Measuring Surgical Trainee Perceptions to Assess the Operating Room Educational Environment

Gouri B. Diwadkar, MD, and J. Eric Jelovsek, MD

Obstetrics, Gynecology and Women’s Health Institute, Cleveland Clinic, Cleveland, Ohio

OBJECTIVE: To determine measurable differences in the perception of learning between junior and senior residents in the operating rooms of an obstetrics and gynecology (OBGYN) residency program.

DESIGN, SETTING, AND PARTICIPANTS: Using a cross-sectional design, the Operating Room Educational Environment Measure (OREEM), a 40-item educational environment inventory, was administered to 28 OBGYN residents from 1 training program, who train at 3 hospital sites. The OREEM measures a trainee’s perceptions of the teaching surgeon, learning opportunities, operating room atmosphere, and workload. The primary outcome was total OREEM scores and secondary outcomes were OREEM subscale scores, global impression of education, and internal consistency and validity of the OREEM scale. Group sample sizes of 10 and 10 achieved 80% power to detect a 10% difference between group mean OREEM scores with a significance level of 0.05.

RESULTS: Twenty-four residents including 11 junior (postgraduate years 1 and 2) and 13 senior (postgraduate years 3 and 4) residents were included in the analysis. Total OREEM scores, learning opportunities, and workload/support subscale scores were significantly lower for junior residents compared with senior residents across all sites. Perceptions of learning at a multispecialty tertiary referral hospital were lower than the community and regional hospitals. This was secondary to complexity of cases, subspecialty fellows, and decreased opportunities to first-assist in the operating room. The OREEM demonstrated acceptable reliability and construct validity.

CONCLUSIONS: There are measurable differences in perception of the operating room educational environment between junior and senior OBGYN residents using the reliable and valid Operating Room Educational Environment Measure. (J Surg 67:210-216. © 2010 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: educational environment, surgical trainee perceptions

COMPETENCY: Professionalism, Interpersonal and Communication Skills, Practice-Based Learning and Improvement, Systems-Based Practice

INTRODUCTION

The Accreditation Council for Graduate Medical Education’s (ACGME) mission is to improve health care in the United States through educational accreditation. The Accreditation Council for Graduate Medical Education recently began exploring 1 of the most important variables to impact a trainee’s learning—different learning environments—to determine why some residency programs succeed better than others at providing high-quality patient care and high-quality resident learning. Learning environments, atmospheres in which learning takes place, play a crucial role in an institution’s educational effectiveness and are frequently perceived by trainees to introduce both constructive and dysfunctional challenges linked to academic achievement and learning.

The operating room is a unique environment in which learning is shaped by the people (surgeon, anesthetist, nurses, students, and other trainees), complexity of the case, and distractions, such as pagers and telephone calls. The operating room environment is an important variable to measure because learning new surgical procedures requires mental assimilation of a large amount of new incoming sensory data within existing data in a stressful setting. A background of distractions, particularly early in training, may significantly impact a trainee’s perception of learning skills. Additionally, training programs are frequently mandating residents to work and learn in multiple environments; climates that potentially influence learning, exposing trainees to a variety of training opportunities with unique fiscal demands and patient populations. Some learning environments may be stressful or outright dysfunctional, exposing trainees to negative experiences such as gender bias, sexual harassment, or poor role-modeling. These sometimes hidden variables in a training curriculum demonstrate the need for reliable quality
assurance of obstetrics and gynecology (OBGYN) educational programs, taking into account important, hidden relationships and associations that exist between trainees responsible for patient care and their learning environments. These conditions, unless measured and improved, can create or allow negative educational experiences to continue.1–4

The overall objective of this study was to measure and compare the educational environment of our operating rooms, with the goal to improve the educational experiences of our trainees and identify variables that may affect learning. The primary aim was to determine whether there are measurable differences in the perception of learning between junior and senior residents. Our hypothesis was that there would be a less favorable perception of learning among junior residents compared with senior residents, because junior residents frequently report the need for improved learning in the operating room compared with senior residents. Secondary aims were to compare trainees’ perceptions of learning among the 3 different training medical centers in which our OBGYN residents operate during their 4 years of training, and identify specific areas where the operating room environment supports versus limits our trainees’ perception of learning. Because this was an initial report using an instrument to measure the operating room learning environment in a United States training program and in OBGYN trainees, we also wanted to examine the reliability and validity of the assessment scale in this sample.

