 3). With faculty members, there were 7 of 35 item responses showed significant difference at $p < 0.05$ level (Table 2), with all of these differences reflecting a more negative perspective. These items included Likert scale responses to the following items: (1) my resident and I got along well; (2) I expected my resident's surgical skills would be better than they actually were; (3) the atmosphere in the OR was pleasant; (4) the level of supervision I provided in the OR was adequate for my resident's level; and (5) the emergency case variety gave my resident(s) appropriate exposure to emergency surgery. With Bonferroni correction, no significant differences were identified among any individual survey item responses obtained before and after PASS implementation for either residents or faculty (i.e., all $p > 0.0014$).

DISCUSSION

General surgery residency programs are entrusted with the formidable task of producing graduating surgical residents who can perform a wide variety of operative procedures with competence and autonomy.[1,11] However, obtainment of surgical autonomy is by necessity a graduated process, with surgical residents going through different stages of progressive autonomy during their training.[2,6,7] Complicating this, there is currently no standardized method for assessing and monitoring residents' operative performance.[12,13] Most programs formally evaluate an individual resident's operative performance periodically based on the surgical faculty's collective recall of their operative experiences with that resident. Therefore, these operative performance assessments are typically conducted remotely from most OR experiences that the evaluating faculty members have had with the resident being assessed. As most surgical faculty evaluators perform large numbers of operations with several different residents in the time periods between evaluation sessions, memory decay can be a significant problem with this approach and the validity of remote assessments is highly questionable.[14] Although some programs try to assess

TABLE 2. Comparison of Faculty OREEM Survey for 2012 (Baseline Period) and 2013 (Study Period)

	2012 M(m)IQR	2013 M(m)IQR	p*
My resident(s) and I got along well	3.8 (4) 4–4	3.5 (3) 3–4	**0.01**
My resident(s) was enthusiastic about learning	3.3 (3) 3–4	3.2 (3) 3–4	0.54
I had a genuine interest in my resident(s)' progress	3.6 (4) 3–4	3.4 (3) 3–4	0.17
My resident(s) understood what I was trying to teach him or her.	3.4 (3) 3–4	3.1 (3) 3–4	0.14
I gave my resident(s) time to practice their surgical skills in the OR.	2.8 (3) 2–4	2.9 (3) 3–4	0.74
When the resident(s) did not perform well, I took over the procedure.	2.8 (3) 2.2–4	3.0 (3) 3–4	0.64
Before the OR, my resident(s) and I discussed the planned procedure.	2.6 (3) 1–3	2.5 (3) 2–3	0.43
Before the OR, my resident(s) and I discussed what part of the procedure they would do.	1.6 (1) 1–3	1.8 (2) 1–2.5	0.25
I expected my resident(s)' surgical skills would be better than they actually were.	1.3 (1) 1–2	2.0 (2) 1–3	**0.03**
I gave my resident(s) feedback on their performance.	3.0 (3) 3–3	2.8 (3) 2.5–3	0.45
My feedback was constructive.	2.9 (3) 2–3	3.0 (3) 3–3	0.69
The atmosphere in the OR was pleasant.	3.8 (4) 4–4	3.4 (3) 3–4	**0.01**
My resident(s) was too slow for me to let them operate.	0.9 (1) 0–1	1.5 (1) 1–2	0.08
In the OR, I had to correct my resident(s) often	1.9 (2) 1–3	1.6 (1) 1–3	0.42
Anesthesia pressured me to operate myself to reduce OR time.	0.5 (0) 0–1	0.7 (1) 0–1	0.11
The staff in the OR was friendly to my resident(s).	3.5 (4) 3–4	3.2 (3) 3–4	**0.03**
My resident(s) felt like part of a team in the OR.	3.6 (4) 3–4	3.2 (3) 3–4	**0.02**
My resident(s) was so stressed in the OR that they did not learn as much as they could have.	0.8 (1) 0–1	1.1 (1) 1–1	0.08
When my resident(s) was in the OR there was nobody to cover the service.	1.4 (1) 1–2	1.4 (1) 1–2	0.87
The level of supervision I provided in the OR was adequate for my resident(s)' level.	3.5 (4) 3–4	3.1 (3) 3–3	**0.02**
Operations done on my rotation were generally too complex for my assigned resident(s).	1.7 (1) 1–3	1.7 (1) 1–3	0.89
The elective OR case list had an appropriate mix of cases to suit my resident(s).	3.0 (3) 3–3	2.6 (3) 2–3	0.13
My resident(s) had plenty of opportunity to assist.	3.5 (3) 3–4	3.1 (3) 3–4	0.10
There were enough OR days/week for my resident(s) to get the appropriate experience.	3.5 (4) 3–4	3.0 (3) 3–4	0.10
Senior residents or fellows took away my resident(s)' opportunities to operate.	1.1 (1) 0.2–1	1.1 (1) 0–1	0.98
The # of emergency OR cases provided my resident(s) with adequate OR experience.	2.4 (3) 2–3	1.9 (2) 1–3	0.14
The emergency case variety gave my resident(s) appropriate exposure to emergency surgery.	2.6 (3) 2–3.8	1.8 (1.5) 1–2.2	**0.03**
My resident(s) missed out on operative experience because of restrictions on work hours.	2.1 (3) 1–3	2.0 (1) 1–3	0.79
My resident(s) had the opportunity to develop the skills required at their stage.	3.2 (3) 3–4	2.9 (3) 3–3	0.06
The nursing staff disliked when my resident(s) operated as the operations took longer.	1.1 (1) 0–2	1.2 (1) 0–1	0.57
My resident(s) were often too busy doing other work to go to the OR.	0.8 (0.5) 0–1	1.1 (1) 0–1	0.27
My resident(s) were often too tired to get the most out of teaching in the OR.	0.5 (1) 0–1	0.7 (1) 0–1	0.34
My resident(s) were asked to perform operations alone that they did not feel competent at.	0.3 (0) 0–1	0.6 (0.5) 0–1	0.20
My resident(s) were paged during operations.	3.1 (3) 3–4	3.2 (3) 3–4	0.89
The operative cases were too long to facilitate a positive learning experience.	0.9 (1) 1–1	1.1 (1) 1–1	0.43
Total score	77.5 (76) 68–79	76.3 (75) 71–80.5	0.69

