Surgery 165 (2019) 1065–1068

Contents lists available at ScienceDirect

Surgery

journal homepage: www.elsevier.com/locate/surg

Performance variability during training on simulators is associated with skill transfer

John Rhodes Martin, MD[a,*], Nicholas Anton, MS[a,b], Lava Timsina, PhD[a], Jake Whiteside, BS[a], Erinn Myers, MD[b], Dimitrios Stefanidis, MD, PhD[a,b]

[a] Indiana University Department of Surgery, Indianapolis
[b] Carolinas Medical Center, Charlotte, NC

ARTICLE INFO

Article history:
Accepted 18 January 2019
Available online 18 March 2019

ABSTRACT

Background: Expert performance is characterized by consistency. The degree of variability of performance from repetition to repetition during proficiency-based simulator training could potentially indicate acquisition of expertise. We hypothesized that learners with less variability in performance during simulator training would achieve greater performance at the end of training and improved transfer of skills to a live, anesthetized, porcine model.

Methods: The performance of 93 subjects (surgery residents and medical students) who had participated in 3 randomized controlled trials was analyzed for variability. All participants had trained in laparoscopic suturing on the Fundamentals of Laparoscopic Surgery (FLS) simulator. Their performance had been assessed on the simulator before (baseline) and after training (posttest) and on a live, anesthetized, porcine model (transfer test). We computed the coefficient of variations of suturing scores during training for each participant. Linear regression was used to assess whether variability in performance during training predicted posttest and transfer-test scores.

Results: Decreased practice variability in performance was associated with greater scores in posttests and transfer tests. For each percent decrease in variability performance, posttest scores increased by 3.8 points ($P < .001$) and transfer-test scores increased by 3.0 points ($P < .001$). Greater mean scores during practice were associated with greater scores on the transfer test ($P < .001$).

Conclusion: Decreased variability in performance during practice on simulators is associated with improved performance at the end of training and during transfer to a live, anesthetized, porcine model. These findings suggest that variability in performance during simulator training may be used to track the progress and readiness of a trainee for the clinical environment. Further studies are needed to verify the robustness of this potentially new metric of performance.

Introduction

Despite the proven efficacy of proficiency-based simulator training for improving operative performance, simulator-trained learners have failed to attain expert levels of operative performance in the operating room during transfer tests.[1–6] As a potential explanation for this finding, we speculated that the metrics of simulator performance may be insufficient to reliably determine when simulator learning is complete.[7] Other authors have also recognized the need to identify additional, reliable measures of expert performance for simulator training.[8–12]

Bryan and Harter[13] of Indiana University first described consistency as a mark of expert performance in 1897 while studying skill acquisition among telegraph operators. They measured variability in performance across multiple transmissions of the same message by an operator and made the following observation: "Variation is an inverse measure of skill. An operator can repeat the same action more exactly the more expert he is."[13] Subsequent studies of motor learning conducted in the twentieth century observed task-specific association between the improvement in skill and a decrease in variability of performance.[14–16] This association was more pronounced when learners engaged in deliberate, distributed practice.[17] In more recent times, Ericsson[18] and others[19,20] have elaborated on the importance of deliberate practice

Presented at the 2018 American College of Surgeons Surgical Simulation Conference.

* Reprint requests: John Rhodes Martin, MD, 545 Barnhill Drive, Emerson Hall 125, Indianapolis, IN 46202.
E-mail address: jrm32@iu.edu (J.R. Martin).

https://doi.org/10.1016/j.surg.2019.01.013
0039-6060/

in skill acquisition and mastery learning and the decreased variability in performance observed among experts.

For surgery trainees, the degree of variability of performance from repetition to repetition could serve potentially as an indicator of expert performance during proficiency-based simulator training. The model of Fitts and Posner[21] describes the acquisition of technical skill in three sequential stages and provides a framework for understanding the development of consistent performance that is characteristic of expertise.[22,23] The first stage is the cognitive stage, during which learners gain knowledge of the steps and mechanics of a task. The second stage is the associative stage, during which learners translate the knowledge of a task into the physical execution of a task. The associative stage involves performing motor actions and experiencing and interpreting sensory feedback to refine subsequent motor actions. We would expect the variability of performance to be greater during this stage, when musculoskeletal movements are initially erratic. With refinement of motion that comes with deliberate practice, learners may progress eventually to the final stage, the autonomous stage, in which psychomotor movements become consistent, automatic, and require minimal attention to execute. According to *The Cambridge Handbook of Expertise and Expert Performance*[20]:

> The lower variability of experts is consistent with the literature on motor-skill acquisition and deliberate practice. As individuals refine their skills over time, their coordination improves, significantly reducing the variability with which an action is executed, allowing performance to be more reliably reproduced with greater control.

