Using Individual Residents’ Learning Trajectories to Better Understand the Impact of Gaps in Practice
Stephanie A. Scott, MD, MHPPE, Nancy Van Eyk, MD, MSc, and Kevin Eva, PhD

Abstract

Purpose
To optimize learning, health professional training programs need to achieve the right balance between depth of practice (gaining more experience with particular skills) and breadth of practice (spreading experience across an array of activities). Better understanding how training for a particular skill set is impacted by periods of focus on a different skill set would allow improved curriculum and assessment design, thereby enhancing the efficiency of training and effectiveness of care. To this end, learning curves were used to compare performance in surgery after prolonged periods of practice to performance after gaps in surgical training.

Method
Daily operative assessments from the Dalhousie obstetrics and gynecology program were analyzed retrospectively and learning curves were generated. In addition to examining the variability in learning trajectories, the impact of gaps was systematically assessed by comparing resident scores after 2 successive months in which they were not assessed operatively to those collected after 2 successive months in which they were assessed at least once.

Results
Four thousand four hundred sixteen scores for 33 residents over a 10-year period were analyzed. Trajectories and peak performances were identified. Residents performed better during their third sequential month of being assessed (mean = 4.40, 95% CI = 4.33–4.46) relative to months following a period of being away from the operating room for at least 2 months (mean = 4.21, 95% CI = 4.13–4.29; P < .01; d = 0.7). However, maximum performance achieved was more strongly related to the number of times residents experienced a gap in training (r = 0.50) than to the number of times residents experienced 3 consecutive months of training (r = 0.25).

Conclusions
Distinct patterns of development exist for individual residents. Time away from surgical practice and assessment negatively impacted short-term performance, but may improve long-term learning trajectories. This speaks to the value of spaced education and is important for the design of longitudinal skills-based training programs.

Health professional education programs worldwide are commonly composed of a variety of rotations aimed at providing trainees with experiences that are important for achieving competence in the array of roles and tasks that will be expected in their future practice. For example, trainees in a postgraduate specialty program might take a month to a year away from clinical training to complete a research or education elective. Within the realm of clinical training, surgical programs offer another example given that they routinely include nonoperative rotations such as when gynecologic surgery residents are expected to spend a large portion of time gaining experience in obstetrics.1 On one hand, these gaps create cause for concern because they remove trainees from concentrated opportunities to engage in deliberate practice through ongoing repetition.2 To continue the above example, if one knows they are on the career track of a gynecologic surgeon, time away to fulfill broader program requirements can be seen as a delay-inducing hoop one must jump through rather than a valuable broadening of one’s background. On the other hand, research in both basic psychology and medical education has clearly shown benefits of “spaced training” (i.e., practice in a way that requires learners to reactivates their mental models after a time away from the to-be-learned material) relative to encouraging repetition of study/practice in a time compressed manner.3,4 This tension creates considerable uncertainty, frustration, and conflict for trainees and program designers alike because intuitions and priorities are not consistent and it remains unclear how educational practices might best establish a balance between depth of practice and breadth of experience. Put in the more colloquial terms popularized by the book Range, if one wants to become a successful tennis player, is it better to concentrate solely on tennis or to take time to engage in a variety of sports?5

Randomizing residents into different training models to generate a deeper understanding of how to achieve an effective balance is unlikely to be feasible or ethical considering the additional tension between learning and service that is inevitably present at advanced stages of training.6 Modern movements toward more routine assessment of clinical skills, however, offer an opportunity to tackle this problem by exploring the learning curves of individual residents. Learning curves, defined as graphical representations of the association between experience and proficiency, document the relationship between the number of events or amount of time engaged in the learning activity.
and the outcomes of learning (be they general performance scores or targeted skills). While most commonly used to explore individual differences in learning trajectories, systematic examination of learning curves in relation to differentially encountered aspects of a training program may provide clearer insight into the impact of curricular planning decisions such as when to require or allow time away from intense clinical activity.

To that end, the purpose of this study was to analyze assessments of surgical performance in trainees over the span of a 5-year surgical residency program. The natural interruptions in surgical training that occur during this 5-year period, including both nonsurgical rotations and absences from training, allowed us to explore if gaps in practice disrupt, positively or negatively, observed learning trajectories.

