

Teaching Surgical Skills: What Kind of Practice Makes Perfect?

A Randomized, Controlled Trial

Carol-Anne E. Moulton, MB, BS, FRACS, MEd, Adam Dubrowski, PhD,
Helen MacRae, MD, MA, FRCSC, FACS, Brent Graham, MD, MSc, FRCSC, Ethan Grober, MD,
and Richard Reznick, MD, MEd, FRCSC, FACS

Objective: Surgical skills laboratories have become an important venue for early skill acquisition. The principles that govern training in this novel educational environment remain largely unknown; the commonest method of training, especially for continuing medical education (CME), is a single multihour event. This study addresses the impact of an alternative method, where learning is distributed over a number of training sessions. The acquisition and transfer of a new skill to a life-like model is assessed.

Methods: Thirty-eight junior surgical residents, randomly assigned to either massed (1 day) or distributed (weekly) practice regimens, were taught a new skill (microvascular anastomosis). Each group spent the same amount of time in practice. Performance was assessed pretraining, immediately post-training, and 1 month post-training. The ultimate test of anastomotic skill was assessed with a transfer test to a live, anesthetized rat. Previously validated computer-based and expert-based outcome measures were used. In addition, clinically relevant outcomes were assessed.

Results: Both groups showed immediate improvement in performance, but the distributed group performed significantly better on the retention test in most outcome measures (time, number of hand movements, and expert global ratings; all P values <0.05). The distributed group also outperformed the massed group on the live rat anastomosis in all expert-based measures (global ratings, checklist score, final product analysis, competency for OR; all P values <0.05).

Conclusions: Our current model of training surgical skills using short courses (for both CME and structured residency curricula) may be suboptimal. Residents retain and transfer skills better if taught in a distributed manner. Despite the greater logistical challenge, we need to restructure training schedules to allow for distributed practice.

(*Ann Surg* 2006;244: 400–409)

From the Department of Surgery, University of Toronto, Toronto, Ontario, Canada.

Supported by the PSI Foundation (from the physicians of Ontario) and the Royal Australasian College of Surgeons (Fellowship in Surgical Education).

Reprints: Carol-anne E. Moulton, Wilson Centre, 200 Elizabeth St., Eaton South 1-565, Toronto, Ontario, Canada, M5G 2C4. E-mail: Carol-Anne.Moulton@uhn.on.ca.

Copyright © 2006 by Lippincott Williams & Wilkins

ISSN: 0003-4932/06/24403-0400

DOI: 10.1097/01.sla.0000234808.85789.6a

Changes in the surgical training curriculum, precipitated by limitations to resident work hours,^{1,2} concerns over patient safety,^{3,4} and budgetary constraints in the operating room^{5,6} have compelled surgical educators to search for more effective and creative means of teaching surgical skills. Emerging technologies have also created a need for strategies to deliver technical education to surgeons already in practice. Consequently, laboratories dedicated to teaching technical aspects of surgical skill have become increasingly popular worldwide. These laboratories deploy an increasing array of training models, both low and high fidelity, many of which have been validated as effective teaching tools.^{7,8} Much work has been done on the development of performance metrics of assessment that can chart the progress through training regimens, demonstrate effects of intervention, and differentiate between varying levels of experience.^{9,10} Recently, there has been an important focus on ensuring that skills taught in laboratory environments can translate to the real world of human operations.^{11,12}

Despite the important work on model development, simulators, and assessment methodologies, there is needed effort in the realm of curricular development.¹³ The ways in which the training is delivered (how, what, when, and how often) are equally important. For example, in many programs, junior residents learn a particular skill 1 week and return the next week to learn a different skill, with very little opportunity structured within the curriculum to rehearse what has previously been taught. Likewise, many CME programs offer courses that are short and intensive (1-day or weekend), not allowing for rehearsal of skills after a period of rest (and forgetting) has elapsed. Commonly, there is a delay between the time the skill is learned in the laboratory and the time the skill is needed in the operating room. Performance may be adequate immediately following training, but how much is retained is uncertain. The ultimate value of skills courses should be measured not by performance immediately after training, but by performance after a time delay, preferably in a realistic setting.

Drawing upon the motor skill learning principle of massed versus distributed practice found in the domains of psychology and athletics, there is good evidence that practice interspersed with periods of rest (*distributed practice*), leads

to better acquisition and retention of skill compared with practice delivered in continuous blocks with little or no rest in between (*massed practice*).^{14–16} How these principles transfer to the domain of surgical skill acquisition have yet to be demonstrated. This randomized controlled trial was designed to assess the impact that scheduling of practice (massed versus distributed) has on surgical skill acquisition. The study was designed to have maximal relevance to current residents' surgical skills courses and curricula. Using previously validated outcome measures and microsurgery as the technical skill domain, this study evaluates not only the *immediate* impact of each training schedule, but more importantly, the *durability* of skill acquisition ("retention" testing 1 month later) and the clinical *transferability* of the skill to a more realistic setting (testing on live anesthetized rats).

METHODS

Participants

Thirty-eight surgical residents (33 male, 5 female) in postgraduate training years 1 (n = 16), 2 (n = 15), and 3 (n = 7) volunteered to participate in the study. Specialties included general surgery (n = 9), urology (n = 6), orthopedics (n = 12), plastic surgery (n = 3), otolaryngology (n = 4), neurosurgery (n = 3), and cardiac surgery (n = 1). Thirty-six participants were right-handed, and 2 were left-handed.

Each participant signed informed consent approved by the local institutional research ethics board. All participants completed a questionnaire to determine baseline demographic characteristics, level of surgical training, and previous exposure to microsurgery. Each were stratified according to their postgraduate year level, and then randomly assigned to one of 2 experimental groups: massed training (n = 19) or distributed training (n = 19). All animals used in this project were covered under a protocol approved by the SLRI's Animal Care Committee and all animal use falls under the guidelines of the Canadian Council on Animal Care and the Ministry of Agriculture and Food. All animal work in the project was carried out by certified Animal Health Technologists.