It is also known that individual trainees make progress in skill acquisition at different rates.[12,24] Therefore, we hypothesized that trainees who demonstrate less variability in performance while engaging in deliberate practice on a surgical simulator would achieve greater performance at the end of an allowed training period and would demonstrate improved transfer of skills to a live, anesthetized, porcine model.

Methods

The data concerning laparoscopic suturing performance of trainees who had participated previously in 3 randomized controlled trials were analyzed to detect the variability of their performance.[25–27] Overall, 38 surgery residents and 55 medical students from 2 training institutions had enrolled voluntarily in the same proficiency-based laparoscopic suturing-skills curriculum.[28,29]

Participants completed a baseline questionnaire, detailing their demographic information and laparoscopic and simulator experience, and underwent a baseline performance assessment of laparoscopic skill, using the Fundamentals of Laparoscopic Surgery (FLS) peg transfer, pattern cut, and intracorporeal suturing tasks.[30]

Trainees then attended weekly, small group ($n = 2$ or 3) training sessions during a period of 3–5 months, depending on the participants' clinical availability. During these sessions, all participants received 45 minutes of FLS proficiency-based training on peg transfer, pattern cut, and intracorporeal suturing tasks. To reach proficiency on these tasks, participants were asked to meet published levels of expert performance (ie, 48 seconds without errors for the peg transfer task on 2 consecutive repetitions followed by 10 nonconsecutive repetitions; 98 seconds without cutting more than 2 mm outside of the marked circle for pattern cut task on 2 consecutive repetitions followed by 10 nonconsecutive repetitions; and 112 seconds without errors of accuracy or knot security for intracorporeal suturing on 2 consecutive repetitions followed by 10 nonconsecutive repetitions).[3–5,29] The performance of the participant on a task of laparoscopic simulator suturing was assessed at baseline (pretest) and at the end of the training period (posttest). Trainees were asked to place a suture and tie a surgeon's knot followed by 2 square knots. Two blinded raters assessed the performance of the participant during every testing session. Knots were tested for accuracy and security, and an objective performance score for each repetition was calculated using the following published formula: Performance score = cutoff time (600 s) − task completion time (in seconds) − [10 × accuracy error (mm outside the marked targets on Penrose)] − [10 × security error (where knot slippage = 1 and knot breakage = 2)].[3–5,29] A greater score indicates a better performance. Given the published expert performance level of 112 seconds without errors of accuracy or knot security, according to this score formula, a score of 488 points or greater was set as the expert-derived proficiency goal.

At the conclusion of the training period and after the posttest on the simulator, trainees were then assessed in a transfer test of intracorporeal suturing on a live, anesthetized, porcine model of fundoplication. The porcine models were prepared by an expert laparoscopic surgeon (D.S.) according to a standardized model for testing, which has been described elsewhere.[5] Trainees were asked to create a fundoplication laparoscopically by placing three interrupted, gastrogastric sutures at premarked target locations on the porcine stomach. Trainees were asked to secure these sutures in the same fashion as in the pretest and posttest by tying a surgeon's knot followed by 2 square knots. The same two blinded raters again assessed the performance of the participant according to the objective score formula for laparoscopic suturing described earlier in this report. Accuracy error for placement of the suture was measured in millimeters outside the marked targets.

Given that the training and testing protocols were the same across the three independent studies, data were pooled in an effort to increase sample size and evaluate the effect of variability of practice performance on skill transfer to a live, anesthetized, porcine model. Coefficient of variation (C_v) was used to determine the variability in practice performance for each trainee, as observed across all of an individual's practice repetitions during the training sessions. C_v is a ratio of standard deviation (σ) to mean (μ) and was calculated by dividing the standard deviation of the complete set of practice scores of each trainee by the mean of those practice scores ($C_v = \sigma/\mu \times 100$). For this study, C_v therefore represents a global measure of the variability of practice performance for each trainee. Linear regression analysis was carried out to determine whether a decrease in the variability of practice performance from the training sessions was associated with increased posttest and transfer test scores. In stratified analysis, linear regression was performed separately for the medical student group and the resident group to account for any effects related to participant differences.