Strongly agree = 4; agree = 3; unsure = 2; disagree = 1; strongly disagree = 0. M(m)IQR, mean (median) interquartile range.

*The reported p values were non-Bonferroni-corrected. The bolded p values reflect significant differences based on $p < 0.05$. With Bonferroni correction, the level of significance was adjusted from 0.05 to 0.05/34 = 0.0014, and no significant differences were found between survey responses obtained before and after PASS implementation.

trainees' operative performances in a more timely manner, their success is consistently hindered by the fact that most existing performance assessment tools are cumbersome[15,16] and assessments are either not performed at all or are deferred to a more convenient time when their validity is questionable. Ultimately, summative judgment of a resident's operative competence is based largely on the number of procedures they have independently added to their ACGME procedural log, which currently does not include reliable supportive information to define the level of competence the resident achieved in the performance of any of the procedures they have logged.[17,18]

We have used the Zwisch scale of progressive procedural autonomy[6,7] to monitor each resident's progression of

TABLE 3. Comparison of Resident OREEM Survey for 2012 (Baseline Period) and 2013 (Study Period)

	2012 M(m)IQR	2013 M(m)IQR	p*
My attending and I got along well.	3.6 (4) 3–4	3.6 (4) 3–4	0.57
My attending was enthusiastic about teaching.	3.3 (3) 3–4	3.3 (3) 3–4	0.88
My attending had a genuine interest in my progress.	3.4 (3) 3–4	3.2 (3) 3–4	0.41
I understood what my attending was trying to teach me.	3.4 (3) 3–4	3.4 (4) 3–4	0.91
My attending's surgical skills were very good.	3.4 (4) 3–4	3.5 (4) 3–4	0.51
My attending gave me time to practice my surgical skills in the OR.	3.2 (3) 3–4	3.0 (3) 2.5–4	0.60
My attending immediately took the instruments away when I did not perform well.	1.4 (1) 1–2	1.3 (1) 1–1	0.50
Before the OR, my attending discussed the surgical technique planned.	1.8 (1) 1–3	1.7 (1) 1–3	0.80
Before the OR, my attending discussed what part of the procedure I would do.	1.3 (1) 1–1	1.3 (1) 1–2	1.00
My attending expected my surgical skills to be as good as his/hers.	1.6 (1) 1–2	1.2 (1) 1–1	0.09
My attending gave me feedback on my performance.	2.6 (3) 2–3	2.5 (3) 1.5–3	0.67
My attending's feedback was constructive.	2.8 (3) 2–3	3.0 (3) 3–4	0.54
My attending was in too much of a rush to let me operate.	1.3 (1) 1–2	1.5 (1) 1–2	0.48
The atmosphere in the OR was pleasant.	3.1 (3) 3–4	3.3 (3) 3–4	0.42
In the OR, I was corrected in front of medical students, nurses, or other residents or all of them in an unprofessional manner.	0.7 (1) 0–1	0.5 (0) 0–1	0.12
Anesthesia pressured my attending to operate him/herself to reduce OR time.	0.5 (0) 0–1	0.5 (0) 0–1	0.94
The staff in the OR was friendly.	3.2 (3) 3–4	3.6 (4) 3–4	0.13
I felt part of a team in the OR.	3.3 (3) 3–4	3.6 (4) 3–4	0.07
I was so stressed in the OR that I did not learn as much as I could have.	1.3 (1) 1–2	0.9 (1) 0.5–1	0.18
When I was in the OR, there was nobody to cover the service.	1.4 (1) 1–3	1.4 (1) 1–2	0.80