Training

The massed group received 4 training sessions in 1 day and the distributed group received the same 4 training sessions, 1 session per week, over 4 weeks. In the first practice session, all participants, after watching a video on the principles of microsurgery, practiced microsuturing on a slit in a Penrose drain. In the second session, participants watched a demonstration video and practiced microvascular anastomoses using a 2-mm PVC (polyvinyl-chloride) artery model. During the third and fourth sessions, participants practiced microvascular anastomoses using arteries of a turkey thigh.^{17,18} Experts in the field of microsurgery were available to each participant (ratio 1 surgeon to 4–5 residents)¹⁹ at every session for informal teaching and feedback. Each group received equal time in training (330 minutes total training time, 155 hands-on practice time).

Testing

All participants had the opportunity to orient themselves to the surgical microscope and practice basic microsurgical exercises on microsurgical practice cards prior to the pretest. The *microsurgical drill* was used for all testing throughout the training phase. This drill has previously been validated as a measure of microsurgical skill.^{20,21} It requires the participant to pass 2 interrupted sutures (size 9/0) through a slit in synthetic tissue (Penrose drain) and tie a square surgeons' knot, followed by 2 additional square knots. Testing throughout the training phase was performed at the beginning and end of each training session (8 tests in total) for the purposes of defining the skill acquisition curve. The pretest, to assess participants' basic level of skill, is defined as the first test at the beginning of the first training session and the post-test is defined as the final test of the fourth training session. The retention test (using the microsurgical drill) and the rat transfer test were performed 1 month following the last training session of each participant. For the rat transfer test, each rat aorta was dissected out, cleaned, clamped, and transected by one of the investigators (C.E.M.). The participant then performed an infra-renal aortic anastomosis on the live anesthetized rat. Performing the anastomosis demanded an understanding of tissue fragility, and respect for surrounding tissues. Pulsations of the vessel throughout the anastomosis served to increase the mental and technical challenges of microsurgery. Transferability of the microsurgical skills acquired in the training phase to a more realistic setting (the ultimate goal of any training program) was thereby assessed.

Performances for each test were videotaped through a side port in the microscope to allow for assessment at a later date by 2 expert raters *blinded* to group, PGY level, and surgical specialty. To control for any extra experience obtained during the study period, all participants, returning for their retention test, were asked to fill in a questionnaire regarding microsurgical exposure since the start of trial.

Outcome Measures

Outcomes for pretesting, post-testing and retention testing (microsurgical drills) included 2 types of previously validated outcome measures: expert-based evaluations of performance (global ratings, competency rating, checklists, and final products)^{10,22–24} and computer-based evaluations of performance (time taken and motion efficiency).^{9,20} The acquisition of skill (performance curve) was assessed using computer-based measures only. Outcomes for the rat transfer test included computer-based measures, expert-based measures, and clinically relevant measures such as bleeding, patency, narrowing, and completion rates.

Expert-Based Evaluations of Performance

Videorecordings of performances were assessed by 2 blinded experts (trained in microsurgery) using checklists and global ratings scales adapted to microsurgery. The checklists were detailed, task-specific, dichotomous, 27-item (drill) or 32-item (rat) evaluation instrument whereby one score was given for each correctly performed step in the procedure. The global ratings scale consisted of 5 items; each rated on a behaviorally anchored 5-point scale. Comple-

tency was defined as the level with which the surgeon would allow the resident to perform microsurgery (supervised) in the operating room, based on the residents' performance on the videotapes. This was assessed as a global rating from 1 to 5.

Computer-Based Measures

Time to task completion was calculated from the moment the needle was grasped in the needle driver until both ends of the final stitch were cut. Hand motion analysis was captured using the Imperial College Surgical Assessment Device device⁹ with sensors applied to the dorsum of the participants' index fingers (proximal phalanges). The parameter of interest collected in this study was the number of dominant hand movements (as performance improves, the number of hand movements decreases).

Clinically Relevant Outcome Measures

The clinically relevant outcome measures included *patency* of anastomosis (the identification of an arterial pulse by palpation distal to the anastomoses), *narrowing* of anastomosis (resistance to a standard size 25 olive tip dilator placed into the lumen of the vessel through a distal arteriotomy after the rat was killed), *bleeding* (initially defined as bleeding occurring after the removal of the clamp, but redefined as bleeding lasting for more than 3 minutes after the removal of the clamp), and *completion* (ability to finish anastomosis).

Statistical Analysis

The data were tested for goodness of fit to normal distribution (Kolmogorov-Smirnov), and given its non normality, a number of independent nonparametric tests (Kruskal-Wallis) were used to assess the impact of practice distribution on: 1) the performance curve of skill acquisition with group as the between subject factor (massed and distributed) and test as the within subject factor (tests 1–8) and 2) change of performance as a result of practice with group as the between-subject factor and test as the within subject factor (pretest, post-test, and retention test). Significance for all tests was set at $P < 0.05$. Subsequently planned comparisons using the Mann-Whitney U test were performed to address each objective. All clinically relevant outcome measures (yes/no) were evaluated using the χ^2 test. Interrater reliability between the 2 expert examiners was assessed by determining Cronbach's α coefficient.

RESULTS

Subject Demographics

No participants were excluded on the grounds of extensive prior experience with microsurgery, defined as having performed greater than 5 microsurgical cases as the primary surgeon (ie, >80% of the procedure). There was no attrition throughout the study. Time from post-test to retention test was not significantly different between groups: massed = 28 days (range, 21–43 days) and distributed = 29 days (range, 22–44 days); Mann Whitney $U = 153$, $P = 0.42$. There were

no significant differences in mean age: massed = 29, distributed = 28; $t(35) = 0.44$, $P = 0.66$, gender distribution: massed 17/19 males, distributed 16/19 males; $\chi^2(1) = 0.23$, $P = 0.63$, or level of training [postgraduate year level 1/2/3 ratio: massed 9/6/4, distributed 7/9/3; $\chi^2(3) = 0.99$, $P = 0.61$].

Interrater Reliability

Interrater reliability, as assessed by Cronbach's alpha, for the 2 expert examiners varied between 0.67 and 0.89 on all expert-based outcome measures.