Results

All trainees included in the analysis completed the baseline evaluation (pretest), posttest, and transfer test and participated in the training sessions. In addition, all trainees demonstrated improvement in their laparoscopic suturing performance from baseline to posttest and transfer test by participating in this curriculum, as reported elsewhere.[25] The total number of completed practice repetitions for the intracorporeal suturing task varied substantially from trainee to trainee (range = 17 to 63 repetitions, mean = 40 repetitions). Although 58 of 93 trainees (62%) (16 residents and 42 medical students) achieved the expert-derived intracorporeal suturing score at least once, only 15 trainees (16%) (4 residents and 11 medical students) reached the expert-derived

proficiency goals (2 consecutive + 10 nonconsecutive) for intracorporeal suturing during the allowed training period. The mean C_v was 21.3 ± 3.2 in this cohort compared with 28.77 ± 1.7 among trainees who did not reach the expert-derived proficiency goals ($P = .043$). Trainees who reached expert-derived proficiency goals during practice also scored greater on average on the posttest (429.2 ± 32.0 points) compared with trainees who did not achieve proficiency (320.1 ± 21.5 points; $P = .018$), but did not score greater on the transfer test. After controlling for differences in C_v, the authors observed that, as the number of practice repetitions completed during the training sessions increased, the more likely a trainee was to score better on the posttest. Each completed practice repetition was associated with a 2.9 ± 1.3 point increase ($P = .025$) in the posttest score; however, increasing the number of completed practice repetitions did not predict improved score on the transfer test.

Of note, for each percent decrease in variability in performance during practice, intracorporeal suturing scores increased by 3.8 ± 1.1 points ($P < .001$) on the posttest and by 3.0 ± 0.6 points ($P < .001$) on the transfer test. Among the residents, for each percent decrease in practice performance variability, intracorporeal suturing scores increased by 2.0 ± 0.7 points ($P = .009$) on the posttest and increased by 2.6 ± 0.7 points ($P < .001$) on the transfer test. In contrast, the variability in practice performance of the medical students did not demonstrate a predictive association with posttest or transfer test scores (1.0 ± 3.3; $P = .755$ and 0.8 ± 1.3; $P = .541$, respectively).

Among all trainees, those whose overall mean score from the training sessions was greater tended to score greater on the posttest and on the transfer test as well. For each unit increase in the mean practice score, the posttest scores were greater by 1.1 ± 0.3 points ($P < .001$) and the transfer-test scores were greater by 0.9 ± 0.2 points ($P < .001$). For each unit increase in the mean practice score, the mean posttest scores of the residents were greater by 0.6 ± 0.2 points ($P = .024$) on the posttest and greater by 0.6 ± 0.3 points ($P = .027$) on the transfer test.

Discussion

In this study, decreased variability in performance during simulator training was associated with better performance at the end of the practice period and during transfer to a live, anesthetized, porcine model. This finding was true among surgery residents and among all trainees combined (surgery residents and medical students). These findings suggest that the variability in practice performance during simulator training could be used potentially to track trainee progress as trainees ascend their learning curve and as an indicator of trainee readiness to transition to the operating room. Although proficiency-based training should only allow trainees to advance once they have achieved proficiency goals, the use of scores from posttests and transfer tests administered before all trainees reaching proficiency was purposeful and served to cross-cut individual learning curves—regardless of experience—and capture differences in variability in practice performance that might not have been seen had every trainee been allowed sufficient time to reach proficiency. Similarly, baseline differences among participants are problematic in case-control studies of the effectiveness of proficiency-based training, but the variability in practice performance is an individual-level metric. The aim of this study was to capture those individual-level differences and determine their association with posttest and transfer-test performance.

Detection of increased variability in practice performance early in training may help identify learners in need of more intense training and closer monitoring of their progress toward achieving proficiency. To achieve this, one proposal the authors of this study considered was to calculate C_v on a rolling basis during training for a most-recent set of practice repetitions. This proposal that needs further implementation and study.

One explanation for why variability in performance of a task decreases as trainees ascend their learning curve may be that as learners discover and accrue successful strategies to accomplish the task—they tend to experiment less with alternative strategies.[31] Therefore, increased variability in the early stages of skill acquisition may be a good thing for learning, but may also signify delayed recognition of the most effective and efficient strategies to accomplish the task. The results of this study appear to support the latter supposition because those whose practice performance variability was decreased at the end of the study period were able to reach greater performance levels in the posttest and on the transfer test. Further study will be needed to determine the optimal cutoffs in variability in practice performance that indicate preparedness to advance to the operating room and to know how such variability practice performance might be interpreted in conjunction with other performance metrics.