METHODS

This study was approved by the Institutional Review Board at our institution. The 40-item Operating Room Educational Environment Measure (OREEM) questionnaire,4 designed to assess the operating room educational environment, was administered online, via Survey Monkey® (www.surveymonkey.com) to all 28 OBGYN residents, ranging from postgraduate years (PGYs) 1 to 4, in our program. The survey was anonymous, and confidentiality was assured at the start of the survey. Each resident was required to only indicate his or her level of training; gender was not identified due to 1 male resident in the program. Each resident was instructed to refer to their most recent operative experience with an attending surgeon. All residents were asked to complete a separate OREEM questionnaire for each training site, because residents receive surgical training at 3 different hospitals: Training Site A (PGY 1 to 4), Training Site B (PGY 1 to 4), and Training Site C (only PGY 2). Site A is a large, multispecialty, tertiary referral healthcare center that trains OBGYN residents and board-approved fellows in Gynecologic Oncology and Female Pelvic Medicine and Reconstructive Surgery. Site B, a regional medical center, located in the city of Cleveland, serves patients of the greater Cleveland area, including a majority of patients in underserved communities. In addition to training residents, this institution also trains fellows in Maternal Fetal Medicine. Site C, located in the eastern suburbs of Cleveland, Ohio, is a 424-bed hospital that serves the communities surrounding the city of Cleveland. At this hospital, residents often operate 1-on-1 with attending physicians, and because training includes only second year residents, it is rare for senior trainees to compete for surgical cases. Site C also serves several private practices which may shape the educational experiences of residents in both positive and negative ways.

Historically, the OREEM originated from the 50-item Dundee Ready Education Environmental Measure (DREEM) to assess environmental factors that affect medical students’ learning.3 The Operating Room Educational Environment Measure (OREEM) includes 40 items divided into 4 subscales that measure the perceptions of the attending surgeon (items 1–13); learning opportunities in the operating room (items 14–24); operating room atmosphere (items 25–32); and workload, supervision, and support in the operating room (items 33–40) (Figure 1A). The OREEM scale and subscales have demonstrated acceptable internal consistency (overall Cronbach’s alpha coefficient = 0.87; teaching and training of attending surgeon = 0.86; learning opportunities, = 0.73; atmosphere = 0.59; and workload/supervision/support = 0.63) and validity during the assessment of the educational operating room environment for Canadian general surgical residents.4 This scale has not been used in obstetrics and gynecology trainees or surgical trainees in the US. All items are answered using a 5-point Likert scale with answers ranging from strongly agree to strongly disagree. Each item is given a score of 1 if the resident answered “strongly disagree,” 2 for “disagree,” 3 for “undecided,” 4 for “agree,” and 5 for “strongly agree.” The scoring was reversed for negative statements (items 8, 11, 14, 16, 22, 23, 26–28, 31, 33–38, 40, and 90) when determining the total score. The minimum and maximum possible total scores were 40 and 200, with higher scores indicating a more positive perception of the learning environment.

All responses from the OREEM were anonymous, and trainees were assured that their identity could not be tracked. Additional items included training location, the resident’s year of training, and a single global impression of the training environment that stated, “In general, would you say the learning environment at (location) is (poor, fair, good, very good, excellent)” We did not compare responses by gender because there was only 1 male resident in the residency program. The online questionnaire was programmed so that each question had to be answered in order to proceed to the next question, preventing missing data in each completed questionnaire.

There are a total of 28 residents (14 junior and 14 senior; 7 per year) in our program. Due to the fixed sample size, residents were grouped a priori as juniors (PGY 1 and 2) and seniors (PGY 3 and 4). We estimated that 2 residents in each training year would not participate in the survey leaving 10 junior residents and 10 senior residents available for analysis. Group sample sizes of 10 and 10 achieved 80% power to detect a difference of 10% between group mean OREEM scores with estimated group standard deviations of ± 10% in each group and with a significance level (alpha) of 0.05 using a 2-sided Mann-Whitney test.