Pretest Results

Expert and Computer-Based Measures

Pretest, post-test, retention test, and transfer test results are summarized in Figures 1 and 2. Outcome measures have been divided into expert-based measures (Fig. 1), computer-based measures (Fig. 2), and clinically relevant measures (below). No significant differences existed in any outcome measure between the 2 groups on pretesting.

Expert-Based Measures

There were no significant differences between the 2 groups on any of the 4 expert-based outcome measures on pretesting (global ratings scores $U = 158$, $P = 0.67$; checklist scores $U = 140$, $P = 0.34$; final product scores $U = 115$, $P = 0.08$; competency scores $U = 169$, $P = 0.69$) (Fig. 1).

Computer-Based Measures

Neither the microsurgical drill times, nor the number of dominant hand movements were significantly different between the 2 groups on pretesting ($U = 178$: $P = 0.94$ and $U = 136$: $P = 0.58$, respectively) (Fig. 2).

Immediate Post-Test Results

No significant differences existed in any outcome measure between the 2 groups on post-testing.

Expert-Based Measures

There were no significant differences between the 2 groups on any of the 4 expert-based outcome measures on immediate post-testing (global ratings scores $U = 111$, $P = 0.11$; checklist scores $U = 100$, $P = 0.13$; final product scores $U = 94$, $P = 0.13$; competency scores $U = 122$, $P = 0.19$) (Fig. 1).

Computer-Based Measures

Neither the microsurgical drill times, nor the number of dominant hand movements were significantly different between the 2 groups on immediate post-testing ($U = 175$: $P = 0.87$ and $U = 114$: $P = 0.45$, respectively) (Fig. 2).

Pre- to Post-Test Results

Both groups showed a significant improvement in performance from pretest to post-test (microsurgical drill) when assessed by global ratings, competency, time taken, and motion efficiency (Figs. 1 and 2). Only the distributed group demonstrated a significant improvement in performance us-

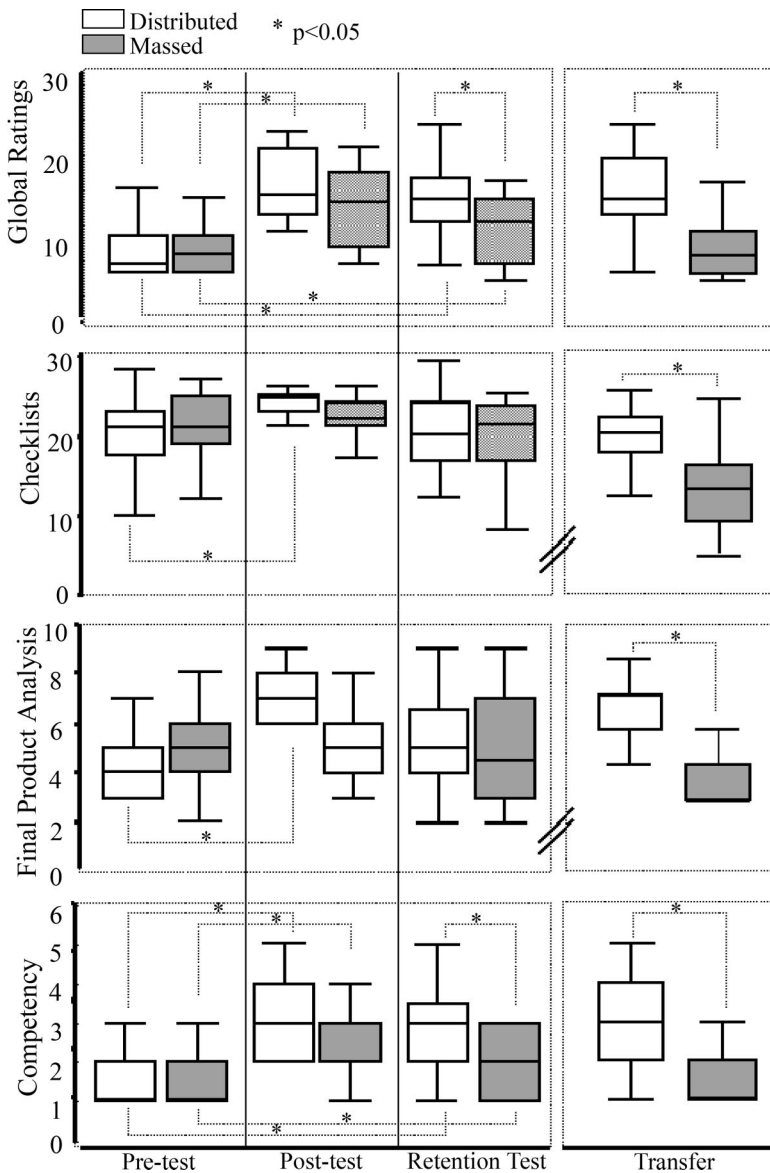


FIGURE 1. Box plots of all expert-based measures. The bar represents median, the box 25th to 75th percentile, and the whiskers the range of the data. Microsurgical drills (pre, post, and retention tests) and live rat (transfer) performances are plotted for global ratings, checklists, final product analysis, and competency for the distributed (clear) and massed (shaded) groups. Significant (set at $P < 0.05$) differences between tests and between groups are highlighted with an asterisk.

ing the checklist scores ($U = 82$, $P < 0.05$) and the final product analysis ($U = 65$, $P < 0.05$).

Acquisition Curves

The acquisition curves of both groups were calculated using the microsurgical drills conducted immediately before and after each of the 4 training sessions. Computer-based metrics were plotted against test (test 1–8). The inverse function accounted for the most variance ($R^2 = 0.69$ – 0.76) of the performance curves and was thus used for data analysis. All participants displayed improvements in time and number of movements as evidenced by participants' acquisition curves (Fig. 3). There were no significant differences with the rate of skill acquisition (slope of curve) between the 2 groups in either dominant hand movements ($U = 138$, $P = 0.62$) or time ($U = 127$, $P = 0.83$). Similarly, no significant differences existed between the 2 groups on the eventual

plateau (asymptote) of skill level on either the number of dominant hand movements ($U = 146$, $P = 0.82$) or mean time ($U = 129$, $P = 0.88$).

