

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/125143/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Brown, Chris, Robinson, David, Egan, Richard, Hopkins, Luke, Abdelrahman, Tarig, Powell, Arfon , Pollitt, M. John and Lewis, Wyn G. 2019. Prospective cohort study of haptic virtual reality laparoscopic appendicectomy learning curve trajectory. *Journal of Laparoendoscopic and Advanced Surgical Techniques* 29 (9) , pp. 1128-1134. 10.1089/lap.2019.0332

Publishers page: <http://dx.doi.org/10.1089/lap.2019.0332>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Prospective cohort study of haptic virtual reality laparoscopic appendicectomy learning curve trajectory

Brown C MRCS¹, Robinson D MRCS¹, Egan R MRCS², Hopkins L MRCS¹,
Abdelrahman T MD¹, Powell A PhD³, Pollitt MJ FRCS⁴, Lewis WG DSc¹.

¹ Department of Surgery, University Hospital of Wales, Cardiff, CF14 4XW.

² Department of Surgery, Morriston Hospital, Heol Maes Eglwys, Swansea, SA6 6NL.

³ Division of Cancer and Genetics, Cardiff University School of Medicine, Heath Park, Cardiff CF14 4XW.

⁴ Department of Surgery, Prince Charles Hospital, Gurnos Road, Merthyr Tydfil, CF47 9DT.

Correspondence to: Professor WG Lewis, Head of School Surgery, Wales PGMDE Deanery, University Hospital of Wales, Heath Park, Cardiff, CF14 4XW, UK.

E: wyn.lewis4@wales.nhs.uk

T: 029 2074 3268

F: 029 2074 4553

Declarations

This study was supported by a grant from Health Care Research Wales.

Key words: laparoscopy, appendicectomy, procedural training, haptic, simulation, virtual reality

Word count: 3,114 (abstract 236)

Abstract (236 words)

Background: Simulation training is strongly advocated by 24/7 risk-rich professions, because swift learning curve inflection point attainment, delivers earlier competence; the left shift effect. The aim of this study was to determine the value of Haptic Laparoscopic Virtual Reality Simulation (HLVRS), by iterative benchmark exercise (n=8), prior to Simulated Laparoscopic Appendicectomy (SLA); the hypothesis was that favourable benchmark learning curve trajectories would be associated with improved SLA competence when compared with consultant expert performance.

Methods: A 28-trainee cohort completed 1,349 LHVRS tasks, during which 19 ergonomic variables were assessed by virtual interface including force feedback (Surgicalseience.com), prior to 153 SLAs. Primary outcome measure was SLA composite competence score related to 6 consultant trainer experts.

Results: Of the 8 LHVRS tasks, the 3 with the steepest learning curve trajectories correlated with better median overall SLA competence scores, namely: tissue grasping/lifting (rho 0.362, p=0.049), fine dissection (rho 0.388, p=0.028), and camera navigation (rho 0.518, p=0.007); fine dissection was the only HLVRS task that predicted a SLA score within a Youden index defined, 70% of the consultant expert level (AUC 0.803, p=0.028). A significant SLA learning curve emerged, with a learning curve trajectory inflection point at the 4th SLA attempt (1st SLA 30.5% vs. 4th SLA score 76.0%, gradient 76⁰, p=0.010).

Conclusion: Learning curve trajectory can be measured, influenced and accelerated significantly; a pronounced left shift effect, with translational potential for enhanced shorter training time and improved patient safety.