In proportion, more medical students than residents were able to achieve expert-derived proficiency goals during training. Surgery residents have claimed elsewhere that their clinical duties and lack of protected time are barriers to full participation in simulation training, which would limit opportunities for deliberate practice.[32] These claims may well have been the case in this study because the mean number of practice repetitions was only 29.5 among the resident group vs 47.1 repetitions among the medical student group. Despite these differences, increasing the number of completed practice repetitions was not associated with a greater score on the transfer test, nor was achieving proficiency associated with a greater score on the transfer test. It has been postulated that current metrics of performance are insufficient to detect when learning is complete, and for this reason, other studies have observed a decline in performance on transfer—the more demanding environment of the transfer test exposes incomplete learning.[7] The results of this study further support these claims.

Of note, in the stratified analysis, the variability in practice performance was not predictive of posttest or transfer test among medical students. In addition, although the mean practice scores were predictive of posttest and transfer-test scores for the resident group and for all trainees when grouped and analyzed together, this observation did not hold true for the medical students when analyzed separately. Medical students are in an earlier phase of training than surgery residents, and thus as a cohort, medical students may represent a more novice moment along their learning curve or different learning curves altogether. In addition, at their given stage of training, medical students may have undergone less self-selection than surgery residents, based on their innate or perceived technical abilities, although the authors must note that earlier research questions this notion of self-realized “weeding-out” among prospective surgery trainees.[33] In our study, given that the relationship of the mean practice score with the performance on the transfer test of medical students approached but did not reach significance ($P = .06$), the possibility of a type-II error is also entertained.

Repeated achievement of expert level performance across time has been shown to prevent skill decay, and successful proficiency-based curricula have relied on the concept of overtraining to maximize skill retention.[1,34] The recognition of this requirement of expertise is reflected in the authors’ own challenge to trainees to meet expert-derived benchmarks in 2 consecutive + 10 nonconsecutive repetitions. Yet, the authors also acknowledge that this recommended number of repetitions is arbitrary. The 2 + 10 paradigm does not account for the considerable variation in psychomotor abilities among individual

trainees. For example, Grantcharov and Funch-Jensen[12] tested 37 surgery trainees on a laparoscopic, virtual-reality trainer. Trainees repeated 6 basic skills during 10 distributed repetitions. From the performance data, the authors described 4 groups of trainees: those who were innately proficient, those who improved to achieve proficiency, those who improved but did not achieve proficiency, and those who performed poorly and failed to demonstrate improvement. It remains unresolved whether this last subset of trainees is indeed untrainable or whether this subset requires a more substantial or dedicated investment in training. In either case, looking at the plots of the learning curves from the study of Grantcharov and Funch-Jensen,[12] the greatest variability in performance for the repetition to repetition for "error score" and "economy of motion score" appeared to occur within the poorest performing group.

Limitations of our study include its retrospective nature and the combination of data from multiple studies occurring during different timeframes. Although training and testing protocols were the same across studies, other unknown and unaccounted factors may have influenced participant performance.

In conclusion, the findings of this study suggest that the variability in scores of practice performance could be used to track trainee progress and readiness for transfer from the simulator. Additional studies are needed to further validate these findings and to determine whether the variability in practice performance is associated with readiness to perform in the operating room.

Funding

This work was funded by the Agency for Healthcare Research and Quality (AHRQ) [grant number R18HS022080]. The AHRQ was not involved in any element of the study design, manuscript preparation, or in the decision to submit the article for publication.

Disclosures

The authors report no disclosures or financial conflicts of interest.