Retention and Transfer Tests

Microsurgical Drill

One month following their last training session, each resident returned to perform the microsurgical drill and a live rat anastomosis. Again, expert-based and computer-based outcome measures were used to evaluate performances.

Expert-Based Measures

A significant difference in performance was noted on the retention test (drill) between the 2 groups when assessed with global ratings and competency (Fig. 1). Learning (significant improvement between pretest and retention test) was

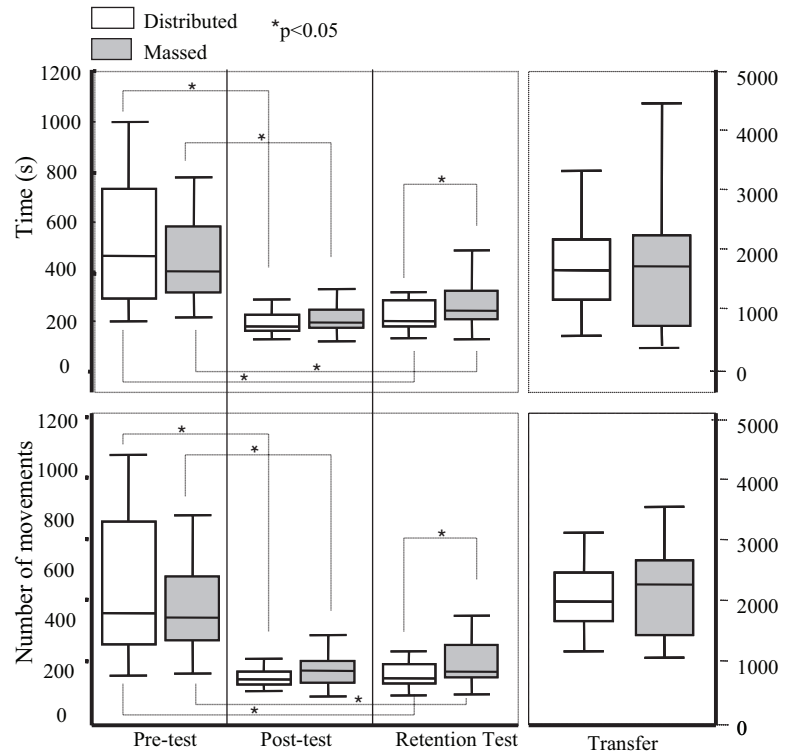


FIGURE 2. Box plots for both computer-based measures. The bar represents median, the box 25th to 75th percentile, and the whiskers the range of the data. Microsurgical drills (pre, post, and retention tests) and live rat (transfer) performances are plotted for time and number of dominant hand movements for the distributed (clear) and massed (shaded) groups. Significant (set at $P < 0.05$) differences between tests and between groups are highlighted with an asterisk.

evident in both groups when assessed by global ratings and competency. Checklist scores and final product analyses scores failed to demonstrate similar or consistent patterns. Although the distributed group demonstrated significantly improved performance on both scores in their post-test, there was no significant improvement on retention testing (checklist $U = 173$, $P = 0.83$; final product $U = 121$, $P = 0.12$).

Computer-Based Measures

A significant difference between the 2 groups were demonstrated for both time and number of movements, with the distributed group being superior (Fig. 2). Learning was evident in both groups with a significant improvement in performance from the pretest to the retention test for both computer-based measures.

Transfer Test (Rat Anastomosis)

Expert-Based Measures

On the transfer test, the distributed group outperformed the massed group on *all* expert-based outcome measures (Fig. 1).

Computer-Based Measures

No significant differences were demonstrated between the 2 groups in the transfer test for either time taken or number of dominant hand movements (Fig. 2).

Clinically Relevant Outcome Measures

Various clinically relevant outcome measures were used for the transfer test. The entire distributed group completed the anastomosis, whereas 3 of the massed group tore

the vessel beyond repair and were unable to complete the anastomosis. This did not reach statistical significance ($\chi^2(1) = 3.257$, $P = 0.07$) but arguably could be considered of clinical importance given the significant morbidity that could result in a real-life setting. Other clinically relevant outcome measures were not significantly different between the 2 groups [bleeding: massed (9/19) and distributed (12/19), $\chi^2(1) = 0.958$, $P = 0.33$; narrowing: massed (10/16) and distributed (13/19), $\chi^2(1) = 0.135$, $P = 0.71$; patency: massed (14/16 patent) and distributed (16/19 patent); $\chi^2(1) = 0.077$, $P = 0.78$].

DISCUSSION

Worldwide, surgical skills laboratories are becoming an important venue for the technical training of surgical residents. Over the last decade, much has been learned regarding the acquisition of surgical skill in this environment.^{7–12} However, the optimal timing and means of acquiring and retaining surgical skills to provide maximal transfer to the operating room requires further study. In this regard, much can be learned from other domains such as athletics and psychology, wherein motor learning has been studied more extensively resulting in the generation of current theories (ie, schedule of practice, learning specificity, schedule of feedback, etc.) on motor skill acquisition. Recently, surgical educators have begun testing these theories to see how well they apply to surgical skill acquisition.^{25–30} Surgical skill may be more complex than other motor skills, in part because of the greater degree of cognitive involvement than many other skills. Therefore, assumptions about how well existing motor learn-