Introduction

Simulation training has long been strongly advocated by professions with inherent high-risk profiles in their job descriptions and person specifications (1-3), especially so in first world economies that demand around the clock access to essential services, dependent on human resource. Running the numbers suggests that effective simulation should produce a left shift in any given learning curve, such that the inflection point and standard of competence is reached earlier than by chance. Learning curves are often referred to in the context of medical education, although their trajectories and natures are a matter of debate. Serial evaluation of operation specific outcomes can plot a surgeon's position on a curve, with competence deemed to be the point at which the curve trajectory reaches a plateau phase (inflection point), consistent with satisfactory quality (4). Curve trajectory or gradient equates to the rate of improvement of performance and may serve as an alternative metric of skill progression. Appendicectomy remains one of the commonest emergency surgical procedures and is one of six index operations in which UK general higher surgical trainees (HST) must demonstrate competency prior to achieving a certificate of completion of training (CCT) (5). The Joint Committee on Surgical Training (JCST) requires evidence of a minimum operative case load of 80 appendicectomies prior to the award of CCT and published UK Deanery learning curve data reports a median case load of 95 (IQR 83 - 137) to achieve the 3rd level 4 competence validation(6).

The aim of this study was to determine the value of a haptic laparoscopic virtual reality simulation model (LapSim®, Surgical Science Sweden AB) (7) by iterative benchmark exercise prior to simulated laparoscopic appendicectomy; the hypotheses were two-fold: first, that favourable benchmark learning curve

trajectories would be associated with improved simulated laparoscopic appendicectomy competence and trajectory when compared with consultant expert performance; second, that a measurable significant improvement in performance would be observed with iterative simulated laparoscopic appendicectomy performance.

Methods

The study was undertaken in a hybrid boot camp blended with clinical operative training, as part of a single UK Deanery national surgical training programme. The design was a prospective cohort study examining simulator performance over time.

Participants and setting

A total cohort of 28 self-selecting surgical trainees (19 Core Surgical Training (CST) and 9 Higher Surgical Training (HST)), representing 28.6% of contemporary general surgery trainees from CT1 to ST5 level within a single UK deanery, underwent a standardised programme of haptic laparoscopic virtual reality appendicectomy simulation training. Prior individual trainee experience was captured using feedback questionnaires to determine baseline levels of experience, including both simulation and operative experience, together with their contemporary Procedural Based Assessment (PBA) level for laparoscopic appendicectomy.

Description of Simulator, Procedural Module, and Simulator Measurements

A Simcart, table-mounted and height-adjustable LapSim virtual reality laparoscopic simulator with integrated haptic technology was used (LapSim®, Surgical Science Sweden AB). The system consisted of a software program run on an Intel Core i7 processor (Intel Corporation, Santa Clara, CA) using Windows 10 Pro (Microsoft Corporation, Redmond, WA). The computer was equipped with 8 GB of internal RAM, a NVIDIA GeForce GTX 1050 graphics card (NVIDIA Corporation, Santa

Clara, CA), a 27-inch monitor, and a virtual laparoscopic interface including 2 laparoscopic instruments with haptic force feedback and a camera. In this study the 2017.9 version of the system was utilized including LapSim Basic Skills and LapSim Appendectomy programs. Four identical LapSim units were provided, running simultaneously in a single isolated training room in the Wales Institute of Minimal Access Therapy (WIMAT).

Simulation procedural modules

The training was separated into 3 modules.

Module 1. Three basic exercises were performed to build system familiarity.

Module 2. Eight Haptic Laparoscopic Virtual Reality Simulation (HLVRS) benchmark tasks were performed as a programme to develop and individually assess the component skills necessary for Simulated Laparoscopic Appendectomy, ranging from camera navigation to more advanced skills such as coordination, clip application, lifting, grasping and cutting.

Module 3. The final procedural module consisted of performing a simulated laparoscopic subcaecal-appendectomy, using the skills practiced beforehand, a hook diathermy electrode and ligating loops (Figure 1). Before removal of the appendix, adhesions along the length of the appendix had to be removed. The mesoappendix then had to be divided to the base of the caecum, and 3 ligating loops had to be placed correctly (2 proximally on the base of the appendix, and 1 distally). Finally, the appendix had to be divided between the ligating loops, and the specimen removed in an extraction bag. In case of perforation of the appendix or caecum, which could be caused by either excessive pressure or with hook electrodes or scissors, the procedure could not be completed and the attempt was ended. The recorded outcomes were the following 19 simulator parameters: total procedure time

