



# HHS Public Access

Author manuscript

*Surg Endosc.* Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

*Surg Endosc.* 2015 November ; 29(11): 3154–3162. doi:10.1007/s00464-014-4042-4.

## Deliberate Practice Enhances Quality of Laparoscopic Surgical Performance in a Randomized Controlled Trial: from Arrested Development to Expert Performance

Daniel A. Hashimoto, M.D., M.S.<sup>1,5</sup>, Pramudith Sirimanna, MBBS<sup>1</sup>, Ernest D. Gomez, M.D., M.S.<sup>2</sup>, Laura Beyer-Berjot, M.D.<sup>1</sup>, K. Anders Ericsson, Ph.D.<sup>3</sup>, Noel N. Williams, MBBCh, FRCSI<sup>2</sup>, Ara Darzi, KBE, PC, M.D., FMedSci, FRCS, FACS<sup>1</sup>, and Rajesh Aggarwal, MBBS, M.A., Ph.D., FRCS<sup>4</sup>

<sup>1</sup>Department of Biosurgery and Surgical Technology, St. Mary's Hospital, Imperial College London

<sup>2</sup>Perelman School of Medicine, University of Pennsylvania

<sup>3</sup>Department of Psychology, Florida State University

<sup>4</sup>Department of Surgery, McGill University

<sup>5</sup>Department of Surgery, Massachusetts General Hospital

### Abstract

**Background**—This study investigated whether deliberate practice leads to an increase in surgical quality in virtual reality (VR) laparoscopic cholecystectomies (LC). Previous research has suggested that sustained DP is effective in surgical training.

**Methods**—Fourteen residents were randomized into deliberate practice (n=7) or control training (n=7). Both groups performed 10 sessions of two VR LCs. Each session, the DP group was assigned 30 minutes of DP activities in between LCs while the control group viewed educational videos or read journal articles. Performance was assessed on speed and dexterity; quality was rated with global (GRS) and procedure-specific (PSRS) rating scales. All participants then performed five porcine LCs.

**Results**—Both groups improved over 20 VR LCs in time, dexterity, and global rating scales (all  $p < 0.05$ ). After 20 LCs, there were no differences in speed or dexterity between groups. The DP group achieved higher quality of VR surgical performance than control for GRS (26 vs. 20,  $p = 0.001$ ) and PSRS (18 vs. 15,  $p = 0.001$ ). For VR cases, DP subjects plateaued at GRS=25 after 10

---

CORRESPONDING AUTHOR: Rajesh Aggarwal, MBBS, MA, PhD, FRCS, Associate Professor of Surgery, Director, Arnold & Blema Steinberg Medical Simulation Centre, Faculty of Medicine, McGill University, 3575 Parc Ave, Suite 5640, Montreal, QC H2X 3P9, rajesh.aggarwal@mcgill.ca.

Preliminary results of this paper were presented at the Surgical Forum of the 2013 Annual Clinical Congress of the American College of Surgeons in Washington, DC, on October 9, 2013.

#### Disclosures

Drs. Hashimoto, Sirimanna, Gomez, Beyer-Berjot, Williams, and Darzi have no relevant financial disclosures.

Dr. Aggarwal is a consultant for Applied Medical. Dr. Ericsson receives royalties from the publication of a textbook on expertise and receives honoraria for lectures on the topic of expertise.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

cases and control group at GRS=20 after five cases. At completion of VR training, 100% of the DP group reached target quality of performance (GRS  $\geq$  21) compared to 30% in the control group. There were no significant differences for improvements in time or dexterity over five porcine LCs.

**Conclusion**—This study suggests that DP leads to higher quality performance in VR LC than standard training alone. Standard training may leave individuals in a state of “arrested development” compared to DP.

---

## Introduction

Current changes within modern delivery of health care have been driven by an increased awareness of patient safety [1]. Reports that surgical adverse events account for two-thirds of all adverse events are only surpassed by data suggesting that half are preventable [2]. The literature has demonstrated that technical faults account for a large majority of these errors [3, 4]. These facts, along with an increase in public and political expectations, have led to the development of strategies to improve surgical proficiency and effectiveness of training.

The dominant paradigm in surgical training has largely been based on an apprenticeship model in which learning occurs within the operating room with subjective assessment methods[5]. Virtual reality (VR) simulation has repeatedly been demonstrated in the literature to enhance surgical performance and reduce error [6–8], and training curricula on VR simulators have been developed allowing surgeons to train to a proficient level [9, 10]. Additional studies have demonstrated that other pedagogical techniques such as mental practice and stepwise training can enhance surgical performance [11, 12]. Despite the development of these promising strategies, implementation has been sparse thus far [13].

Currently, technical proficiency is based on completing a pre-defined number of cases as outlined by the ACGME [14]; yet, these targets are not necessarily based on evidence that suggests sufficient caseload to guarantee superior performance. The non-medical literature has attempted to quantify the number of procedures one has to perform in order to be elevated to “expert” status [15]. However, Ericsson has suggested that those who consistently exhibit superior performance, rather than those with the most experience, should be identified as experts and that the acquisition and maintenance of expert performance requires sustained Deliberate Practice (DP) [15].