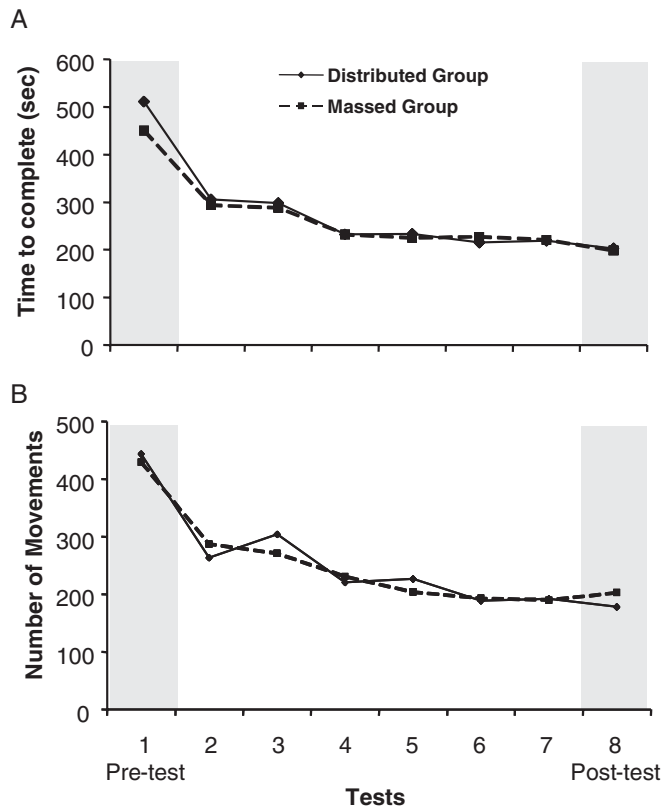


FIGURE 3. Learning curve of outcome measure (A, time to complete drill; B, number of dominant hand movements) versus test for both groups. Shaded areas represent the tests that were subsequently used as pre and post tests. There were 4 training sessions, and each participant was tested on the microsurgical drill immediately before and after each session.

ing principles apply to surgery cannot be confidently made without testing them in the surgical skills arena.

This study assessed the impact of practice distribution (massed and distributed training schedules) on surgical skill acquisition in the laboratory. Presently, we are teaching in a predominantly massed curriculum, but criticism regarding its effectiveness lingers.^{13,31} Mackay et al have previously addressed the question of massed and distributed practice.²⁵ Novices, performing the “transfer place” task on the MIST-VR, practiced for either 20-minute blocks with no rest (massed) or 5-minute blocks interspersed with 2.5-minute periods of rest (distributed). Unfortunately, all participants became very proficient at the skill and, as a result, the applied metrics were not capable of detecting differences in performance. This made interpretation difficult, although results favored the distributed group (median test).

Our study design was specifically chosen to test a skill that is often taught under massed practice conditions. Additionally, we chose to study surgical residents, a group who are not rank novices in surgical skills but who are yet to be considered experts. The results of this study demonstrate that surgical skill, acquired under distributed practice conditions, is more robust and shows superior transferability to a lifelike model than surgical skill acquired under massed practice

conditions. Retention of skill and transferability of skill, rather than skill level immediately following training, should be the ultimate aim of any surgical skills curriculum.^{15,32,33}

Impact of Practice Scheduling on Performance During Skills Training

Residents in massed and distributed practice groups performed at similar levels immediately after training, with no differences identified in either the rate of skill acquisition or the eventual plateau of skill level. In other domains, individuals practicing in a distributed practice schedule often outperform those practicing in a massed practice schedule immediately after training.^{14,34} Predominantly, this is attributed to a phenomenon called reactive impedance, where the performance is hindered by factors such as fatigue and boredom. Once the reactive impedance has dissipated (following a period of rest), the group that practiced under the massed regimen improves (reminiscence), though usually not to the performance level of the distributed group.^{16,34} We did not see this in this study, and one might infer that reactive impedance was not a factor in the massed learning approach to this surgical skill.

There are several possible explanations for this finding. Skills studied in the psychological literature often involve repetitive movements of limited numbers of muscle groups (rotor pursuit, ball tossing) leading to fatigue. In the present study, refreshment breaks for the massed group were mandatory between each training session. This was intended to better simulate what would occur in surgical skills courses in real life where it is not the intention to bore or fatigue the participants. Another plausible factor that may explain the lack of reactive impedance was the relative simplicity of the test (microsurgical drill) compared with the functionally more complex practice model of the last 2 sessions (turkey thigh artery). It is possible that, if the test model was of equal functional difficulty to the practice model, the expected reactive impedance effect would have been present.³⁵

Impact of Practice Scheduling on Retention and Transfer Tests

Performance during practice or training sessions may not be the best indicator of true learning. Increasingly, attention is being paid to evidence of retained knowledge and skill; and importantly, the transfer of knowledge from an artificial environment to real world settings.³³ *Retention tests* attempt to remove the effects of temporary modulators on performance, such as fatigue, and rely on the retrieval of skill from memory. It is therefore a better indicator of learning than performance during practice trials. *Transfer tests* are important in assuring that what has been learned in the acquisition of a new skill bears direct relevance to a real task. Many studies that have been used to assess the efficacy of a new technical skill learning tool, for example, the use of virtual reality simulators, have been criticized with regards to lack of significance. There is the often heard criticism about some studies that all they demonstrate is training on a simulator makes you better at performing on a simulator. To mitigate this potential criticism, we took the added step of testing performance in a live animal model, and feel that the finding

for superiority of distributed training is strengthened by the finding of an effect in this ultimate outcome task.

Although both groups demonstrated significant improvement in skill as a result of the training regimen, the distributed group significantly outperformed the massed group on the retention test, suggesting increased robustness of skill learned under distributed practice conditions. The significant difference between the 2 groups on the transfer test, when other factors such as bleeding, vessel fragility, and rat movements added to the functional complexity and difficulty of the skill, adds further support to the increased robustness of skill learnt under distributed conditions.³⁵

Why Is Distributed Practice Superior?

Superiority of distributed practice over massed practice has been shown in many domains. Explanations for why this occurs have centered on the concept of consolidation of learning.³⁶ It refers to 2 phases that occur during the learning of a new skill or behavior: within the session and between the session. During each phase, it is thought that different brain regions become activated, each considered necessary for the relative permanent retention of that skill. In addition to this distributed practice allows for cognitive preparation and mental rehearsal both considered key factors in the performance of surgical procedures.^{37–39} Having to retrieve from memory (with each practice session), key aspects of the skill being learned, more deeply encodes the particular skill into memory.