The total OREEM score and 4 subscale scores were calculated for each resident by training level and by hospital site.
Given the structure of our rotations, we anticipated that some first-year residents would be able to provide responses for 1 or 2 sites, and none of the first-year residents would be able to provide responses for Site C because residents start this rotation in their second year. Furthermore, as the inventory was administered in the midst of the academic year, we assumed some second-year residents would not have rotated at Site C. Item responses were normally distributed based on the Shapiro-Wilk test for normality. Mean scores and 95% confidence intervals for each item on the questionnaire were calculated. Mean scores between junior and senior residents were compared using the Student t test, and mean scores among the different hospitals were compared using 1-way analysis of variance (ANOVA) testing. Tukey Honestly Significant Differences (HSD) test was used for multiple comparisons. Based on a 5-point Likert scale, a score of 3 was “undecided” and therefore corresponded to the 50th percentile; a score of 4 (“agree”) corresponded to the 75th percentile; and a score of 5 (“strongly agree”) to the 100th percentile. Therefore, the educational environment was determined to be less than satisfactory if any mean score was below 4, or 75%.

Overall and subscale OREEM scores were correlated with responses on the global impression question using Spearman correlation coefficient to assess convergent validity, a type of construct validity. Spearman correlation coefficient values were considered excellent if absolute values $\geq 0.60$, adequate between 0.31–0.59, and poor $\leq 0.30$. Internal consistency was evaluated using Cronbach’s alpha, corrected item total, and Cronbach’s alpha if item were deleted. The internal consistency was rated as positive if Cronbach’s alpha was $> 0.7$ (adequate) and $> 0.8$ (excellent), when the corrected item total was between 0.3 and 0.9, and when Cronbach’s alpha if item were deleted did not exceed the overall Cronbach’s alpha. A positive rating indicates that all items evaluate different aspects of the same construct. All statistical analyses were performed using JMP 7.0 (SAS Institute, Cary, NC), and internal consistency analyses were performed using NCSS Version 2005 (Kaysville, UT).

RESULTS

Twenty-seven out of 28 residents responded to the questionnaires resulting in a 96% response rate, with data missing on a single third-year resident. Three first-year residents were excluded from the study because they had not participated in a surgical rotation at the time in the academic year when the survey was administered. A total of 24 residents were included in the analysis, 11 junior (4 PGY 1, 7 PGY 2) residents and 13 (6 PGY 3, 7 PGY 4) senior residents. One first-year resident had not yet completed a surgical rotation at Site A, none of the first-year residents had completed a surgical rotation at Site B, and 3 out of 7 second-year residents had not rotated at Site C. Therefore, 60 questionnaires were available by hospital site including 23 residents at Site A (multispecialty, tertiary referral center), 20 from Site B (inner city medical center), and 17 from Site C (community hospital).

Comparisons by Level of Training

Mean total scores [Juniors 137 (standard deviation 129-144), Seniors 152 (147-157), p < 0.001], learning opportunities subscale scores [Juniors 31(28-34), Seniors 38 (36-40), p < 0.0003] (p < 0.0003), and workload and support subscale scores [Juniors 27(25-29), Seniors 31(29-32), p < 0.004] were significantly suboptimal in junior residents compared with senior residents at all sites. We were unable to detect significant differences in mean scores between training levels on the attending [Juniors 46 (43-50), Seniors 50 (48-52), p < 0.08] and atmosphere subscales [Juniors 32 (30-34), Seniors 34 (33-35), p < 0.13]. Additionally, scores between junior and senior residents at Site C could not be compared because only juniors rotate at that site.