Method

Context

The specialty of obstetrics and gynecology provides an ideal context within which to study this issue because surgical performance is a time-specific moment during which ratings of ability can be collected and aggregated over time. Further, residents routinely rotate in and out of surgical phases of training to obtain the experience needed to master both gynecologic surgery and nonoperative obstetrics. While these topics commonly remain together within training programs, they are sufficiently distinct that whether mastery of the full breadth of both domains is truly possible has been debated.

The obstetrics and gynecology training program at Dalhousie University is one of 17 obstetrics and gynecology residency training programs in Canada, all of which are currently 5 years in length and accredited by the Royal College of Physicians and Surgeons of Canada. Relative to most programs of its type, Dalhousie has a long history of collecting daily operative assessment forms for all residents that span the length of their tenure in the program. When not in operative rotations, trainees spend the majority of their time in the practice of obstetrics which has a much smaller surgical component that is limited primarily to cesarean sections. In addition, residents spend time on research rotations and clinical electives that may involve operative or nonoperative aspects of the specialty depending on the career goals of the individual resident. While the number, timing, and length of gaps from gynecological surgery rotations are not randomized or constant and should, therefore, be considered a “complex intervention” for the sake of this study, the opportunity to practice gynecological surgery procedures is fairly constrained to such rotations, thereby making performance on such cases a useful distinct outcome relative to the broader skill set that gaps are intended to enable. Further, by examining the learning trajectories of individual residents, each trainee effectively serves as their own control when making comparisons across different points of the curriculum. That makes this context experimentally advantageous given that variability in when gaps take place and what is done during the gaps reduces the risk that a particular gap-filling activity could be the sole cause of any performance changes that follow for the group as a whole.

Data collection and participants

We examined daily operative assessment forms for all residents entering the program from 2008 to 2017. These forms were completed by surgeon educators for every resident in the program with entries completed for approximately 80% of all surgeries performed on core general gynecology rotations. The surgical skills coordinator during this time (N.V.E.), tracked individual procedures and ensured completion of forms on a weekly basis for each resident on an operative gynecology rotation as part of their rotational assessment. During the study period, 33 residents entered and completed the program and are represented in the assessment database. Residents who had not completed the training program were excluded to avoid the potential of identification because noncompletion of the program is rare, whether attributable to failure or other reasons.

The data associated with each assessment include year of training and the procedure performed. Scores were assigned, using 1–5 rating scales on which 5 indicated optimal performance, to each resident’s tissue handling, time and motion, instrument handling, knowledge of instruments, flow of the operation, and knowledge of the procedure. These parameters have been validated in a number of contexts, including gynecologic surgery, as important measures of technical ability in the assessment of surgical skills.

Given that our data revealed a very high internal consistency (Cronbach’s alpha = 0.95), the individual components were averaged to create an overall performance score (out of 5), on which all analyses were conducted. Entrustment ratings and pass–fail opinions, both dichotomous judgments offered by the supervisors, were also captured and used as validity checks for the ratings analyzed.

Data analysis

We created learning curves by mapping assessment scores against time in the program for each of the 33 residents who completed the postgraduate training program. To do so, dates were coded relative to postgraduate training year and month such that Time 0 marks when each individual started the program and month 60 indicates the end of the 5-year training program. To increase the psychometric robustness of the data, an average performance score was calculated for each month during which each resident had assessment data available.

To explore whether the apparent trajectories were mediated by more senior trainees taking on more difficult cases, the assessment data were subdivided into procedures based on relative difficulty (i.e., minor vs major procedure), as outlined in Table 1. This coding structure was created (and implemented) by the primary author (an experienced gynecologic surgeon) in an effort to group procedures based on a dichotomy commonly used by operating room staff to indicate level of complexity, invasiveness, and anticipated length of a procedure, thus allowing for planning of pre-, intra-, and postoperative care and resources.

Learning curves were then generated for each resident for both minor and major procedures to determine if distinct patterns emerged dependent on the relative difficulty of the case. Curves were examined visually to detect patterns, and peak scores were used for the purposes of this study to define the degree of competence achieved. To quantify and systematically assess the
impact of gaps in practice on learning trajectory, we isolated all scores received by each resident after 2 successive months in which they were not assessed and after 2 successive months in which they were assessed at least once. This cutoff was chosen to maximize the amount of data in 2 distinct circumstances that reflect “accumulated recent practice” and a “considerable gap from practice.” A repeated-measures ANOVA was then used to compare the mean performance score achieved in both of these conditions.