Outcome Measures

The expert-based performance measures used both task specific checklists and global rating forms. These 2 approaches are fundamentally different, with checklists focusing on surgical maneuvers, and global scales focusing on surgical behaviors. In this study, global ratings seemed to be a more accurate measure of surgical performance than checklists. Improvements in checklist scores did not seem to follow any consistent pattern. Our group has shown that checklists do not show as high concurrent or construct validity as global ratings.²² Checklists may not accurately reflect the true value of an item if not weighted adequately. Very important items may be given only one point, and numerous less important items when added up may contribute significantly to the overall score. In addition, as a learner matures, the psychometric properties of checklists falter. It is likely that, with experience, comes differing ways of accomplishing the same goal. A checklist approach to scoring will reward only one way, and not allow the examiner the leeway of rewarding good, but different ways of accomplishing a task. Checklists can be of value as an independent outcome measure, as they help to keep the examiner focused and ensure the observation of the entire task. As well, in this study, the checklists were able to demonstrate significant differences between the 2 groups on the rat anastomosis, consistent with the other outcome measures. This could be a result of better weighting of individual items on the anastomotic checklist compared with the microsurgical drill checklist, or it could be that the increased difficulty of the anastomotic task was better able to differentiate between the 2 skill levels.

Similar to the inconsistencies seen in the checklist scores, final product analysis scores did not follow the patterns of most other outcome measures when assessing the microsurgical drills. As with the checklists, the only significant difference found was between the pretest scores and the post-test scores for the distributed group. Again, as with the checklists, the final product analysis demonstrated a significant difference between groups on the rat anastomosis. Previous work on final product analyses has not shown the same degree of consistency or validity as the global rating scores and motion efficiency.^{23,24}

Although both global ratings and hand motion analysis were able to detect significant differences between the 2 groups on the retention test, only global ratings were able to detect differences on the transfer test. The rat anastomosis was more complex, with many confounding factors (number of sutures, rat movement, time to perform the anastomosis) influencing the performance metrics of time and number of dominant hand movements. We provided guidance for the rat anastomosis in terms of how many sutures should be placed but left it to the discretion of the surgical trainee to ultimately choose the exact number in an effort to emulate real life. The exact number chosen was obviously influenced by the spacing of the sutures and the quality of the sutures. Some sutures, not considered adequate, were removed. This therefore leads to a wide variation in time taken to complete the procedure and number of hand movements. A recent publication highlights the limitations of using hand motion analysis for more complex, whole procedures, rather than discrete tasks.⁴⁰

Clinically relevant outcome measures did not prove helpful in distinguishing between the 2 groups in this study. None of the measures had previously been validated in other studies but were included in an attempt to measure what is considered clinically important when performing vascular anastomoses. Lessons learned from this study would change our practice for future studies of this nature. The size of the needle, adequate for the training sessions, was too large for the 1- to 2-mm diameter aorta in the rat, resulting in bleeding in *all* rat anastomoses through the needle holes. Bleeding was then redefined as bleeding lasting more than 3 minutes after the clamp was removed (manual compression with gauze was performed and bleeding was checked for at 30 second intervals). Other factors that may influence bleeding, such as rat body temperature and amount of previous bleeding during the dissection of the aorta, would need further exploration. Narrowing, as tested by an olive tip dilator, similarly had not been previously validated and may be prone to subjectivity. Patency did not demonstrate differences between the 2 groups, possibly because of the small numbers involved. The difference between groups in completion rates did not quite reach statistical significance ($P = 0.07$), although the authors felt this difference to be of some clinical import. A vessel torn beyond repair during a microvascular anastomosis in a human could result in serious morbidity.

This study specifically assesses the acquisition of a novel surgical skill in junior residents and favors distributed practice regimens. Although, theoretically, the same principle should be applicable to surgical skill acquisition in the more

experienced surgical population (eg, CME), verification of this with further research would add support to our findings.

Implications for Clinical Practice

The results from this study suggest that programs using a massed method of teaching a new skill, which at present predominates in skills courses, should undergo curricular reform to allow for distributed practice. CME courses for ongoing surgical development and training are organized to enhance participation by busy surgical personnel, and thus are also predominantly conducted using massed practice schedules. Creative ways of adopting a more distributed regimen so that the program remains feasible need to be explored.

As we continue to develop and encourage the model of more structured training outside the operating room, we must vigilantly scrutinize the methods we use to do this. One advantage of teaching surgical skills in this environment is the application of educationally (pedagogically) sound theories to enhance skills learning. As technical skills training continues to evolve in surgical education, it is important to incorporate those methodologies shown to enhance motor learning into the curricula. It is mandatory that time spent away from clinical duties, especially in an increasingly shortened training program, be most effective in the acquisition of transferable surgical skill. Furthermore, in an era of increasingly scarce healthcare resources, it will become imperative to develop the most effective means of training future surgeons.