(seconds); right and left instrument angular path (degrees); right and left instrument path length (meters); left and right instrument outside view (number); left and right instrument outside view (seconds); blood loss (ml); energy applied in air (seconds); total time burned on appendix (seconds); number of times burned on appendix (number); percent of adhesions ablated; percent of mesentery ablated; percent of adhesions ripped; percent of mesentery ripped; appendix evacuated from body (yes/no). Training sessions lasted a maximum of 4 hours and simulator measurements were stored automatically after each attempt on the simulator. The principal investigator was present throughout; for technical assistance during the training sessions and registered simulator related irregularities, in which case the attempt was terminated.

Data collection

The basic exercise tasks were not scored, but the 8 benchmark-tasks together with the laparoscopic appendicectomy procedures were rated and assessed by means of composite competence scores derived from the individual simulator parameters. An overall pass mark was calculated and defined for all tasks, including imulated laparoscopic appendicectomy, based on the performance of a cohort of six consultant expert trainers, which completed the full programme before the trainees. Allowance buffers were created based on a review of the relevant published literature (8-10); a pass mark or score for the variables in the 8 benchmark training tasks was defined by comparison with median and lower quartile consultant scores.

Learning curve trajectory

Rates of improvement in module-2 task performance scores were determined by subtracting a participant's first score, from the best attempt score, for each of the 8 Haptic Laparoscopic Virtual Reality Simulation tasks. A task improvement ratio was

then calculated by dividing the above score differential by the number of task attempts performed to achieve the best task score; defined as a final task improvement / attempt ratio. Composite scores for each simulated task were plotted graphically versus the number of attempts to develop learning curve trajectories for each task. Learning curve trajectory gradients related to number of attempts (first to second, second to third, third to fourth, fourth to fifth, fifth to sixth) to allow for arbitrary, objective comparisons between tasks were calculated using standard trigonometric techniques (inverse function of tan) (11). Primary outcome measures were:

1. SLA composite competence score and iterative trajectory related to the cohort of six consultant trainer experts.
2. Whether performance in the eight Haptic Laparoscopic Virtual Reality Simulation benchmark tasks predicted SLA performance and iterative trajectory, and the 19 simulator ergonomic variable parameters outlined above.

Statistical Analysis

Sample size calculations were based on a pre-study literature survey, which indicated that general surgeons in training achieve a third level 4 competence PBA after a median of 95 laparoscopic appendicectomies (IQR 54, SD +/- 45). Thus, it was calculated that a minimum of fifteen participants would need to be enrolled into the study, providing 80% power with alpha set at 0.05 (6, 12).

Statistical analysis appropriate for non-parametric data (Kruskal-Wallis, Mann-Whitney U, Spearman's rank correlation) and binomial logistic regression including ROC curve analysis, was performed using SPSS 25 (IBM Corp, Armonk, NY).

Simulate Laparoscopic Appendicectomy composite scores were dichotomised using the point closest to (0,1) corner in the ROC plane approach, with the cut point based

on the Youden Index (13).

Formal ethical approval was not required for this study since it did not involve a National Health Service organisation site and participants were recruited voluntarily by virtue of their role as postgraduate trainees within the Wales Deanery School of Surgery rather than as NHS employees.

Results

A total of 28 participants (9 female, 19 male) comprising 19 Core Surgical (CST) and 9 General Higher Surgical trainees within a single UK deanery completed the study.

Median total operative caseload prior to commencement of the study was 463 (range 35 to 2461) and appendicectomy experience as per logbook entry: assisted 10 (2 to 51), supervised trainer scrubbed 13 (0 to 55), supervised trainer unscrubbed 0 (0 to 36), and performed 0 (0 to 137). Contemporary Intercollegiate Surgical Curriculum Programme (ISCP) Procedural Based Assessment (PBA) levels for appendicectomy were available for 20 participants: level 2a/b = 4, level 3a/b = 10 and level 4a/b = 6.