Comparisons by Training Site

Mean total OREEM scores [Site A 136 (standard deviation 130-142), Site B 158 (151-165), Site C 151 (144-158), p < 0.0001], attending [Site A 44 (42-47), Site B 54 (51-57), Site C 49 (46-52), p < 0.0001], learning opportunities [Site A 33 (30-36), Site B 39 (36-42), Site C 36 (33-39), p < 0.02], and atmosphere [Site A 31 (29-32), Site B 35 (34-37), Site C 35 (33-37), p < 0.0001] subscale scores were significantly different between the 3 sites. Site B had the highest perceived learning environment with significantly higher mean total scores, attending, and learning opportunities subscale scores. On the atmosphere subscale, scores were similar between Sites B and C; both were superior to Site A. There were no significant differences between mean workload and support scores among the 3 medical centers [Site A 29 (26-30), Site B 30 (28-32), Site C 31 (29-33), p < 0.07].

When comparing levels and sites, senior residents at Site B had the highest mean total score, and junior residents at Sites A and B had the lowest mean total scores. Tukey Honestly Significant Differences results revealed significant difference in total scores between seniors and juniors at Site B and between seniors at Site B and juniors at Site A.

Satisfactory Scores (Figure 1A and B)

The 9 items with “satisfactory” mean scores (mean item score $\geq 4$) were, “My attending has a pleasant personality,” “I get along well with my attending,” “My attending is enthusiastic about teaching,” “My attending has a genuine interest in my progress,” “I understand what my attending is trying to teach me,” “My attending’s surgical skills are very good,” “I feel discriminated against in the operating room because of my sex,” “I feel discriminated against in the operating room because of my race,” and “I am asked to perform operations alone that I do not feel competent at performing.”

The remainder of the questionnaire items were found to be less than satisfactory (mean item score $< 4$) indicating areas in need of further investigation and improvement. The 3 lowest scoring items (mean item score $< 2$) were, “The number of
emergency procedures is sufficient for me to gain the right operative experience,” “The variety of emergency cases gives me the appropriate exposure,” and “I get paged during operations.” These items take priority in future program evaluations.

**Validity and Reliability**

An excellent correlation was found between global impression scores and total, attending subscale, and atmosphere subscale scores, and an adequate correlation was found between global impression and the learning opportunities subscale scores (Table 1). There was poor correlation for workload subscale scores, indicating that either the scale does not adequately assess this domain, or there is a paradox in trainee perception that although trainees rate the workload and support in their operating room learning environment as less than satisfactory, they rated the overall learning environment as conducive to learning. Internal consistency was excellent based on Cronbach’s alpha of 0.97, Cronbach’s alpha if item deleted of 0.967, and corrected item total between 0.3–0.9, except for item 16 (Table 2). 

**DISCUSSION**

This study has demonstrated the feasibility of administering the OREEM in an OBGYN training program to identify areas for improvement in the operating room learning environment. Junior residents are busier than senior residents because they simultaneously cover the wards when in the operating room, field more pages from nursing and support staff on the wards, and report more fatigue compared with senior residents. Additionally, junior residents perceive that the cases are too complex for their level of training, failing to understand the material being taught, resulting in fewer opportunities to first-assist and develop their surgical skills. Moreover, when given the opportunity to operate, junior residents perceive that the nurses are more frustrated, possibly from pages that need to be answered, or secondary to increased operative time. However, interestingly, it did not appear that junior residents’ global learning perception was adversely affected by having fewer learning opportunities, a higher workload, or less support.

Residents perceived the atmosphere to be more conducive to learning at Training Sites B and C compared with Site A. Upon further analyzing individual item scores, the data revealed that residents appear to have increased opportunities to practice surgical skills for emergent and elective cases that are more appropriate for their level, resulting in first-assisting and having an earlier chance to develop their skills. At our community hospital, residents are less likely to get paged in the operating room because the surgeon’s office staff handle patient calls throughout the day, and consults are not solely driven by a resident service. Because Gynecologic Oncology and Female Pelvic Medicine and Reconstructive Surgery fellows receive the majority of their training at Site A, this may provide some explanation why scores for case complexity and opportunities to first-assist were lower. Opportunities to work 1-on-1 with the attending surgeon may have a large part in determining the overall global impression of a teaching institution, explaining the lower overall scores at Site A. This finding will require further investigation.