Finally, Pearson’s correlation coefficients were used to explore associations between the maximum performance achieved by individual residents and (1) the number of times they experienced a gap in gynecological surgical training, (2) the number of times they experienced 3 consecutive months of that training, (3) the total number of assessments received, (4) when their first assessment was received, (5) the minimum score received, (6) first score received, and (7) standard deviation in scores received. The first 2 analyses were aimed at addressing the primary research question regarding the influence of gaps in practice. Correlations with the latter variables were used to explore other factors that may help to account for which trainees’ learning trajectories achieved the heights expected by the end of their 5 years in residency training.

Table 1
Procedural Coding Used to Differentiate Cases Encountered Based on Complexity, Invasiveness, and Anticipated Length of Procedure

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Procedures, no. (%)</th>
<th>Average performance (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minor procedures</td>
<td>2,122 (48.1)</td>
<td>4.39 (4.37-4.41)</td>
</tr>
<tr>
<td>Dilatation and curettage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cystoscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic hysteroscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IUD insertion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEEP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Operative hysteroscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diagnostic laparoscopy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laparoscopic tubal ligation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minor vaginal/vulvar repairs/excisions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major procedures</td>
<td>2,294 (51.9)</td>
<td>4.26 (4.23-4.29)</td>
</tr>
<tr>
<td>Laparoscopic USO/BSO or ovarian cystectomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior and/or posterior repair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TVT(O)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaginal hysterectomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal USO/BSO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal hysterectomy (+/- BSO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal myomectomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total laparoscopic hysterectomy (+/- BSO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laparoscopic-assisted vaginal hysterectomy (+/- BSO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laparoscopic myomectomy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced urogynecologic procedures</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tubal reanastomosis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; IUD, intrauterine device; LEEP, loop electrosurgical excision procedure; U(B)SO, unilateral (bilateral) salpingo-oophorectomy; TVT(O), tension-free vaginal tape (obturator).

Residents deemed to have Passed (mean = 4.33, SD = 0.58) the clinical encounter were assigned higher scores relative to those deemed to have Failed (mean = 3.19, SD = 1.03; t = 11.8; P < .001; d = 1.92). Similarly, comparison of ratings for those judged to be Entrustable (mean = 4.66, SD = 0.43) relative to those deemed NOT Entrustable (mean = 3.97, SD = 0.55) support the validity of using observers’ ratings as indicators of the learning progress of these trainees (t = 44.7; P < .001; d = 1.14).

Learning curves

Figure 1 illustrates the average rating assigned to all residents within a given month of training. The first 5 months of training and any months > 60 have been truncated because fewer than 3 residents had data in each of those months. Furthermore, the latter instance represents individuals whose training was extended beyond the normal 5 years. Although the curve is jagged, there is clear linearity in the growth of ratings assigned with average ratings beginning around 3.8 at month 5 and ending around 4.7 at month 60. When similar curves were created for minor procedures and major procedures separately, the trend lines were parallel, so that distinction was not used in further analyses.

Impact of gaps in practice and assessment

To systematically assess the impact of gaps in practice, we isolated all scores received by each resident after 2 successive months in which they were not assessed and after 2 successive months in which they were assessed at least once. Calculating the average of both conditions revealed that candidates performed better during their third month of being assessed (mean = 4.40;
Workplace Assessment

Academic Medicine, Vol. 96, No. 11S / November 2021 Supplement

Table 2
Performance on Assessed Procedures as a Function of Postgraduate Training Year

<table>
<thead>
<tr>
<th>PGY</th>
<th>Total no. of procedures</th>
<th>Major procedures, no. (%)</th>
<th>Average performance (95% CI)</th>
<th>Procedures where resident was deemed entrustable, no. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>225</td>
<td>109 (48.4)</td>
<td>3.88 (3.82–3.96)</td>
<td>30 (13.3)</td>
</tr>
<tr>
<td>2</td>
<td>1,028</td>
<td>478 (46.4)</td>
<td>3.98 (3.94–4.02)</td>
<td>284 (27.7)</td>
</tr>
<tr>
<td>3</td>
<td>971</td>
<td>483 (49.7)</td>
<td>4.25 (4.23–4.29)</td>
<td>479 (49.3)</td>
</tr>
<tr>
<td>4</td>
<td>1,093</td>
<td>611 (55.9)</td>
<td>4.49 (4.46–4.52)</td>
<td>709 (64.4)</td>
</tr>
<tr>
<td>5</td>
<td>964</td>
<td>544 (56.4)</td>
<td>4.62 (4.59–4.65)</td>
<td>824 (85.5)</td>
</tr>
</tbody>
</table>

Abbreviations: PGY, postgraduate year; CI, confidence interval.