DP is based on the theory that expert performance results from deliberately engaging in and choosing activities that improve and maintain high performance. It involves repeated practice on tasks and immediate feedback on performance that allows individuals to focus their training on weaknesses while also refining other aspects of performance. In other domains, such as music, chess and sport, the achievement of expert performance is related to the extent of DP performed [16].

DP has been shown to improve the quality of surgical technical skills when compared to standard training [17]; however, this initial study was designed only to assess whether or not DP improves performance. While the results are able to demonstrate a learning curve for quantitative metrics of time and dexterity, methodological limitations prevented an assessment of surgical quality at every operative case performed. The lack of data prevented

an accurate determination of rate of learning for quality of performance. The current study investigated whether deliberate practice can result in improved performance in VR laparoscopic cholecystectomy (LC) as measured by consistency of performance, whether such skills transfer to tissue in the form of porcine LC, and how DP affects the rate of improvement and attained skill level in surgical quality.

## Methods

### Participant Selection

Twenty (n=20) junior residents from various London training programs were recruited to participate. All trainees had limited laparoscopic surgical experience (performed 0 but observed > 1 LCs in the operating room). At recruitment, participants were randomized into one of two groups - Deliberate Practice group or Control group – using a random number generator (STATA, College Station, TX).

### Baseline Assessment of Laparoscopic Skills

Each participant underwent a validated baseline skills assessment on the LapMentor VR (Symbionix; Cleveland, OH) laparoscopic simulator on Basic Skills tasks 5 (Clip and Cut) and 6 (Two Handed Maneuvers) [10].

### Virtual Reality Laparoscopic Simulator

The LapMentor VR laparoscopic simulator (Symbionix Corporation; Cleveland, OH) and the LapSim VR laparoscopic simulator (SurgicalScience, Inc.; Göteborg, Sweden) were used for this study. Four procedural tasks (clip and cut – retracted gallbladder, clip and cut – 2 hands, Calot’s triangle dissection, and gallbladder fossa dissection) and a full procedure LC were utilized on the LapMentor. Four basic tasks (Clip and Cut, Lift and Grasp, Grasping, Coordination) were utilized on the LapSim. Construct validity has previously been demonstrated for each of the tasks on these simulators [10, 18].

### Didactic and Proficiency Training

Participants underwent a validated laparoscopic skills training program as previously described [10].

### Deliberate Practice Group Training Sessions

Participants in the DP group underwent 10 VR training sessions comprising a total of 20 VR LCs. At the beginning of each session, participants completed a VR LC on the LapMentor simulator, and their performance was assessed in real time by a qualified observer using two previously validated rating scales of surgical technical skill [OSATS global rating scale (GRS) and a procedure-specific rating scale for LC (PSRS)] [7, 19]. Participants were given immediate post-procedure feedback on their performance based on rating scales. The feedback was guided based on the OSATS GRS and PSRS. For example, if a participant scored poorly on “instrument handling,” feedback was provided on how to improve handling of instruments. If a participant scored poorly on “Cystic duct clipping and transaction” on the PSRS, feedback was given on how to improve placement of clips.

Each subject then performed 30 minutes of deliberate practice using the LapSim VR or LapMentor VR laparoscopic simulators, focusing on the weakest skills as assessed by either the GRS or PSRS. Deliberate practice here is defined as effortful practice that is guided by a coach (the qualified observer) and focused on addressing specific weaknesses in performance. Assignment of DP tasks based on weakest rating scale performance is described in depth in Crochet et al [17]. Briefly, participants were assigned a practice task on the LapSim based on the lowest OSATS GRS dimension until a minimum of 3 out of 5 was reached for each dimension. If all dimensions on the OSATS GRS were 3 out of 5, a practice task on the LapMentor was assigned based on the lowest PSRS dimension (Table 1). If one or more dimensions were equally the lowest, the participant was given a choice of practice activity.

After DP, each participant performed a second VR LC on the LapMentor simulator and received post-procedure feedback based on the rating scales as described above.

### **Control Group Training Sessions**

The control group also underwent 10 training sessions comprising a total of 20 VR LCs. At the start of each session, participants performed a VR LC and were then assigned 30 minutes of control activity (viewing a TED talk or reading a journal article) unrelated to laparoscopy or cholecystectomy. After control activity, participants were asked to perform another VR LC. This group did not receive any commentary or advice during or after their performance on the simulator, and no practice tasks were assigned. However, their performance was assessed in real time using the same rating scales, and participants were able to review quantitative performance on time, number of movements, and path length on the VR simulator screen.

### **Cadaveric Porcine Model for Laparoscopic Cholecystectomy**

Following VR training sessions, all subjects performed 5 LCs on a cadaveric porcine model using real surgical instruments, including the use of diathermy as previously described [20]. Performance was assessed with in-house motion tracking software, with all participants in both groups being rated on the OSATS GRS and PSRS and receiving feedback after each LC. No practice was allowed between porcine LCs.

### **Post-training Assessment**

After completing all porcine LCs, each participant was evaluated on an exit assessment on the LapMentor VR simulator using Basic Skills tasks 5 and 6, repeated twice each (Figure 1).