REFERENCES

- Mittal V, Salem M, Tyburski J, et al. Residents' working hours in a consortium-wide surgical education program. *Am Surg*. 2004;70:127-131.
- Reznick RK, Folse JR. Effect of sleep deprivation on the performance of surgical residents. *Am J Surg*. 1987;154:520-525.
- Saied N. Virtual reality and medicine. From the cockpit to the operating room: are we there yet? *Mo Med*. 2005;102:450-455.
- Brewster LP, Risucci DA, Joehl RJ, et al. Management of adverse surgical events: a structured education module for residents. *Am J Surg*. 2005;190:687-690.
- Mitton C, Donaldson C, Shellian B, et al. Priority setting in a Canadian surgical department: a case study using program budgeting and marginal analysis. *Can J Surg*. 2003;46:23-29.
- Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. *Am J Surg*. 1999;177:28-32.
- Hamstra SJ, Dubrowski A. Effective training and assessment of surgical skills, and the correlates of performance. *Surg Innov*. 2005;12:71-77.
- Grober ED, Hamstra SJ, Wanzel KR, et al. The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. *Ann Surg*. 2004;240:374-381.
- Datta VK, Mackay SD, Mandalia M, et al. The comparison between motion analysis and surgical technical assessments. *Am J Surg*. 2002.
- Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84:273-278.
- Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg*. 2002;236:458-463.
- Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg*. 2004;91:146-150.
- Anastakis DJ, Wanzel KR, Brown MH, et al. Evaluating the effectiveness of a 2-year curriculum in a surgical skills center. *Am J Surg*. 2003;185:378-385.
- Lee TD, Genovese ED. Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. *Res Q Exerc Sport*. 1988;59:277-287.
- Schmidt RA, Bjork RA. New conceptualization of practice. *Psych Sci*. 1992;3:207-217.
- Donovan JJ, Radosevich DJ. A meta-analytic review of the distribution of practice effect: now you see it, now you don't. *J Appl Psychol*. 1999;84:795-805.
- Hino A. Training in microvascular surgery using a chicken wing artery. *Neurosurgery*. 2003;52:1495-1497.
- Galeano M, Zarabini AG. The usefulness of a fresh chicken leg as an experimental model during the intermediate stages of microsurgical training. *Ann Plast Surg*. 2001;47:96-97.
- Dubrowski A, MacRae H. Randomised, controlled study investigating the optimal instructor: student ratios for teaching suturing skills. *Med Educ*. 2006;40:59-63.
- Grober ED, Hamstra SJ, Wanzel KR, et al. Validation of novel and objective measures of microsurgical skill: hand-motion analysis and stereoscopic visual acuity. *Microsurgery*. 2003;23:317-322.
- Starkes JL, Payk I, Hodges NJ. Developing a standardized test for the assessment of suturing skill in novice microsurgeons. *Microsurgery*. 1998;18:19-22.
- Regehr G, MacRae H, Reznick RK, et al. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Acad Med*. 1998;73:993-997.
- Szalay D, MacRae H, Regehr G, et al. Using operative outcome to assess technical skill. *Am J Surg*. 2000;180:234-237.
- Datta V, Mandalia M, Mackay S, et al. Relationship between skill and outcome in the laboratory-based model. *Surgery*. 2002;131:318-323.
- Mackay S, Morgan P, Datta V, et al. Practice distribution in procedural skills training: a randomized controlled trial. *Surg Endosc*. 2002;16:957-961.
- Rogers DA, Regehr G, Howdieshell TR, et al. The impact of external feedback on computer-assisted learning for surgical technical skill training. *Am J Surg*. 2000;179:341-343.
- Rogers DA, Regehr G, Yeh KA, et al. Computer-assisted learning versus a lecture and feedback seminar for teaching a basic surgical technical skill. *Am J Surg*. 1998;175:508-510.
- Dubrowski A, Xeroulis G. Computer-based video instructions for acquisition of technical skills. *J Vis Commun Med*. 2005;28:150-155.
- Dubrowski A, Backstein D, Abughaduma R, et al. The influence of practice schedules in the learning of a complex bone-plating surgical task. *Am J Surg*. 2005;190:359-363.
- Dubrowski A, Backstein D. The contributions of kinesiology to surgical education. *J Bone Joint Surg Am*. 2004;86:2778-2781.
- Rogers DA, Elstein AS, Bordage G. Improving continuing medical education for surgical techniques: applying the lessons learned in the first decade of minimal access surgery. *Ann Surg*. 2001;233:159-166.
- Grober ED, Hamstra SJ, Wanzel KR, et al. Laboratory based training in urological microsurgery with bench model simulators: a randomized controlled trial evaluating the durability of technical skill. *J Urol*. 2004;172:378-381.
- Dubrowski A. Performance vs. learning curves: what is motor learning and how is it measured? *Surg Endosc*. 2005;19:1290.
- Bourne LE, Archer EJ. Time continuously on target as a function of distribution of practice. *J Exp Psychol*. 1956;51:25-33.
- Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *J Mot Behav*. 2004;36:212-224.
- Luft AR, Buitrago MM. Stages of motor skill learning. *Mol Neurobiol*. 2005;32:205-216.
- Hall JC. Imagery practice and the development of surgical skills. *Am J Surg*. 2002;184:465-470.
- Bohan M, Pharmed JA, Stokes A. When does imagery practice enhance performance on a motor task? *Percept Mot Skills*. 1999;88:651-658.
- Whitley JD. Effects of practice distribution on learning a fine motor task. *Res Q*. 1970;41:576-583.
- Aggarwal R, Moorthy K, Darzi A. Laparoscopic skills training and assessment. *Br J Surg*. 2004;91:1549-1558.

Discussions

DR. AJIT K. SACHDEVA (CHICAGO, ILLINOIS): This is another landmark contribution to surgical skills education from Dr. Reznick's group in Toronto. The study enhances our comprehension of surgical education conducted in skills laboratories, an area that is receiving great attention in the current milieu of health care.

Over the past 8 years, a fair amount of work has gone into studying structured teaching, learning, and assessment of surgical skills and a variety of methods have been validated. Also, transfer of technical skills from skills laboratories to the OR has been demonstrated. However, insufficient attention has been focused on studying the mode of delivery of the educational content in surgery, using the structured approaches.

This elegant study clearly demonstrates that distributed learning involving sequenced sessions is more effective in retention and transfer of surgical skills learned in laboratory settings as compared to massed learning. Also, the authors point out that global ratings are more consistent than checklists in evaluating performance, a finding that has been reported previously by Dr. Reznick's group and other investigators as well.

The impact of these findings is very significant. Although the study was conducted within the context of residency education, I believe the greatest impact of these findings is on continuing education and continuous professional development of surgeons. This is because, during residency training, there are often opportunities for repetition in the learning process, and sequenced learning can occur either by design or by default. But when it comes to continuing education, as you saw the very telling video clip, the educational intervention is often an isolated event, sometimes offered over a weekend, and usually with little or no follow-up.

I have 3 questions for Dr. Reznick. First, how do you and your coauthors plan to redesign your graduate and continuing surgical education programs based on these important findings? Second, are you considering adding reinforcing and enabling strategies to this distributed learning model to further enhance the retention and transfer of the skills learned in laboratory settings? This has been shown to be a very valid strategy in continuing education. Finally, what should be the role of preceptoring within the context of this distributed learning model? Where do you see preceptoring fitting in?