Individual simulated tasks completed numbered 1,502, consisting of 1349 basic laparoscopic skills exercises, and 153 simulated appendicectomies (table 1).

The relationship between individual Haptic Laparoscopic Virtual Reality Simulation task score improvement ratios and overall median simulated laparoscopic appendicectomy performance scores are shown in table 2.

Receiver Operating Characteristic (ROC) curves for each task that demonstrated significant correlation with overall simulated laparoscopic appendicectomy performance score were calculated and plotted. Fine dissection was the only module-2 task that predicted a simulated laparoscopic appendicectomy score within

70% of the consultant expert level: AUC 0.803, $p=0.028$ (figure 2), lifting and grasping (AUC=0.656, $p=0.269$), camera navigation (AUC=0.733, $p=0.084$).

Participant training grade was unrelated to fine dissection task scores ($p=0.530$).

Univariable analysis of individual parameters related to an overall median composite LSA score of $\geq 70\%$ is shown in table 3.

Multivariable analysis

The covariates found to be significant on univariable analysis at the $p<0.10$ level were entered into a binomial logistic regression model. No factors emerged as independent predictors of achieving an overall median composite simulated laparoscopic appendectomy score of within 70% of the consultant expert level (Step 0 Constant, $B = -0.916$ (SE 0.48) Wald 4.798, df 1, $p=0.028$, Exp (B) 0.400).

Learning curve trajectory

The overall learning curves for the fine dissection task and simulated laparoscopic appendectomy are shown in figures 3 and 4 respectively. Calculation of the trajectory of the simulated laparoscopic appendectomy learning curve demonstrated a significant inflection point at the 3rd attempt (table 4).

Discussion

Learning is most effective when people are able to practise new skills and key elements are: unequivocal definition of learning objectives and evaluation; self-paced learning; appropriate feedback; and testing that the expert phase has been achieved. Haptic, force, and tactile feedback are important laparoscopic surgical concepts, yet the additional value of haptic feedback in virtual reality training is controversial. This is the first study to examine the value of haptic laparoscopic virtual reality simulation as a predictive training tool related to laparoscopic appendicectomy; one of the most commonly performed emergency surgical operations. The principal findings were that more demanding simulated tasks predicted laparoscopic appendicectomy performance, and that simulated composite competence scores increased more than two-fold (from 30% to 76%), over an average of four operative attempts; pushing any theoretical learning curve inflection point to the left by a significant margin. Of the surgeons in training, eight (28.6%) achieved simulated laparoscopic appendicectomy composite scores within 70% of the expert standard, completing the simulated surgery 80% faster, with up to 3-fold tighter instrument control, and 20% more focused use of diathermy energy, when compared with the 20 (71.4%) of participants beneath this level. Yet all of the participants demonstrated improved performance, with task improvement ratios improving almost 4-fold (3.8 to 14.7), simulated laparoscopic appendicectomy lower quartile performance improving 2.5-fold in four attempts, and simulated laparoscopic appendicectomy median performance improving 2-fold in two attempts. Consequently, the hypothesis that simulation enhances training and is associated with a learning curve trajectory left-shift effect, with a corresponding reduction in the number of procedures and time required to achieve 3 level 4 competence procedural based assessments is plausible.

Judging clinical performance is demanding and methods of measuring of learning operative performance can be categorised into surgical process and patient outcome. Any correlation between experience and competence is controversial, and a number of reports have questioned the validity of critical indicative numbers exemplifying competence (12, 14). The operative attainments of the 2013 UK CCT cohort varied broadly; two-thirds achieved elective goals, but only half emergency target experience, and only 5 per cent non-operative technical skills (15). The operative experience to prove level 4 competences has been reported to vary over four-fold, resulting in the concept of competence ratios; the ratio of case number at which 3L4 competence proven to target indicative number). This ratio has been calculated to 1.34 for appendicectomy, but ranged from 0.76 (emergency laparotomy) to 3.40 (Hartmann's procedure) (12).