The supervision level my attending provided in the OR was adequate for my level.	3.0 (3) 3–3	3.1 (3) 3–4	0.33
Operations performed on this rotation were generally too complex for my level.	1.1 (1) 1–1	1.2 (1) 0.5–1	0.88
The elective OR case list had an appropriate mix of cases to suit my training.	2.9 (3) 3–3	3.1 (3) 3–4	0.25
I had plenty of opportunity to assist.	2.9 (3) 3–3	3.1 (3) 3–4	0.18
There were enough OR days/week for me to get the appropriate experience.	2.8 (3) 3–3	3.2 (3) 3–4	0.10
Senior residents or fellows took away my opportunities to operate.	1.5 (1) 1–2.5	1.1 (1) 1–1	0.12
The # of emergency cases was enough for me to gain adequate OR experience.	2.3 (3) 1–3	2.2 (2.5) 1–3	0.61
The emergency case variety gave me appropriate exposure to emergency surgery.	1.9 (1) 1–3	1.9 (2) 1–3	0.83
I missed out on OR experience because of restrictions on work hours.	1.5 (1) 1–2.2	1.2 (1) 0.5–1	0.28
I had the opportunity to develop the skills required at my stage.	3.0 (3) 3–3	2.9 (3) 2.5–3	0.70
The nursing staff disliked, when I operated as the operation took longer.	1.2 (1) 1–2	1.2 (1) 0.5–1.5	0.68
I was too busy doing other work to go to the OR.	1.2 (1) 0–1	1.3 (1) 0–2.5	0.88
I was often too tired to get the most out of teaching in the OR.	1.2 (1) 1–1	1.0 (1) 1–1	0.28
I was asked to perform operations alone that I did not feel competent at.	0.9 (1) 0–1	0.5 (1) 0–1	0.09
I was paged during operations.	3.1 (3) 3–4	3.1 (3) 3–3	0.49
The operative cases were too long to facilitate a positive learning experience.	0.9 (1) 1–1	0.8 (1) 1–1	0.47
Total score	75.2 (76.5) 71–82.2	75.6 (74) 69.5–83.5	0.91

Strongly agree = 4; agree = 3; unsure = 2; disagree = 1; strongly disagree = 0. M(m)IQR, mean (median) interquartile range.

*The reported p values were non-Bonferroni-corrected. Based on $p \leq 0.05$ no significant differences were found between survey responses obtained before and after PASS implementation. With Bonferroni correction with the level of significance adjusted from 0.05 to 0.05/34 = 0.0014, no significant differences were found between survey responses obtained before and after PASS implementation.

autonomy with core surgical procedures. We believe that autonomy provides a very clear and relevant long-term goal for the trainee that helps put their learning needs into perspective for the faculty members.[2] The Zwisch scale serves to break the progression to autonomy into defined stages based on the amount of guidance that is provided by the attending surgeon during the procedure.[6,7] Diminishing faculty guidance reciprocally correlates with increasing resident autonomy. It is also important to clarify that the Zwisch scale distinguishes between guidance and supervision. Guidance is physical or verbal help provided to the trainee, whereas supervision is oversight of the trainee. In a

training environment supervision is always required, but faculty can provide supervision without providing guidance (i.e., supervision only) in circumstances where a resident has demonstrated competence in performing a procedure independently.