These results do allow for all sites to learn from each other, verify areas of strength, and identify areas for improvement realizing that not all weaknesses can be resolved immediately. For example, learning appears limited by residents’ suboptimal experience with emergency cases (ectopic pregnancies, ovarian torsion, etc.). These rare emergency cases may need to be simulated or discussed using other methods. Residents may need to be assigned to cases more appropriate for their level of training rather than second-assisting fellows on complex cases. Having a resident or other provider solely assigned to cover the wards, while the other residents on the team are operating can reduce frequent pages and disruptions in the operating room. An additional option is to hire assistants such as physician’s assistants to cover the wards and consults during the daytime.

A strength of this study is that the different training environments appear to complement one another. Residents participate in more complex cases at Site A and more common gynecologic cases (emergent and elective) at the other 2 hospitals. One way to optimize the residents’ experience in complex cases is to design the curriculum so that fellows enhance the residents’ learning experiences rather than becoming detrimental to it. For example, during a laparoscopic hysterectomy, the junior resident can insert the trocars and start the dissection, followed by the fellow performing the more complex dissection and suturing. For a vaginal prolapse case, the resident can perform the vaginal hysterectomy, and the fellow can perform the vaginal suspension and urethral sling.

Residency directors can also use the OREEM in conjunction with the common program information form (PIF) to determine exactly where weaknesses lie. Due to a recent site visit at our program, this form happened to be distributed to our residents during the same time period that they were asked to complete the OREEM. The PIF results revealed that 68% residents felt that trainees that were not part of the residency (e.g., fellows) interfered with their education “some of the time.” Only 29% of residents felt that clinical education was emphasized over service obligations at all times. We present these findings, not to determine the relationship between the PIF and the OREEM, but to demonstrate how

### Table 1. Validity of the Operating Room Education Environment Measure (OREEM) in Obstetrics and Gynecology Trainees

<table>
<thead>
<tr>
<th>OREEM Scores</th>
<th>Spearman Correlation Coefficient</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>0.63</td>
<td>0.0001</td>
</tr>
<tr>
<td>Attending surgeon</td>
<td>0.60</td>
<td>0.0001</td>
</tr>
<tr>
<td>Learning opportunities</td>
<td>0.40</td>
<td>0.0023</td>
</tr>
<tr>
<td>Learning atmosphere</td>
<td>0.60</td>
<td>0.0001</td>
</tr>
<tr>
<td>Workload/traineer support</td>
<td>0.30</td>
<td>0.0524</td>
</tr>
</tbody>
</table>

*Significance was set at the 0.05 level.
the OREEM can be used to target specific areas of improvement based on responses from the PIF.

The limitations of this study include the small sample size, limiting the power to detect differences between individual training levels (PGY1, 2, 3, and 4) and the 3 sites. The most significant limitation is that these results are specific to the curriculum, sites, and residents in a single residency program. However, our goal was to demonstrate the value of using an instrument such as the OREEM to assess factors in a program’s learning environment that may need improvement, in addition to reassuring trainees that the context in which learning is taking place is optimal. Furthermore, given the cross-sectional study design, the results may vary slightly at a different time in the academic year due to different perceptions of workload, skills, or fellow involvement.

Future studies include determining the variance of OREEM scores across different groups of residents in the same institution with similar operating room environments or readministering the OREEM to the same group of residents after a curriculum change. The OREEM appears to be a feasible tool to assess and potentially modify variables in the learning environment that affect trainees in the operating room.

REFERENCES


**FIGURE 1A. Operating Room Education Environment Measure (OREEM)**

The purpose of this survey is to learn more about the learning environment in the operating rooms in which you operate. The survey is strictly for trainees in surgical residencies. This questionnaire assesses your perceptions of your attending surgeon, perceptions of learning opportunities and operating room atmosphere, and your perceptions of workload, supervision, and support which contribute to shaping the environment in the operating room. Your participation in this study will help identify areas for improvement that can eventually positively affect your training.