Sutherland et al in a meta-analysis of simulation training in 2006, including 30 randomised control trials with 760 participants, reported that trial quality was often poor (16). Computer simulation was better than no training at all, but not convincingly superior to standard training, or video simulation. It was concluded that while there may be compelling reasons to reduce reliance on patients, cadavers, and animals for surgical training, none of the methods of simulated training had yet been shown to be better than other surgical training. Gurusamy et al in a 2009 Cochrane review of simulation training included 23 trials (612 participants), and most were at high risk of bias. In trainees without prior surgical experience, virtual reality training shortened the time taken to complete a task, increased accuracy, and reduced errors when compared with no training. In the participants with limited laparoscopic experience, virtual reality training reduced operating time and error, and improved composite operative performance score the most (17). The most recent reevaluation of simulated

laparoscopic appendectomy as a procedural and assessment tool published by Bjerrum et al (Copenhagen, Denmark), reported that novice surgeons had significantly higher risks of tissue damage compared with experienced surgeons, and concluded that although simulator models may be useful, further development was necessary for assessment (18). Overtoom et al from the Netherlands reported a systematic descriptive overview of 87 articles related to haptic feedback during simulated laparoscopic training in a variable raft of procedures such as cholecystectomy, herniorrhaphy and basic laparoscopic tasks, with variable results, and concluded that haptic feedback had a small positive performance effect, most prominently related to complex tasks where improved learning curve trajectories were apparent (19).

Four phases have been defined in any hypothetical learning curve which is typically S-shaped: commencement of training, followed by an ascending trajectory, with the gradient indicating the rate of performance change; a third phase is reached commensurate with competence (4), with additional experience improving outcomes by small amounts, until a fourth plateau phase. A compound intricate procedure is often termed erroneously to have a steep learning curve, arguably because steepness in common parlance equates to gaining height quickly, suggesting that skills are acquired rapidly because of simplicity. In fact, complex operations are more likely to be associated with gradual learning curve trajectories with small, iterative improvements, such that competence is achieved only after considerable experience. Trajectory shift, either left or right, equating to steeper or shallower angles, or easier or more tough procedures, will have many fundamental reasons including: trainee insight, trainer skill, hospital quality, interpersonal engagement related to the drive at target attainment.

This study has a number of potential limitations. Overall procedure related performance may depend on multiple variables, not least task complexity; a trainee assessed as competent on three easy cases, may be contrary when faced with more demanding patients. Because all trainees were from a single deanery, the data must be translated with caution, although all were appointed via UK national selection, subject to exacting quality warranty. The influence of trigonometric factors with respect to *x*-axis scale in calculating trajectory gradients, may risk replicating the results if alternative formats were used, but procedure axis uniformity theoretically counteracts this effect, enabling comparison. Pass mark scores for the benchmark tasks were arbitrarily set, with allowance buffers applied from the available published literature and the studies six consultant experts, but time constraints of the four-hour training window no doubt also influenced results, as fatigue will have played a role, and rest period time was limited. Finally, simulation training, per se, comes at a financial cost and data as to whether this may be reclaimed, in terms of better patient safety and improved clinical outcomes has not yet been evoked. Moreover, there remains the spectre that participants are always aware that any given simulated scenario does not carry the inherent true-life risks of clinical training procedures on real patients. To provide balance, the strengths of the study are its statistical power, and in particular relate to the engineering advantages associated with the type of virtual reality simulation examined, which is considered optimal when operating at six degree of freedom (6DoF). This term refers to the freedom of movement of a body in three-dimensional space. Specifically, the body is free to change position as forward or backward (surge), up or down (heave), left or right (sway) translation in three perpendicular axes, combined with changes in orientation through rotation about three perpendicular axes, often termed yaw (normal axis), pitch (transverse axis),

and roll (longitudinal axis). The LapSim provided the six requisite degrees of freedom with haptic feedback provided in all domains. The most telling and tangible measure of outcome will come with the next phase of this research, regarding how the 28 participants perform as defined by clinical practical PBA performance and ARCP outcome, regarding transferable operative technical skills.