### **Quality of Performance Assessment**

An OSATS GRS score of 21 was arbitrarily defined by surgical educators at Imperial College London as a target interim level of performance to be achieved, based upon a score of 3 in each measured domain. The OSATS GRS is a 7-domain rating scale with each domain being rated on a 5-point Likert scale; thus, a score of 3 on each domain was selected as a target for interim performance as these are anchored by descriptors suggesting acceptable performance [7]. Subjects were aware of this target at the start of the study.

## Statistical Analysis

Statistical analysis was performed using STATA Intercooled 12 (College Station, TX). The Shapiro-Wilk test showed the nature of the data to be nonparametric. The Mann-Whitney U-test was employed to compare intergroup baseline to post-test laparoscopic basic task performance, VR training session performance, and porcine model performance. The Wilcoxon signed-rank test was utilized for intra-group comparison. VR training session and porcine model data was analyzed using nonlinear regression to assess learning curves and plateaus [21]. The proportion of participants obtaining interim performance level in each group was compared using the  $\chi^2$  test. Results were reported as median (interquartile range). Levene's test was utilized to compare the consistency in performance of the deliberate practice group versus the control group in both VR training sessions and porcine model sessions. A  $p < 0.05$  was considered statistically significant.

In addition to live ratings, videos of VR and porcine LC were assessed by an independent, blinded rater using the GRS and PSRS. Intraclass correlation coefficient (ICC) was calculated to assess the blinded inter-rater reliability of both the GRS and PSRS. ICC of live and video raters was found to be 0.84 for the OSATS GRS and 0.9 for the PSRS.

Sample size was based on detecting at least a 25% difference in quality of VR surgical performance as assessed by the OSATS global rating scale with alpha of 0.05 and beta of 0.8.

## Results

### Subjects

Fourteen of the twenty (70%) recruited junior doctors completed the study with those who dropped out citing scheduling conflicts for their inability to complete the study. Five of twenty dropped out after baseline testing, and one participant in the DP group dropped out at the completion of VR LC training. The data from the participant who dropped out after VR training was included in the VR data analysis. Training level of the participants ranged from post-graduate year 1 (PGY1) to post-graduate year 3 (PGY3). There were seven participants in the control group and seven in the DP group who completed the study. All study participants were right-handed.

### Baseline Assessment of Laparoscopic Skill

At baseline, for Basic Skills task 5, there was no difference in time to completion between groups. For Basic Skills task 6, participants in the DP group were faster and more dexterous than those in the Control group (Table 2). After proficiency training, there were no differences between groups on task 5 in time ( $p=0.06$ ), or during task 6 for time ( $p=0.22$ ), movements ( $p=0.90$ ), or path length ( $p=0.10$ ). Thus, both groups were of similar proficiency before beginning VR LC training sessions.

### Completion of Virtual Reality Training Session Performance

At the end of VR training, there were no significant differences in speed or dexterity (Table 3); however, the DP group had significantly higher ratings than the control group for quality

of surgical performance in global (26 vs. 20;  $p=0.001$ ) and procedure-specific rating scales (18 vs. 15;  $p=0.001$ ) (Figure 2). Both groups improved over 20 VR LCs in time, dexterity, and global rating scales. The DP group improved in scores on the PSRS over the course of VR training while the control group did not. There were no significant differences in variance between groups in time, dexterity, or quality of performance.

### Virtual Reality Laparoscopic Cholecystectomy Learning Curves

Analysis of learning curves for the 20 VR LCs for the control group showed a plateau at 400 seconds ( $p<0.001$ ), 364 movements ( $p<0.001$ ), and 777 cm path length ( $p<0.001$ ). The control group plateaued at OSATS GRS of 20 ( $F_{1,138}=49.9$ ,  $df=139$ ,  $p<0.001$ ) after 5 cases.

These results compare to the DP group where analysis showed a plateau in time of 424 seconds ( $p<0.001$ ), 367 movements ( $p<0.001$ ), and 603.1 cm of path length ( $p<0.001$ ) over the course of 20 LCs. The DP group on average plateaued at an OSATS GRS of 25 ( $F_{1,158}=73.7$ ,  $df=159$ ,  $p<0.001$ ) after 10 cases.

There were no significant differences in the learning curves of time, movements, or path length between the two groups. There was a significant difference in the OSATS GRS learning curves of the two groups ( $F_{3,296}=132.17$ ,  $p<0.001$ ) (Figure 2). At the fifth case when the control group plateaued, the DP group had a significantly higher median OSATS GRS of 21.5 (19.5–24) compared to control of 19 (16–20;  $p=0.03$ ). This difference was maintained at the tenth LC when the DP group plateaued with the DP group again having a higher median OSATS GRS of 24.5 (22–27) compared to control of 20 (18–20;  $p=0.002$ ).

Analysis of the rate of quality improvement in VR LC performance demonstrated that 100% of the DP group achieved the interim level of performance by the 15<sup>th</sup> VR LC. The proportion of participants in the DP group achieving interim level of performance by the 20<sup>th</sup> and final VR LC (100%) was significantly greater than that of the control group (30%;  $p=0.007$ ) (Figure 3).