SD = 0.43) relative to after being away from the operating room assessment protocol for at least 2 months, with large effect size (mean = 4.21; SD = 0.50; P < .01; d = 0.7). However, the maximum performance achieved was more strongly related to the number of times residents experienced a gap in training (r = 0.50) than to the number of times residents experienced 3 consecutive months of training (r = 0.25).

Additional factors related to trainee achievement

Thirty-two of 33 individuals showed trends indicating growth over time with only 8 appearing likely to need more than 5 years to reach top performance. Those 8 residents achieved a maximum score of 4.73 on average, d = 1.2 standard deviations lower than the maximum of the remaining 24 residents (4.94 on average). They did not vary strongly from the higher achieving residents on their minimum score (3.62 vs 3.51, respectively; d = 0.4 standard deviations different), the score achieved on their first assessment (3.92 vs 3.77, respectively; d = 0.3 standard deviations different), or the standard deviation in their assessments over the first 30 months of their training program (0.40 vs 0.37, respectively; d = 0.3). However, they received assessments in fewer months (13.6 vs 16.2, respectively; d = 0.6) and the month at which they received their first assessment was 9 months later (month 22 on average) than the month at which the higher achievers received their first assessment (month 13; d = 1.1). Of those variables, the maximum score achieved by each resident correlated strongly with only the total number of assessments received (r = 0.55).

Discussion

Through analysis of resident learning curves we were able to identify and explore the impact of gaps in training on gynecologic surgical practice using data that reflect real-world authentic clinical performance measures. While the issue of gaps in training has wide-ranging relevance, it is particularly amenable to study in obstetrics and gynecology because residents in such programs obtain broad training in both gynecologic surgery and nonoperative obstetrics and because surgical encounters provide specific and observed moments that enable accumulation of distinct clinical skill assessments. The scope of practice results in residents spending a large portion of their training away from the operating room, which has been identified as a source of concern regarding whether that strategy is optimal for enabling trainees to obtain the experience needed for competence in complex surgical procedures. Such concerns derive from findings like those generated from a survey of 130 fellowship directors in which it was revealed that, after completion of their residency program, only 52% of trainees were believed to be competent to cover generalist call and only 46% were...
considered able to perform an abdominal hysterectomy independently.\textsuperscript{15} One argument put forward in response to such observations is that more surgical training is required, withdrawing residents from spending as much time in other activities, thereby minimizing gaps in surgical practice.\textsuperscript{1,10}

Given the breadth of skills required to be a good clinician by modern standards, there are a wide array of reasons one might benefit from variable training activities. Even limiting our focus to surgical skill, however, our examination of learning trajectories suggested an unexpectedly complex relationship between gaps in training and performance. Resident achievement was strongly associated with the number of months trainees were assessed in gynecologic surgery rotations. Looking more granularly, they performed better in the operating room after 2 months of regular practice with assessment than they did after 2 months of a nonoperative rotation. Neither finding is surprising given that skill development requires experience, and one might reasonably expect that returning to any task after a period away requires some degree of reacclimation. Put colloquially, “practice makes perfect” and time away can leave one needing to “shake off the rust.” The latter observation, however, suggests that caution needs to be exerted when claiming someone has “achieved competence” given that any skill requires ongoing practice to maintain ability.\textsuperscript{2}

The real intrigue, however, lies in the contrast between our short-term evaluation of the effect of gaps in practice and evaluation of what effect they have over the longer term. Here, we found that maximum performance (i.e., the scores achieved at the completion of training) was positively and strongly correlated with the number of times trainees took time away from surgical training ($r = 0.50$). That is, short-term pain appears to have led to long-term gain, a finding that may have important implications for curriculum design and rotation schedules, especially given that maximum achievement was unrelated to how strongly residents performed during the early stages of the residency program. Although this pattern was not anticipated, the results are consistent with the literature suggesting the value of spaced education\textsuperscript{5,6,16,17} and of gaining a wider foundation of experience from which to draw upon.