To dilute the effect of assessment variability related to individual faculty behaviors, trainee experiences, and procedural circumstances it is important to collect data on as many procedural experiences as possible, ideally from every procedure a resident logs.[19] To facilitate its timely usage[19] with every procedure, we have linked the Zwisch scale to a prototype smartphone-based system, the PASS. PASS is linked to the OR database and is set up to automatically send a prompt to the faculty surgeon's smartphone at the end of every surgical procedure they perform reminding them to assess the performance of the resident with the procedure that they just performed together. The performance assessment requires 2 button pushes by faculty members to complete: one to assign a Zwisch level based on the level of autonomy achieved by the resident in that procedure, and a second one to assign the level complexity or difficulty or both of the procedure based on the faculty assessor's previous experience.

PASS was designed to facilitate the completion of OR performance assessments in just seconds while "on-the-fly," so that the workflow of a busy surgical practice is not interrupted.[15] Although more complex evaluation forms have been the norm with OR performance assessments, studies have repeatedly demonstrated that these are performed infrequently because they are time consuming, not always linked to relevant anchors, and can usually be completed only by interrupting workflow and stopping to complete either a paper or computer-based form.[15] Given that smartphones are now ubiquitous and are commonly used by individuals while in motion, they serve as an effective platform for performing high frequency assessments in a busy workplace environment. Collection of high volumes of assessment data will help to provide data of higher validity that will serve to monitor the progress of individual residents, the teaching behaviors of individual faculty, and the dynamics of the OR teaching environment. With wide scale use, this system could be used to define national norms and set standards for procedural autonomy. We believe that versatile technology-based systems, such as PASS, that focus on essential competency-based performance parameters are needed to facilitate broad sampling of trainee workplace performances with entrustable professional activities such as surgical procedures. This would help facilitate the collection of sufficient performance data to help map progressive milestone statuses for trainees and support a competency-based curriculum.

Despite its potential ease of use, faculty members have raised concerns that, by shifting the emphasis to resident autonomy, PASS implementation would contribute to a prolongation in OR times by increasing the proportion of each procedure that would be done by the resident. Concerns were also raised that this would lead to increased tension in the OR, particularly on behalf of faculty. In this study, our results demonstrate that implementation of PASS for evaluation of residence performance in the OR did not increase mean OR times for most (i.e., 19 of 20) of the procedures evaluated, and although it did not appear to have a negative effect on OR satisfaction as perceived by surgical residents, surgical faculty responses to some individual survey items suggest that it had some negative effect on them.

Finding that mean OR times were not increased during PASS implementation is a critical piece of information. However, this does not eliminate the possibility that prolonged OR times may occasionally be an issue, as the emphasis on progressive resident procedural autonomy increases. With notifications sent after every OR procedure, faculty members will be constantly reminded of the Zwisch operative autonomy–based evaluation system. This will potentially increase their awareness of the importance of facilitating progressive resident autonomy in the OR. As they consider their residents' level of autonomy with every OR case, they may allow residents more time to struggle with difficult parts of the case to enhance their progression to autonomy. It is anticipated that such a change in OR dynamics will be associated with significant improvements in residents performance and a faster progression to competency and autonomy. Surgical trainees finishing their training will be better prepared to independently perform surgical procedures safely and effectively when they enter independent practice[20] or fellowship training.[21] This will potentially have a positive effect on patient outcomes that will likely outweigh any potential negative effect of prolonged OR times in training scenarios. Of course, it will always be absolutely necessary to maintain a delicate balance between enhancing progressive autonomy for trainees and maintaining safety and efficiency in the OR. Practical training strategies, such as focusing on progressive autonomy for only predetermined portions[22] of the procedure within a trainee's zone of proximal development,[23] will need to be invoked to minimize the potential negative consequences of prolonged OR times.

The 7 items on the faculty OREEM survey that were associated with significantly different responses before and after PASS implementation warrant further consideration. This should be prefaced by the fact that with Bonferroni correction, no significant differences were found in comparing any of the individual faculty OREEM survey items before and after PASS implementation. However, because the Bonferroni correction is a very conservative method[9,10] that increases the risk of type II errors and as no significant differences were found with any of the resident OREEM survey items with or without Bonferroni correction, we will further discuss the faculty OREEM survey items where

significant differences were identified without Bonferroni correction.