### Porcine Model Performance

Analysis of motion tracking data of subjects performing laparoscopic cholecystectomy in the porcine model demonstrated that those in the control group did not improve from the first to fifth and final LC in time, movements, path length or OSATS GRS. While the DP group also did not show improvement in time, path length or OSATS GRS from first to fifth and final LC, the DP group did produce significantly fewer movements at the fifth LC. The DP and control group demonstrated no significant differences at the fifth and final LC in time, movements, path length or PSRS. The DP group had significantly higher quality of performance than the control group as assessed by OSATS GRS (24 vs. 19;  $p=0.007$ ). Both groups improved in PSRS scores from first to final porcine LC (Table 4).

### Porcine Laparoscopic Cholecystectomy Learning Curves

Analysis of learning curves for porcine cases for the control group showed no plateau in time, number of movements, total path length, or OSATS GRS over the course of 5 LCs but did demonstrate plateau at a PSRS score of 20 after 4 cases ( $F_{1,33}=11.49$ ,  $df=34$ ,  $p<0.001$ ).

The DP group similarly demonstrated no plateau in time, movements, or path length. For OSATS GRS, the DP group demonstrated a non-significant trend toward plateau at a score of 22 ( $p=0.06$ ) (Figure 4). The DP group plateaued at a PSRS score of 22 after 5 cases ( $F_{1,33}=10.24$ ,  $df=34$ ,  $p=0.003$ ).

Neither group fully achieved the interim performance level on porcine LC, but there was a trend for increasing performance in both groups over the course of five porcine LCs. There was no significant difference between groups in the proportion of participants achieving interim level of performance.

### Post-Training Assessment

In the post-training assessment of basic skills, both groups were significantly faster at Basic Task 5 when compared to baseline. Both groups were also significantly faster and more dexterous at Basic Task 6 when compared to baseline (Table 1). There were no significant differences between groups in time or dexterity in either basic skills task at post-training assessment.

### Discussion

Deliberate practice has been investigated as a key component of the development of expertise, particularly in fields demanding consistent superior performance. Initial acquisition of skills by novices requires concentration and repetition to automate motions to a level acceptable for everyday performance. This initial phase has been hypothesized to require around 50 hours, after which performance will plateau without further effort to improve performance. Those who continue DP will improve performance toward the level of expert; however, a lapse in practice will lead to a premature plateau of performance known as “arrested development” that, though above the level of everyday skill, falls short of expert performance [16, 22, 23]. This phenomenon has been demonstrated repeatedly in the sports science and music literature and may play a similar role in the acquisition and maintenance of surgical skill [16].

While junior surgeons engaging in DP were found to be equivalent to the control group for speed and dexterity, results of this study demonstrated that engaging in DP led to higher quality of surgical performance as assessed by OSATS GRS and PSRS in LC in both VR and live tissue. Furthermore, junior surgeons engaging in DP plateaued in skill after a greater number of cases at a higher quality of surgical performance, suggesting that DP leads to continued performance improvement when compared to standard, non-directed training.

Deliberate practice in this study followed Ericsson’s definition as that which involves concentration and attention directed at effortful practice that is guided by specific feedback on improving performance [17, 22, 24, 25]. As the DP group was receiving directed feedback with performance-specific practice activities, this group would expectedly have higher quality of performance than the control group, which did not receive directed guidance on improving surgical quality. In comparing the DP group’s learning curve for quality of performance to that of the control group, the control group may have been left in “arrested development” where further time spent practicing would not have improved

performance given the plateau in the learning curve (Figure 2) [16]. Further evidence for arrested development can be seen in the VR LC performance rates of each group. By the 15<sup>th</sup> LC, 100% of DP participants achieved the interim level of performance while only 30% of the control group achieved it, suggesting arrested development in the remaining 70% of the control group.

There is a similar trend seen in porcine LCs with 86% of the DP group achieving the interim level of performance by the fifth and final porcine LC versus 43% in the control group. The lack of a statistically significant difference may be due to the relatively small number of porcine procedures in comparison to VR procedures. Given the trend for improving performance, additional sessions on the porcine model may have demonstrated a significant difference in quality of performance between the two groups, but the cost of obtaining porcine models and time constraints prevented additional porcine sessions. However, this trend of improved rate of achieving interim level of performance in the DP group suggests that DP is transferrable from VR to a real tissue model, and its effects on training appear to be sustained as trainees graduate from VR simulation to operating on real tissue.

This study is limited by the high dropout rate of study participants with 30% of the original 20 participants dropping out of the study. Participants cited time demands (the number of site visits required to complete the study ranged from 17 – 22 depending on the number of sessions required to complete the pre-study proficiency curriculum) as the primary reason for dropout. Despite the dropout rate, the study was adequately powered to detect at least a 25% difference in quality of performance between the two groups in VR training as reflected in the power calculation, which was based off preliminary data collected in a VR setting.