UK NHS consultant appointments occur after defined training times; although some opinion contends that this may be shortened (20, 21), experience improves outcomes, and consultants will probably be appointed on an array of trajectories, in some short of the expert phase. In June 2016, the Economist surveyed the realm of artificial intelligence and the adjustment required of workers as jobs changed. It concluded that education and training must be made flexible enough to teach new skills quickly and efficiently, requiring weightier stress on lifelong on-the-job learning, with supplementary use of digital video-game-style simulation (22). Surgical skill set development rather than number acquisition should be the focus, allied to a will to move away from university teaching models that originated in medieval times, based around books and Socratic methods (23), to employ state-of-the-art online teaching strategies that develop universal skills. The findings of this study suggest that learning curve trajectories are susceptible to significant left shift, with implications for enhanced and shorter trainee time, boosted and more efficient trainer (instructor) time, and most of all a commensurate improvement in patient safety and satisfaction, especially during the early learning curve trajectory.

Author Disclosures

C Brown: no conflict of interest or disclosures

D Robinson: no conflict of interest or disclosures

R Egan: no conflict of interest or disclosures

L Hopkins: no conflict of interest or disclosures

T Abdelrahman: no conflict of interest or disclosures

A Powell: no conflict of interest or disclosures

MJ Pollitt: no conflict of interest or disclosures

WG Lewis: no conflict of interest or disclosures

References

1. Cheng, A., Grant, V. and Auerbach, M. Using Simulation to Improve Patient Safety. *JAMA Pediatrics*, 2015; 169(5), p.419
2. Patterson, M., Geis, G., Falcone, R., LeMaster, T. and Wears, R. In situ simulation: detection of safety threats and teamwork training in a high-risk emergency department. *BMJ Quality & Safety*, 2012; 22(6), pp.468-477.
3. *Flight Simulation and Synthetic Trainers. RAF Museum 2018*. [online] Available at: <https://www.rafmuseum.org.uk/research/online-exhibitions/taking-flight/historical-periods/flight-simulation-and-synthetic-trainers.aspx> [Accessed 14 Sep. 2018].
4. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J*. 2007;83 (986):777-779.
5. Joint Committee on Surgical Training (JCST). Guidelines for the Award of CCT in General Surgery. JCST: London, 2014.
6. Brown C, Abdelrahman T, Patel N, Thomas C, Pollitt MJ and Lewis WG. Operative learning curve trajectory in a cohort of surgical trainees. *British Journal of Surgery* 2017;104 (10), 1405-1411.
7. Surgical Science, Sweden. <https://surgicalseience.com/systems/lapsim/> (accessed 17th August 2018).
8. Jensen, K., Bjerrum, F., Hansen, H., Petersen, R., Pedersen, J. and Konge, L. Using virtual reality simulation to assess competence in video-assisted thoracoscopic surgery (VATS) lobectomy. *Surgical Endoscopy* 2016; 31(6), pp.2520-2528.
9. Ahlberg, G., Enochsson, L., Gallagher, A., Hedman, L., Hogman, C., McClusky, D., Ramel, S., Smith, C. and Arvidsson, D. Proficiency-based virtual reality training

- significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. *The American Journal of Surgery* 2007; 193(6), pp.797-804.
10. Larsen, C., Grantcharov, T., Schouenborg, L., Ottosen, C., Soerensen, J. and Ottesen, B. Objective assessment of surgical competence in gynaecological laparoscopy: development and validation of a procedure-specific rating scale. *BJOG: An International Journal of Obstetrics & Gynaecology* 2008; 115(7), pp.908-916.
11. Schmidt P, Ayres F. Schaum's outline of theory and problems of college mathematics: algebra, discrete mathematics, trigonometry, geometry, introduction to calculus. New York: McGraw-Hill 2nd ed.1992.
12. Abdelrahman T, Long J, Egan R, Lewis W. Operative Experience vs. Competence: A Curriculum Concordance and Learning Curve Analysis. *Journal of Surgical Education* 2016; 73(4), pp.694-698.
13. Youden WJ. Index for rating diagnostic tests. *Cancer* 1959; **3**: 32-35.
14. Siqueira JR, Gough MJ. Correlation between operative targets and competence for general surgical certification, *British Journal of Surgery* 2016; **103**: 921–927.
15. Thomas C, Elsey E, Boyce T, Catton J, Lewis W. Data from two deaneries regarding the ability of higher surgical training programmes to deliver curriculum targets. *Bull R Coll Surg Engl.* 2015;97(9):e50-e53.
16. Sutherland, L., Middleton, P., Anthony, A., Hamdorf, J., Cregan, P., Scott, D. and Maddern, G. Surgical Simulation. *Annals of Surgery* 2006; 243(3), pp.291-300.
17. Nagendran M, Gurusamy K, Aggarwal R, Loizidou M, Davidson B. Virtual reality training for surgical trainees in laparoscopic surgery. *Cochrane Database of Systematic Reviews* 2013.
18. Bjerrum F, Stranbygaard J, Rosthoj S, Grantcharov T, Ottesen B, Sorensen JL. Evaluation of procedural simulation as a training and assessment tool in general

surgery – simulating a laparoscopic appendectomy. *J Surg Education* 2017; 74 (2) 243-250.

19. Overtoom E, Horeman T, Jansen F, Dankelman J, Schreuder H. Haptic Feedback, Force Feedback, and Force-Sensing in Simulation Training for Laparoscopy: A Systematic Overview. *Journal of Surgical Education* 2019; 76(1), pp.242-261.

20. Improving Surgical Training: Trainee Prospectue 2018. Royal College of Surgeons of England.

21. Greenaway D. Shape of Training. Securing the Future of Excellent Patient Care: Final Report of the Independent Review Led by Professor David Greenaway; 2013.[http://www.shape oftraining.co.uk/static/documents/content/Shape_of_training_FINAL_Report.pdf_53977887.pdf](http://www.shapeoftraining.co.uk/static/documents/content/Shape_of_training_FINAL_Report.pdf_53977887.pdf)

22. The Economist. Artificial intelligence. March of the machines. June 25, 2016.

23. Frenk, J, Chen L, Bhutta Z, Cohen J, Crisp N, Evans T, Fineberg H, Garcia P, Ke Y, Kelley P, Kistnasamy B, Meleis A, Naylor D, Pablos-Mendez A, Reddy S, Scrimshaw S, Sepulveda J, Serwadda D, Zurayk H. Health professionals for a new century: transforming education to strengthen health systems in an interdependent world. *The Lancet* 2010; 376(9756), pp.1923-1958.

Figure Legends

Figure 1. Simulated laparoscopic appendicectomy, Surgical Sciences.

Figure 2. ROC curve of relationship between fine dissection task and median simulated appendicectomy composite score of $\geq 70\%$ (AUC = 0.803, $p=0.028$).

Figure 3. Fine dissection task learning curve (UQ: upper quartile, LQ: lower quartile).

Figure 4. Simulated laparoscopic appendicectomy learning curve (UQ: upper quartile, LQ: lower quartile).

Table Legends

Table 1. Individual laparoscopic virtual reality simulated task: total attempts and composite score. Values are median (IQR).

Table 2. Correlation between individual laparoscopic virtual reality simulated task score improvement ratio and overall simulated laparoscopic appendectomy performance score (Spearman's rho).

Table 3. Univariate analysis of individual parameters related to overall median LSA composite score of $\geq 70\%$.

Table 4. Laparoscopic simulated appendectomy score variation from 1st to 6th attempt respectively.