In brief, superior learning has been demonstrated in a variety of contexts when trainees practice with material or tasks they are meant to learn in a way that is distributed (i.e., spaced) over time relative to when the same amount of practice is concentrated (i.e., massed) over shorter intervals. Such spacing occurs naturally when one engages in a wider range of activity relative to focusing narrowly on one particular skill set. Common argument for why this is the case is that gaps between learning moments result in the need to relearn (i.e., reactivate the knowledge or skill), which makes the learning task harder (as reflected in the lesser scores we observed immediately after residents returning from 2 months away).

Ultimately, however, the additional effort required results in greater consolidation, strengthening learners’ ability through the accumulation of new memory traces rather than simply using ones that are already activated during massed practice.\textsuperscript{16}

There is little debate regarding the benefit of spaced practice as it has been demonstrated consistently in the literature.\textsuperscript{16,17} In surgery specifically, the benefit of distributed practice has been demonstrated in simulations,\textsuperscript{5,6,17} but to our knowledge, it has not previously been demonstrated over the course of a residency program in surgery or otherwise. In this real-world context, the periods between practice are much longer than those that have been studied in simulation settings. If, upon further study, the observed effect proves replicable, these results create an important challenge that will need to be resolved as surgical educators will need to determine how to optimally balance the importance of surgical volume with the spaced practice model emphasizing the importance of time away. To this point, little is understood regarding the ideal timing or length of gaps in practice for achieving optimal learning.\textsuperscript{3,17,18} Further systematic examination of resident learning curves may offer a way in which to explore optimization of practice schedules in residency training. For now, we conclude that the inclusion of nonsurgical rotations, to the degree used in this program, does not appear to be cause for concern regarding the degree of achievement eventually reached by gynecologic surgical residents.

**Limitations**

The retrospective nature of the data collection and lack of randomization inherent in this study limits our control over some variables that might affect resident performance. Although the procedures performed, and thus relative difficulty, was available to us and does not appear to have played a mediating role in the patterns observed, we do not have information regarding other potential contributors to procedural complexity including patient body habitus, prior surgery, or anatomic distortion from various disease processes. It is expected that such factors will have added noise to the data collected (resulting in the jagged learning curves summarized in Figure 1), but whether or not they influenced skill development or measurement more generally cannot be discerned.

Second, the naturalistic nature of this exploration means that we did not have control over what happened during gaps in practice. It is possible, in fact, that some gaps were filled training on other surgical rotations. If such periods were routinely spent improving (or at least maintaining) general surgical skill in other training contexts, however, the decline in performance in the short term (after time away from the gynecological surgery environment) should not have been observed.

Third, it is noteworthy that our data are based on 1 program within a single institution and, therefore, may not be entirely generalizable to other specialties or centers, which may have different curriculum schedules, elective options, or direct observation protocols. Finally, given the small numbers of procedures performed of each type, we were unable to determine whether or not different tasks may be best described with different learning curves.

**Conclusions**

Our study has provided insight into the application of learning curves using existing procedural assessment tools in the context of a 5-year surgical residency program. By examining the learning trajectories of individual residents, we identified patterns that may be further explored to offer insights into program design. Distinct trajectories toward competence were apparent with positively contributing variables including gaps in...
practice and assessment. Further study is required to refine the implications of these findings and to determine how to make improvements to formalized surgical curricula, competency-based medical education, resident remediation, self-directed learning, and, possibly, resident selection.

Funding/Support: None reported.

Other disclosures: None reported.

Ethical approval: This study received ethics approval from the Dalhousie University Research Ethics Board.

S.A. Scott is assistant professor and surgical skills coordinator, Department of Obstetrics and Gynecology, Dalhousie University, Halifax, Nova Scotia, Canada.

N. Van Eyk is associate professor and clinical obstetrician and gynecologist, Department of Obstetrics and Gynecology, Dalhousie University, Halifax, Nova Scotia, Canada.

K. Eva is associate director, Centre for Health Education Scholarship, and professor and director of educational research and scholarship, Department of Medicine, University of British Columbia, Vancouver, British Columbia, Canada.

References