A porcine model was included in the study to assess for transfer of skill from a VR to a tissue model of LC despite inherent differences when compared to VR. Porcine models are of greater anatomic complexity and track all motions including instrument changes whereas VR models focus on motions involved in operative steps alone, and these differences were reflected in our results where porcine LC required longer times and greater number of movements and path length. The lack of improvement in time/dexterity over the course of the five porcine LCs may be due to plateau in skill demonstrated in the VR model. Overall quality of the procedure remained consistent despite differences in time/dexterity compared to a VR model, suggesting that DP participants maintain principles of a quality operation despite the greater complexity of a live tissue model.

Studies of DP within medicine have been undertaken but focus on the cognitive elements of medicine, such as making a clinical diagnosis or reading radiologic films [26]. There is a paucity of studies regarding the application of DP in the acquisition of technical skills for surgeons [15, 26], but some work has demonstrated that junior surgical trainees and medical students undergoing DP were better than the control group, particularly in quality of performance as assessed by the OSATS GRS [17, 27]. This is the first study to quantify the learning curve of surgical quality in DP versus standard training. It demonstrates not only a transfer of skill in LC from VR to tissue but also continued learning past the arrested development plateau of standard training.

While deliberate practice can be a tool to improve the training of surgeons, it is not without its limitations; and its implementation into surgical training curricula may be challenging. DP necessitates a trainee's regular engagement in practice activities with feedback from an instructor. The surgical resident's multiple responsibilities as clinical care provider and trainee, coupled with duty hour regulations, minimizes the amount of time that can be dedicated to participation in DP-based simulation training. DP would require a significant time commitment from faculty who would need to provide consistent feedback designed to address deficiencies in performance. Preliminary incorporation of DP into training would allow for data collection on its impact on clinical outcomes, after which programs would be able to decide if DP would be cost effective for their institution.

## Conclusion

Results of this study suggest that DP leads to superior quality of performance in junior surgeons. While dropout from several participants limited our study, power calculations demonstrated an adequate sample. Though not every practicing surgeon may attain "expert" status, DP has the potential to serve as a powerful tool in surgical education to improve the performance of surgeons with the goal of improving patient outcomes through the reduction of morbidity and mortality related to technical limitations of the surgeon. Future research should investigate integration of DP into surgical training curricula and its impact on patient outcomes.

## Acknowledgments

The authors would like to thank Steve Marchington and Kenneth Miller for their administrative and logistical support in arranging materials for use in this study. We thank Dr. Steve Siegel for his assistance with the manuscript and Dr. Andrew Cucchiara for his biostatistical support. The authors are grateful to Drs. Emma Meagher and Karen Kerr for administrative support and assistance in facilitating institutional collaboration.

Daniel Hashimoto and Ernest Gomez were partly supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant TL1TR000138.

## References

1. Medicine, Io. To Err is Human: Building a Safer Health System. National Academies Press; Washington, DC: 2000.
2. Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*. 1999; 126:66–75. [PubMed: 10418594]
3. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, Hebert L, Newhouse JP, Weiler PC, Hiatt H. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med*. 1991; 324:377–384. [PubMed: 1824793]
4. Thomas EJ, Studdert DM, Newhouse JP, Zbar BI, Howard KM, Williams EJ, Brennan TA. Costs of medical injuries in Utah and Colorado. *Inquiry*. 1999; 36:255–264. [PubMed: 10570659]
5. Cameron J. William Stewart Halsted: Our Surgical Heritage. *Annals of Surgery*. 1997; 225:445–458. [PubMed: 9193173]
6. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR. Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. *Br J Surg*. 2008; 95:1088–1097. [PubMed: 18690637]
7. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg*. 2008; 247:372–379. [PubMed: 18216547]