References

1. Stefanidis D, Acker C, Heniford BT. Proficiency-based laparoscopic simulator training leads to improved operating room skill that is resistant to decay. *Surg Innov*. 2008;15:69–73.
2. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—A randomized controlled trial. *Am J Surg*. 2010;199:115–120.
3. Stefanidis D, Korndorffer Jr JR, Markley S, Sierra R, Heniford BT, Scott DJ. Closing the gap in operative performance between novices and experts: Does harder mean better for laparoscopic simulator training? *J Am Coll Surg*. 2007;205:307–313.
4. Korndorffer Jr JR, Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J Am Coll Surg*. 2005;201:23–29.
5. Prabhu A, Smith W, Yurko Y, Acker C, Stefanidis D. Increased stress levels may explain the incomplete transfer of simulator-acquired skill to the operating room. *Surgery*. 2010;147:640–645.
6. Dawe SR, Pena GN, Windsor JA, et al. Systematic review of skills transfer after surgical simulation-based training. *Br J Surg*. 2014;101:1063–1076.
7. Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD. Simulator training to automaticity leads to improved skill transfer compared with traditional proficiency-based training: a randomized controlled trial. *Ann Surg*. 2012;255:30–37.
8. Sadideen H, Alvand A, Saadeddin M, Kneebone R. Surgical experts: Born or made? *Int J Surg*. 2013;11:773–778.
9. Schaverien MV. Development of expertise in surgical training. *J Surg Educ*. 2010;67:37–43.
10. Alderson D. Developing expertise in surgery. *Med Teach*. 2010;32:830–836.
11. Zevin B, Aggarwal R, Grantcharov TP. Surgical simulation in 2013: Why is it still not the standard in surgical training? *J Am Coll Surg*. 2014;218:294–301.
12. Grantcharov TP, Funch-Jensen P. Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. *Am J Surg*. 2009;197:447–449.
13. Bryan WL, Harter N. Studies in the physiology and psychology of the telegraphic language. *Psychological Review*. 1897;4:27–53.
14. Stelmach GE. Individual differences and intra-individual variability in motor performance under continuous-practice conditions. *Hum Factors*. 1969;11:201–206.
15. Carron AV, Leavitt JL. Effects of practie upon individual differnces and intra-variability in a motor skill. *Res Q*. 1968;39:470–475.
16. Grose JE. Inter- and intravariability of motor performance. *Res Q*. 1967;38:570–575.
17. Stelmach GE. Distribution of practice in individual differences and intra-variability. *Percept Mot Skills*. 1968;26:727–730.
18. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med*. 2004;79(10 Suppl):S70–S81.
19. McGaghie WC, Kristopaitis T. Deliberate practice and mastery learning: Origins of expert medical performance. In: Cleland J, Durning SJ, eds. *Researching medical education*. New York, NY: John Wiley & Sons, Ltd; 2015:219–230.
20. Ward P, Williams AM, Hancock PA. Simulation for performance and training. In: Ericsson KA, ed. *The Cambridge handbook of expertise and expert performance*. New York, NY: Cambridge University Press; 2006:243–264.
21. Fitts PM, Posner MI. *Human performance*. Belmont, CA: Brooks/Cole; 1967.
22. Ericsson KA, Lehmann AC. Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annu Rev Psychol*. 1996;47:273–305.
23. Stefanidis D. Optimal acquisition and assessment of proficiency on simulators in surgery. *Surg Clin North Am*. 2010;90:475–489.
24. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Learning curves and impact of previous operative experience on performance on a virtual reality simulator to test laparoscopic surgical skills. *Am J Surg*. 2003;185:146–149.
25. Stefanidis D, Anton NE, Howley LD, et al. Effectiveness of a comprehensive mental skills curriculum in enhancing surgical performance: Results of a randomized controlled trial. *Am J Surg*. 2017;213:318–324.
26. Anton NE, Beane J, Yurco A, et al. Mental skills training effectively minimizes operative performance deterioration under stressful conditions: Results of a randomized controlled study. *Am J Surg*. 2018;215:214–221.
27. Anton NE, Bean EA, Hammonds SC, Stefanidis D. Application of mental skills training in surgery: A review of its effectiveness and proposed next steps. *J Laparoendosc Adv Surg Tech A*. 2017;27:459–469.
28. Peters JH, Fried GM, Swanstrom LL, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135:21–27.
29. Ritter EM, Scott DJ. Design of a proficiency-based skills training curriculum for the fundamentals of laparoscopic surgery. *Surg Innov*. 2007;14:107–112.
30. Fundamentals of laparoscopic surgery (FLS) Web site. http://www.flsprogram.org. Accessed April 6, 2017.
31. Alibali MW, Sidney PG. The role of intraindividual variability in learning and cognitive development. In: Diehl M, Hooker K, Sliwinski MJ, eds. *Handbook of Intraindividual Variability Across the Livespan*. New York, NY: Routledge; 2015:84–90.
32. Shetty S, Zevin B, Grantcharov TP, Roberts KE, Duffy AJ. Perceptions, training experiences, and preferences of surgical residents toward laparoscopic simulation training: a resident survey. *J Surg Educ*. 2014;71:727–733.
33. Panait L, Larios JM, Brenes RA, et al. Surgical skills assessment of applicants to general surgery residency. *J Surg Res*. 2011;170:189–194.
34. Stefanidis D, Korndorffer Jr JR, Markley S, Sierra R, Scott DJ. Proficiency maintenance: impact of ongoing simulator training on laparoscopic skill retention. *J Am Coll Surg*. 2006;202:599–603.