DR. RICHARD K. REZNICK (TORONTO, ONTARIO, CANADA): Thank you, Dr. Sachdeva, for those kind comments and important questions. I will answer the first 2 together because I think that they are basically intertwined.

Right now, on the basis of this study, we are taking a very good look at redesigning our curriculum. At present, we have a curriculum where all of our junior residents come

every week for 4 weeks to learn fundamental surgical skills in the laboratory. But we are doing these as one-off sessions where they are coming for 2 or 3 hours and doing something technical on the bowel and then the next week they might be learning how to do something technical on another organ. So we are looking at redesigning our curricula so that they have time to come back to our laboratory over and over again for practice sessions that they can do either on their own or importantly, as your third question alludes to, with mentoring.

Our skills laboratory would not work if our faculty did not come. It is as simple as that. Faculty presence is vital to giving it the notion of skills training their stamp of approval, for teaching the skills, and lastly for role modeling. If we didn't have the faculty come on a repeated basis, it wouldn't work. So our faculty are regularly assigned to the skills lab. They do come, and we think that is why it might be working.

DR. OLGA JONASSON (CHICAGO, ILLINOIS): Thank you, Dr. Reznick, for all of your work over the years in this important area and in particular preparing residents to enter the operating room.

Many of you know the work that came about a year ago from Duke out of the hernia trials of Dr. Neumayer where it was shown that PGY 1 and 2 residents had a much greater recurrence rate despite the presence of an attending at their side during the entire open hernia repair than more advanced residents. Can you speculate on the application of your techniques and your learning laboratories in preparing residents to enter the operating room for the first time?

DR. RICHARD K. REZNICK (TORONTO, ONTARIO, CANADA): I would stress that our skills laboratory should be viewed as, and I think all skills laboratories should be viewed as adjuncts of, not replacements for learning in the operating theater. All of us learned in the operating room, and it still remains the fundamental place to learn how to become a surgeon. This notwithstanding, given all the pressures that we all know about, adjunctive environments are needed.

My feeling is that the answer your question is of the use of regular testing. We now have performance curves of what a PGY 1 looks like, a PGY 2, a PGY 3, and there are significant differences at virtually every level for assessments of operations that we have tested. And we think that can help guide the readiness, if you will.

We are also using this exam for admission of foreign doctors who are already specialists and want to practice in our country. And we are seeing where they test out. For example, an individual might say they are practice ready. Well, if they test out at the level of a PGY 1 or a PGY 2 given our exam, we don't feel they are practice ready and we won't let them.

So I think the regular testing using a variety of different tools like the OSATS and other assessment tools will be the

route to attesting to the competence of trainees to perform actual procedures in the operating room.

DR. GERALD M. FRIED (MONTREAL, QUEBEC, CANADA): Dr. Reznick, that was a beautiful study and very well presented. I think it is important to emphasize a few of the things that you did that are particularly important in these types of educational studies. One of the most important is to demonstrate transferability of skills learned in the simulated environment to the real world. The other is demonstration of retention of the acquired skills over time. Most published studies of simulation training have not looked at retention.

That being said, even though it seems logical and intuitive that distributed learning would be superior, the magnitude of the benefit of distributed learning wasn't quite as great as one would hope to observe. Certainly in an environment where time in the simulation lab is hard to come by, we are all trying to find an educational model for training that is as efficient as possible. We are constrained by resident work hour rules, and for CME programs surgeons may not have the opportunity to have ongoing access to a simulation center for distributed learning.

My questions are: Can we come up with a process for massed training, which is usually easier to organize, that can achieve the same results as a program of distributed learning? Was the curriculum and the amount of time allocated to the 2 groups identical? That is, did both groups work on the same simulated tasks and for the same duration?

DR. RICHARD K. REZNICK (TORONTO, ONTARIO, CANADA): The answer to your last question is: absolutely. We critically organized the trial to have equivalence of time on every single task between the 2 groups. That was one of the most important issues to prove the point.

The outcome measure that we have the most faith in is the global rating form. We have been using it for 9 years, as Dr. Sachdeva said. We have done some studies that have shown the psychometric properties of global rating forms exceed the measures of many other assessment strategies. Using this rating, we found that in the transfer test, with the exception of 2 residents, 17 of the 19 residents didn't come

within 2 standard deviations of the residents who were in the distributed group. So for us, this is very convincing.

As I said, we are about to revamp our curricula. It is going to be very difficult. But it just makes sense that one-off courses, whether for CME or whether it be for technical skills learning in a surgical curriculum, are just not good enough, particularly as we increase the complexity of skills that we are teaching in our skills lab; when we are going from just knot-tying to doing microvascular anastomosis. We need to pay attention to that.

DR. MERRIL T. DAYTON (BUFFALO, NEW YORK): Dr. Reznick, that was a sophisticated and elegant study. I thought, as you were presenting that surgery, education has come a long way in the last 25 years.

The question that I have relates to the time interval between the last distributed session of training and the massed session of training. Was that time interval the same in both groups? Because were it not the same, you might simply be measuring recall rather than improvement in the system. Would you address that?

DR. RICHARD K. REZNICK (TORONTO, ONTARIO, CANADA): It was indeed the same. We took great pains to make sure that it was exactly the same time between their very last opportunity and the test. And the only other comment I would make is fortunately both you and I remember what surgical education was like 25 years ago.

DR. ALFRED CUSCHIERI (PISA, ITALY): I enjoyed this paper. My compliments on the study. There is, however, one crucial issue relating to proficiency in the profusion of a microvascular anastomosis. The 2 groups—was it the first time that they actually did an aortic microanastomosis?

DR. RICHARD K. REZNICK (TORONTO, ONTARIO, CANADA): Yes, for almost all the participants, it was their first time doing a microvascular anastomosis, I think it was for all but 1 or 2. And there was no difference between the 2 groups. These were PGY 1s through PGY 3s. They basically had had no experience in microvascular surgery.