8. Willaert WI, Aggarwal R, Daruwalla F, Van Herzele I, Darzi AW, Vermassen FE, Cheshire NJ. Simulated procedure rehearsal is more effective than a preoperative generic warm-up for endovascular procedures. *Ann Surg.* 2012; 255:1184–1189. [PubMed: 22566016]
9. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, Darzi A. An evidence-based virtual reality training program for novice laparoscopic surgeons. *Ann Surg.* 2006; 244:310–314. [PubMed: 16858196]
10. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *Br J Surg.* 2009; 96:1086–1093. [PubMed: 19672934]
11. Arora S, Aggarwal R, Sirimanna P, Moran A, Grantcharov T, Kneebone R, Sevdalis N, Darzi A. Mental practice enhances surgical technical skills: a randomized controlled study. *Ann Surg.* 2011; 253:265–270. [PubMed: 21245669]
12. Hashimoto DA, Gomez ED, Danzer E, Edelson PK, Morris JB, Williams NN, Dumon KR. Intraoperative resident education for robotic laparoscopic gastric banding surgery: a pilot study on the safety of stepwise education. *J Am Coll Surg.* 2012; 214:990–996. [PubMed: 22521438]
13. Seymour NE, Cooper JB, Farley DR, Feaster SJ, Ross BK, Pellegrini CA, Sachdeva AK. Best practices in interprofessional education and training in surgery: experiences from American College of Surgeons-Accredited Education Institutes. *Surgery.* 2013; 154:1–12. [PubMed: 23809478]
14. ACGME. Committee RR. ACGME Program Requirements of Graduate Medical Education in Surgery. Accreditation Council for Graduate Medical Education; Chicago, IL: 2008.
15. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004; 79:S70–81. [PubMed: 15383395]
16. Ericsson KA, Nandagopal K, Roring RW. Toward a science of exceptional achievement: attaining superior performance through deliberate practice. *Ann N Y Acad Sci.* 2009; 1172:199–217. [PubMed: 19743555]
17. Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T, Ericsson KA, Darzi A. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. *Ann Surg.* 2011; 253:1216–1222. [PubMed: 21516035]
18. Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *Am J Surg.* 2006; 191:128–133. [PubMed: 16399123]
19. Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997; 84:273–278. [PubMed: 9052454]
20. Aggarwal R, Ward J, Balasundaram I, Sains P, Athanasiou T, Darzi A. Proving the effectiveness of virtual reality simulation for training in laparoscopic surgery. *Ann Surg.* 2007; 246:771–779. [PubMed: 17968168]
21. Feldman LS, Cao J, Andalib A, Fraser S, Fried GM. A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining “learning plateau” and “learning rate”. *Surgery.* 2009; 146:381–386. [PubMed: 19628099]
22. De Bruin ABH, Rikers RMJP, Schmidt HG. The influence of achievement motivation and chess-specific motivation on deliberate practice. *Journal of Sport and Exercise Psychology.* 2007; 29:561–583. [PubMed: 18089893]
23. Krampe RT, Ericsson KA. Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J Exp Psychol Gen.* 1996; 125:331–359. [PubMed: 8945787]
24. Cumming J, Hall C. Deliberate imagery practice: the development of imagery skills in competitive athletes. *J Sports Sci.* 2002; 20:137–145. [PubMed: 11811570]
25. Ericsson, KA. *The road to excellence: the acquisition of expert performance in the arts and sciences, sports, and games.* Lawrence Erlbaum Associates; Mahwah, N.J: 1996.
26. Ericsson KA. An expert-performance perspective of research on medical expertise: the study of clinical performance. *Med Educ.* 2007; 41:1124–1130. [PubMed: 18045365]

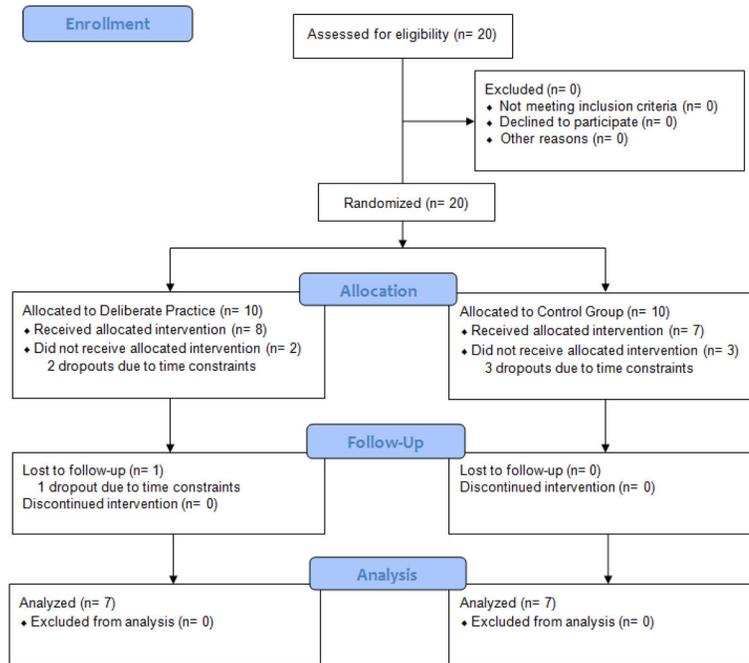
27. Price J, Naik V, Boodhwani M, Brandys T, Hendry P, Lam BK. A randomized evaluation of simulation training on performance of vascular anastomosis on a high-fidelity in vivo model: the role of deliberate practice. *J Thorac Cardiovasc Surg.* 2011; 142:496–503. [PubMed: 21742349]