The information submitted over the internet is completely anonymous and cannot be traced back to you, and confidentiality will be maintained. If you are not sure about an answer, or you can’t remember the answer to a question, just answer as best as you can. Please check that you have not missed any questions as you go. Please complete a separate questionnaire for each of the three hospitals at which you have rotated during your residency. Please reflect on your most recent rotation at each hospital. Please refer to your current or most recent surgical experience in the operating room with respect to site and attending surgeon. If you are on service with more than one attending, refer to your most recent operative experience and answer the following regarding that single attending. The questionnaire will take approximately 10 minutes to complete.

Thank you in advance for your participation.

1. During your training, have you had a surgical rotation (excluding Labor and Delivery) during which you went to the operating room? (Please check the appropriate box.)
   - Yes
   - No

2. Level of training: (Please check the appropriate box.)
   - PGY 1
   - PGY 2
   - PGY 3
   - PGY 4

3. Training site: (Please check the appropriate box.)
   - Hospital A

**FIGURE 1.** [A] Operating Room Educational Environment Measure (OREEM) **4** Modified from the original OREEM. The questionnaire was administered online. Minor modifications to the questionnaire included the exclusion of gender identification, and hospital site and level of training were changed to reflect Obstetrics and Gynecology residency training at our institution. (B) One method to visually display educational quality control data on the learning environment in the operating room is to use a radar plot. The plot demonstrates the relationship between trainee responses (squares and solid line) and targeted satisfactory program values (circles with dotted line) on the Operating Room Educational Environment Survey in an Obstetrics and Gynecology training program (n = 27). Response values are on a Likert scale where 1 = strongly disagree, 2 = disagree, 3 = unsure, 4 = agree, 5 = strongly agree. Any score ≥ 4 indicates a satisfactory mean score.
Hospital B
Hospital C

4. In general, would you say the learning environment in the operating room at Hospital (A, B, or C) is (please check the appropriate box):
   Poor
   Fair
   Good
   Very Good
   Excellent

Please answer the following questions by indicating whether you:
   Strongly agree
   Agree
   Unsure
   Disagree
   Strongly disagree

1. My attending has a pleasant personality.
2. I get along with my attending.
3. My attending is enthusiastic about teaching.
4. My attending has a genuine interest in my progress.
5. I understand what my attending is trying to teach me.
6. My attending’s surgical skills are very good.
7. My attending gives me the time to practice my surgical skills in the operating room.
8. My attending immediately takes the instruments away when I do not perform well.
9. Before the operation, my attending discusses the surgical technique planned.
10. Before the operation, my attending discusses what part of the procedure I will perform.
11. My attending expects my surgical skills to be a good as his/hers.
12. My attending gives me feedback on my performance.
13. My attending’s criticism is constructive.
14. The operations performed on this rotation are too complex for my level.
15. There is a right mix of cases to suit my training.
16. There are too many cases on the operating room schedule to give me the opportunity to operate.
17. I get enough opportunity to assist.
18. There are enough operating room days per week for me to gain the appropriate experience.
19. More senior residents or fellows take away my opportunity to operate.
20. The number of emergency procedures (ectopic pregnancy, ovarian torsion, etc.) is sufficient for me to gain the right operative experience.
21. The variety of emergency cases gives me the appropriate exposure.
22. My attending is in too much of a rush during emergency cases to let me operate.
23. I miss out on operative experience because of restrictions on working hours.
24. I have the opportunity to develop the skills required at my stage.
25. The atmosphere in the operating room is pleasant.
26. In the operating room, I don’t like being corrected in front of medical students, nurses, and residents.
27. The nursing staff dislike when I operate as the operation takes longer.
28. The anesthetists put pressure on my attending to operate him/herself to reduce anesthetic time.
29. The staff in the operating room are friendly.
30. I feel discriminated against in the operating room because of my sex.
31. I feel discriminated against in the operating room because of my race.
32. I feel part of a team in the operating room.
33. I am too busy doing other work to go to the operating room.
34. I am often too tired to get the most out of teaching the operating room.
35. I am so stressed in the operating room that I do not learn as much as I could.
36. I am asked to perform operations alone that I do not feel competent at.
37. When I am in the operating room, there is nobody to cover the ward.
38. I get paged during operations.
39. The level of supervision in the operating room is adequate for my level.
40. The operating cases are too long.