Of these items, 4 items (items 1, 12, 16, and 17) primarily address interpersonal issues relevant to OR ambiance, that is, (1) my resident(s) and I got along well; (2) the atmosphere in the OR was pleasant; (3) the staff in the OR was friendly to my resident(s); and (4) my resident (s) felt like part of a team in the OR. Faculty responses after PASS implantation reflected a lesser degree of agreement with these statements. These responses reflect a more negative perception of the OR environment by faculty members after PASS implementation. One could speculate that PASS implementation, by increasing awareness of progressive autonomy, may have led to increased resident participation that created greater angst among surgical faculty members as they were subjected to increased periods of time watching trainees struggle. This may have created a less relaxed atmosphere in the OR and more formal, mentor-trainee relationship that may have contributed to a decrease in faculty OR satisfaction. Although some increase in faculty members angst may be an unavoidable consequence of enhancing progressive autonomy in surgical residents, it will be important to monitor this carefully to ensure that the potential benefits of enhancing progressive autonomy with surgical trainees outweigh the risks. Finally, although the differences in faculty responses after PASS implementation were statistically significant, the relevance of these differences is less clear as mean faculty responses still remained within Likert scale ranges indicating that overall faculty responders still "agreed" or "strongly agreed" that "My resident and I got along well," and that "The atmosphere in the OR was pleasant," etc.

The other 3 OREEM items (items 9, 20, and 27) that generated more negative faculty responses after PASS implementation address issues relevant to residents' skills and experience that may only indirectly affect OR ambiance, that is, (1) I expected my resident's surgical skills would be better than they actually were; (2) the emergency case variety gave my resident(s) appropriate exposure to emergency surgery; and (3) the level of supervision I provided in the OR was adequate for my resident's level." These differences in faculty responses after PASS implementation reflect an increased level of awareness on behalf of the faculty members. Faculty members were in greater agreement that their residents' surgical skills were less than they expected, and agreed less that their residents had adequate exposure to emergency surgery and that they, the faculty, were providing adequate supervision for their residents in the operating room. It is possible that implementation of a frequently used operative performance assessment instrument such as PASS contributed to greater awareness among the faculty members of their residents' OR performance and of other factors relevant to their relationship with their residents in the OR. Increased awareness should be considered a positive consequence as this indicates that individual faculty surgeons are paying greater attention to the needs of their residents, the performance levels they achieve during a procedure, and the level of supervision they need to provide to their residents during these procedures. Thus, by adopting the Zwisch scale, they may have recalibrated their perspective of OR performance assessment of resident operative performance and faculty supervision.

Limitations

The major limitation of this study is that it represents the experience of a single institution. Mean OR times, faculty-resident relationships in the OR, and faculty members willingness to engage in frequent resident performance assessments that may differ at other institutions. The PASS system may also be difficult to deploy with most surgery residency programs as it requires significant IT support. We have subsequently developed a modified version of PASS called System for Improving and Measuring Procedural Learning that is more versatile and provides additional features including enabling resident to self-assess their performance and faculty members to dictate specific feedback.[24] As with any innovation, we believe it is essential to not only to assess and validate its effectiveness but also to monitor for potential negative consequences.

Additionally, as the OREEM tool asks faculty members and residents to answer via a Likert scale from 1-5, there was no way to collect information on reasons behind agreeing or disagreeing with any given statement. Thus, we can only speculate on reasons for differences in OREEM scores before and during PASS implementation. However, to preserve confidentiality with the OREEM surveys, we were unable to match pre-PASS and post-PASS survey responses for individual faculty members or residents. This limits the power of the statistical analysis pertaining to comparison of OREEM survey item responses before and after PASS implementation. Finally, we could not control for coincidental institutional factors unrelated to PASS, such as changes in reimbursement, OR policy, administrative structure, etc., that may have influenced faculty OR satisfaction with the survey performed after PASS implementation.

CONCLUSIONS

This study demonstrates that PASS implementation does not appear to significantly influence mean OR times for most commonly performed procedures. Overall OR satisfaction as perceived by general surgery residents did not change significantly during the PASS implementation period. With full-time general surgery teaching faculty members, PASS implementation was associated with some changes in their perspectives regarding OR satisfaction.