Author Manuscript

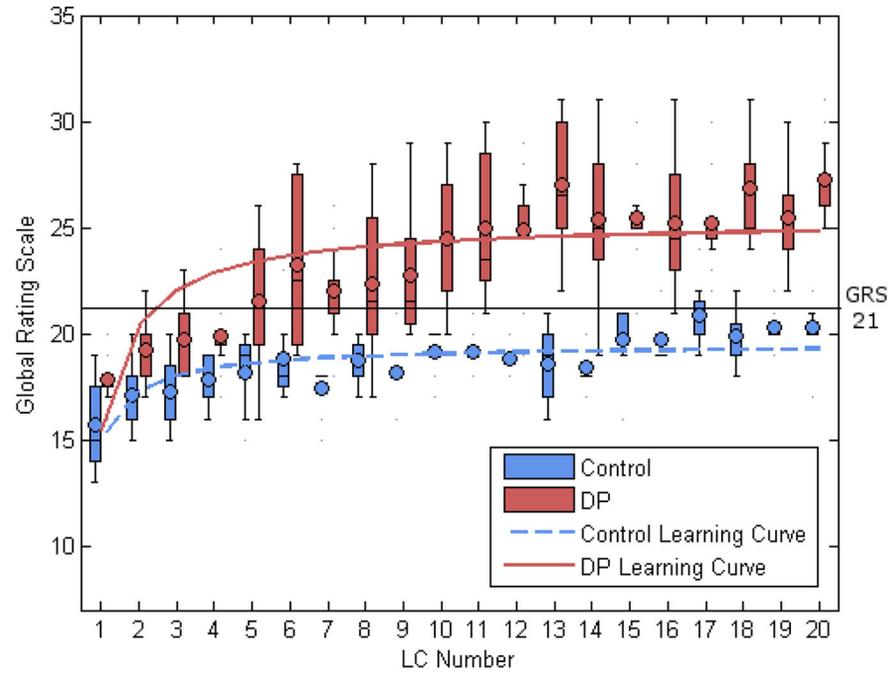
Author Manuscript

Author Manuscript

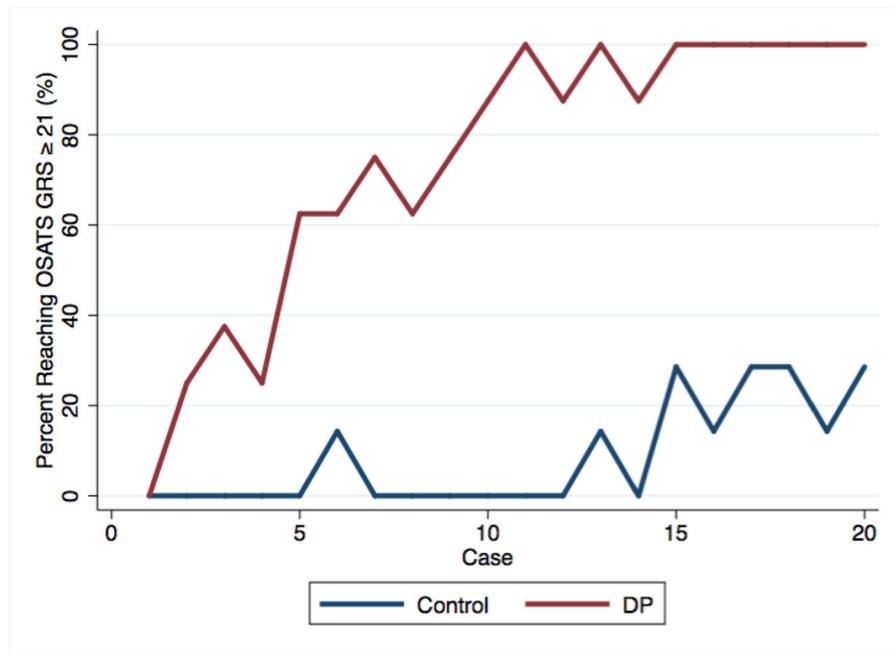
Author Manuscript



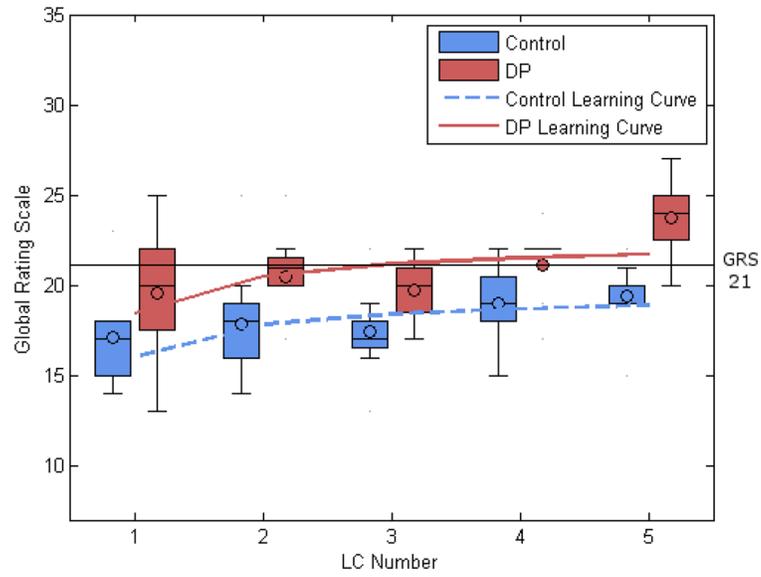
**Figure 1.** Flow chart of study protocol with recruited subjects and drop-outs



**Figure 2.** Learning curve and box-and-whisker plot demonstrating greater improvement of DP group in quality of surgical performance versus control.



**Figure 3.** Quality of performance rate comparison of Control and DP groups showing achievement of 100% interim level of performance (OSATS GRS score  $\geq 21$ ) in the DP group by the completion of VR training.



**Figure 4.** Learning curve comparison and box-and-whisker plot of OSATS GRS demonstrating significantly greater quality of performance in the DP group at the end of porcine training.