Although additional research is required to monitor this and to determine the effect and relevance of these changes, the overall results suggest that an autonomy-based evaluation system can be employed without a major negative effect on the OR environment.

REFERENCES

1. Bell RH. Why Johnny cannot operate. *Surgery*. 2009;146(4):533-542.
2. Meyerson SL, Teitelbaum EN, George BC, Schuller MC, DaRosa DA, Fryer JP. Defining the autonomy gap: when expectations do not meet reality in the operating room. *J Surg Educ*. 2014;71(6):e64-e72.
3. DaRosa DA, Skeff K, Friedland JA, et al. Barriers to effective teaching. *Acad Med*. 2011;86(4):453-459.
4. Kiran RP, Ahmed Ali U, Coffey JC, Vogel JD, Pokala N, Fazio VW. Impact of resident participation in surgical operations on postoperative outcomes: National Surgical Quality Improvement Program. *Ann Surg*. 2012;256(3):469-475.
5. Papandria D, Rhee D, Ortega G, et al. Assessing trainee impact on operative time for common general surgical procedures in ACS-NSQIP. *J Surg Educ*. 2012;69(2):149-155.
6. George BC, Teitelbaum EN, Meyerson SL, et al. Reliability, validity, and feasibility of the Zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 2014;71(6):e90-e96.
7. DaRosa DA, Zwischenberger JB, Meyerson SL, et al. A theory based model for teaching and assessing residents in the operating room. *J Surg Educ*. 2013;70(1):24-30.
8. Kanashiro J, McAleer S, Roff S. Assessing the educational environment in the operating room—a measure of resident perception at on Canadian institution. *Surgery*. 2006;139(2):150-158.
9. Armstrong RA. When to use the Bonferroni correction. *Opthalmic Physiol Opt*. 2014;34(5):502-508.
10. Ludbrook J. Multiple comparison procedures updated. *Clin Exp Pharmacol Physiol*. 1998;25(12)1032-1037.
11. Bell RH, Biester TW, Tabuenca A, et al. Operative experience of resident in US general surgery programs: a gap between expectation and experience. *Ann Surg*. 2009;249(5):719-724.
12. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg*. 2011;202(4):469-480.
13. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess*. 2011;15(1):1-162.
14. Williams RG, Chen XP, Sanfey H, Markwell SJ, Mellinger JD, Dunnington GL. The measured effect of delay in completing operative performance ratings on clarity and detail of ratings assigned. *J Surg Educ*. 2014;71(6):132-138.
15. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ Theory Pract*. 2015;21(2):455-473.
16. Beard J. Workplace-based assessment: the need for continued evaluation and refinement. *Surgeon*. 2011;9(suppl 1):S12-S13.
17. Safavi A, Lai S, Butterworth S, Hameed M, Schiller D, Skarsgard E. Does operative experience during residency correlate with reported competency of recent general surgery graduates? Can J Surg 2012;55(4 suppl 2):S171-S177.
18. Nygaard RM, Daly SR, Van Camp JM. General surgery resident case logs: do they accurately reflect resident experience? *J Surg Educ*. 2015;72(6):e178-e183.
19. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med*. 2003;15(4):270-292.
20. Sachdeva AK, Flynn TC, Brigham TP, et al. Interventions to address challenges associated with the transition from residency training to independent surgical practice. *Surgery*. 2014;155(5)867-882.
21. Grover BT, Kothari SN. Fellowship training: need and contributions. *Surg Clin North Am*. 2016;96(1):47-57.
22. Roberts NK, Williams RG, Kim MJ, Dunnington GL. The briefing, intraoperative teaching, debriefing model for teaching in the operating room. *J Am Coll Surg*. 2009;208(2):299-303.
23. Sadideen H, Kneebone R. Practical skills teaching in contemporary surgical education: how can educational theory be applied to promote effective learning? *Am J Surg*. 2012;204(3):396-401.

24. Bohnen JD, George BC, Williams RG, et al. The feasibility of real-time intraoperative performance assessment with SIMPL (system for improving and measuring procedural learning): early experience from a multi-institutional trial. *J Surg Educ*. 2016;73(6): e118-e130 http://dx.doi.org/10.1016/j.jsurg.2016.08.010.