**Table 1**

Correspondence between rating scale criteria and training task assigned on Lapsim

<b>GENERIC RATING SCALE ITEM</b>	<b>ASSIGNED TASK</b>
Respect for tissue	Clip and cut (LS)
Time and motion	Lift and Grasp (LS)
Instrument handling	Grasping (LS)
Use of assistants	Coordination (LS)

<b>PROCEDURE-SPECIFIC RATING SCALE ITEM</b>	<b>ASSIGNED TASK</b>
Cystic duct dissection	Calot's triangle dissection (LM)
Cystic duct clipping and transection	Clipping and cutting (LM)
Gallbladder fossa dissection	Gallbladder separation (LM)

(LS) Or LapMentor (LM)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Pre-test baseline skills assessment of control and DP groups compared to post-test skills assessment for basic laparoscopic skills in a VR simulator. Values as median (interquartile range). Significant values in bold.

	Control	DP	p-value
<b>Task 5</b>			
Pretest Time (sec)	139.6 (122–149)	134.1 (114–159)	0.84
Posttest Time (sec)	84.8 (78–94)	86.5 (76–102)	0.44
<b>p-value</b>	<b>0.001</b>	<b>0.001</b>	
<b>Task 6</b>			
Pre-test Time (sec)	177.3 (150–265)	142.6 (123–192)	<b>0.04</b>
Post-test Time (sec)	97.1 (80–113)	91.8 (81–106)	0.85
<b>p-value</b>	<b>0.001</b>	<b>0.001</b>	
Pre-test Movements	233 (203–347)	192.5 (133–204)	<b>0.01</b>
Post-test Movements	112 (97–127)	95 (86–114)	0.06
<b>p-value</b>	<b>0.001</b>	<b>0.003</b>	
Pre-test Path Length (cm)	618 (534–813)	486 (451–519)	<b>0.004</b>
Post-test Path Length (cm)	287.7 (274–320)	291 (267–327)	0.96
<b>p-value</b>	<b>0.001</b>	<b>0.002</b>	

**Table 3**

Comparison of first and final VR LC time, movements, path length, and quality of surgical performance between groups. Values as median (interquartile range). Significant values in bold.

	Control	DP	p-value
<b>Time (sec)</b>			
<b>First LC</b>	490.5 (482–555)	554.9 (527–630)	0.13
<b>Final LC</b>	381.7 (324–432)	433.3 (313–443)	0.64
<b>p-value</b>	<b>0.03</b>	<b>0.01</b>	
<b>Movements</b>			
<b>First LC</b>	522 (385–586)	560 (526–610)	0.22
<b>Final LC</b>	360 (320–429)	339.5 (291–365)	0.42
<b>p-value</b>	<b>0.02</b>	<b>0.01</b>	
<b>Path Length (cm)</b>			
<b>First LC</b>	816.1 (679–997)	905 (827–1056)	0.17
<b>Final LC</b>	592.4 (501–759)	537 (510–615)	0.56
<b>p-value</b>	<b>0.02</b>	<b>0.01</b>	
<b>OSATS GRS</b>			
<b>First LC</b>	15 (14–18)	18 (17–19)	0.08
<b>Final LC</b>	20 (20–22)	26 (25.5–30)	<b>0.001</b>
<b>p-value</b>	<b>0.02</b>	<b>0.01</b>	
<b>PSRS</b>			
<b>First LC</b>	15 (14–15)	14.5 (13–16)	0.59
<b>Final LC</b>	15 (14–15)	18 (17.5–20)	<b>0.001</b>
<b>p-value</b>	0.79	<b>0.01</b>	

**Table 4**

Comparison of first and final Porcine LC time, movements, path length, and quality of surgical performance between groups. Values as median (interquartile range). Significant values in bold.

	Control	DP	p-value
<b>Time (sec)</b>			
<b>First LC</b>	3826.3 (1840–4397)	2953.1 (2904–2958)	0.37
<b>Final LC</b>	2187.7 (1814–3430)	2076 (1186–2831)	0.37
<b>p-value</b>	0.74	0.35	
<b>Movements</b>			
<b>First LC</b>	1674 (942–1953)	1679 (1430–1817)	0.86
<b>Final LC</b>	1503 (1171–1594)	1141 (713–1704)	0.23
<b>p-value</b>	0.75	<b>0.04</b>	
<b>Path Length (cm)</b>			
<b>First LC</b>	268.4 (161–380)	257.1 (251–294)	0.86
<b>Final LC</b>	248.8 (201–407)	174.1 (139–253)	0.18
<b>p-value</b>	0.35	0.35	
<b>OSATS GRS</b>			
<b>First LC</b>	17 (14–18)	20 (16–23)	0.22
<b>Final LC</b>	19 (19–21)	24 (22–25)	<b>0.02</b>
<b>p-value</b>	0.13	0.09	
<b>PSRS</b>			
<b>First LC</b>	16 (14–18)	18 (16–19)	0.18
<b>Final LC</b>	20 (18–22)	22 (22–24)	0.07
<b>p-value</b>	<b>0.04</b>	<b>0.03</b>	