Visual behaviour in robotic surgery – demonstrating the validity of the simulated environment.

Article Authors

James Dilley[1*], Harsimrat Singh[1*], Philip Pratt[1], Ismail Omar[1], Ara Darzi[1], Erik Mayer[1]

Author Affiliations

[1]Department of Surgery and Cancer**,** Imperial College London, United Kingdom

*Authors contributed equally.

Conflicts of Interest and Sources of Funding

Sources of funding: This study is independent research funded by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Conflicts of interests and disclosures

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/rcs.2075

James Dilley, Harsimrat Singh, Philip Pratt, Ismail Omar, Ara Darzi, Erik Mayer have no conflicts of interests or financial ties to disclose

Key Words : Robotic, Simulation, Training, Eye metrics

Running head: Visual behaviour in robotic surgery

Corresponding Author & Request for Reprints

Mr Erik Mayer

Department of Surgery and Cancer

10th Floor, QEQM Wing

St Mary's Hospital, Praed Street

London W2 1NY

Phone: +44(0)20 3312 1310

Fax: +44(0)20 3312 6950

Email: e.mayer@imperial.ac.uk

Structured Abstract

Background

Eye-metrics provide insight into surgical behavior allowing differentiation of performance; however have not been used in robotic surgery. This study explores eye-metrics of robotic surgeons in training in simulated and real tissue environments.

Methods

Following the FRS training curriculum novice robotic surgeons were trained to expert derived benchmark proficiency using real tissue on the da Vinci Si and the dVSS simulator. Surgeons eye-metrics were recorded using eye tracking glasses when both 'novice' and 'proficient' in both environments. Performance was assessed using GEARS and NPMTS scores.

Results

Significant($p \leq 0.05$) correlations were seen between pupil size, rate of change and entropy and associated GEARS/NPMTS in 'novice' and 'proficient' surgeons. Only number of blinks per minute was significantly different between pupilometrics in the simulated and real tissue environments.

Conclusions

This study illustrates the value of eye-tracking as an objective physiological tool in the robotic setting. Pupilometrics significantly correlate with established assessment methods and could be incorporated into robotic surgery assessments.

Introduction

The popularity and evidence supporting a robotically assisted approach in modern minimal access surgery is ever increasing(1)(2). In order to ensure safe and effective use specialist and advanced skills need to be developed and achieved. Simulator environments for training have been increasingly been used and valued for this purpose (3) with several virtual reality robotic simulators available commercially (4). However, there is a lack of standardisation in outcome metrics and exercises across different platforms to measure efficiency and performance of surgical skill(5). Furthermore there is paucity of objective assessment for the surgeon both in terms of technical performance longitudinally (before and after using these environments) and the subsequent translation into the clinical environment. Currently there is no method to report evidence of skill transfer from simulation to clinical surgery on real patients (4). This makes it difficult to justify the considerable expense incurred when training takes place *in vivo* (6) and hence, hugely important that the simulated environment is representative of the skill transfer from simulated to real environments.

The surgical community has acknowledged the need for an approved curricula to allow global standardisation of training and subsequent certification in the form of the

Fundamentals of Robotic Surgery (FRS) programme (7)(8). In order to introduce objective assessment of competency in reliable and reproducible manner, robotic surgery has remodeled scoring systems akin to laparoscopic surgery (9) such as Global Evaluative Assessment of Robotic skills (GEARS) (10) and Robotic Skills Assessment Score (RSA-Score) (11). The recently published study proving the effectiveness of the FRS training curriculum also employed a tasks specific 32 point checklist numeric psychomotor test score (NPMTS)(12). This can help provide a quantitative assessment, to which there is an increasing body of evidence suggesting superiority compared to a Likert based assessment(13).

Eye-tracking is an established quantitative objective tool in the assessment of surgical skill (14)(15). Observation of the gaze behaviour in surgeons has made it possible to differentiate between gaze patterns of novices and experts (16)(17)(18), simulated and live operations (19) as well as giving valuable information regarding the surgeon's cognitive load (20). Original work by Hess and Polt (21) found that mean pupil dilation was a function of the level of difficulty, with subsequent studies of pupilometrics showing that pupil diameter increases with increasing task demands (22). Further work has has shown that the pupil entropy is also higher when experiencing higher cognitive demand.(23) Blinks are known to be an indicator of visual attention and mental stress, with surgeons blinking less frequently when experiencing a higher workload and levels of frustration than those who are not(20). However, studies to date predominately focused on the traditional laparoscopic environment. The gaze behaviour in the robotic environment is poorly understood, mainly due to hardware limitations. Isolated studies (24) have suggested that the visual behavior of surgeons in the robotic environment is different to that in a laparoscopic environment. It has been suggested that this may be in response to the lack of haptic feedback in the robotic environment, forcing the surgeon to rely on visual cues to act as a substitute (25). It is therefore pertinent that we understand the parameters associated with eye tracking in the robotic environment.

We hypothesize that there are apparent differences in the eye tracking parameters both when performing complex minimal access tasks in the simulated robotic and real tissue environments and as the surgeons' skills develop from being a novice. This study had two objectives. Firstly to investigate the visual patterns during the 'learning process' of a novice surgeon undergoing simulator-based training to becoming a 'proficient' operator. Secondly, to compare the differences in gaze behaviour highlighting the adjustments a surgeon has to make, when performing the identical complex task in a robotic simulated environment and in a real tissue robotic environment. Understanding the potential differences between real tissue and computer-simulated platforms will be valuable in design and help understand the importance of simulation and improvement in training programs.

Materials and Methods

Participants

A total of 21 trainee surgeons (17 male, 4 female) participated in the study. All surgeons were classified as novice robotic surgeons (5 or fewer procedures as primary console surgeon). Institutional ethical approval was obtained before commencement of the study and all surgeons provided informed consent. All participants completed the Fundamentals of Robotic Surgery (FRS) online learning curriculum (http://frsurgery.org/frs-curriculum) and were given a tutorial on how to operate the da Vinci Si robot and da Vinci skills simulator (dVSS) (Intuitive Surgical, Sunnyvale, CA, USA).

Surgical systems and task

The study followed the FRS ‘validation for novice’ protocol. At the beginning of the training day, each participant completed the continuous suturing task on an avian tissue model on the da Vinci Si as a ‘novice’. All participants then underwent a period of training using the dVSS simulator over the course of the day. Participants were required to perform continuous suturing task on the dVSS on multiple occasions until they all achieved the expert-derived benchmark of proficiency threshold twice consecutively or five times non-consecutively. They were then subsequently deemed 'proficient'. This expert-derived benchmark of proficiency threshold was predetermined and set by work from the Fundamentals of robotic surgery validation trial(12)(26). To establish this benchmark, expert robotic surgeons were asked to complete the continuous suturing task on a simulated platform until no improvement was seen on 2 consecutive attempts. At the end of the training, each participant completed one further 'proficient' continuous suturing task on an avian tissue model using the da Vinci Si.

Measures

Eye tracking data

Eye movements were recorded using the mobile eye tracking glasses (SMI Eye Tracking Glasses 2.0, SensoMotoric Instruments, Germany). The glasses had a sampling rate of 30Hz and a resolution of 1280 x 960p at 24 frames per second. A three-point calibration was performed on initial set up to calibrate the glasses. The glasses were securely fastened to the participant's head using the draw cord and surgical tape to minimise movement (figure 1a). The participant's head was secured to the console using a Velcro band to ensure that the head did not move in relation to the console and that the light levels remained constant throughout the study (figure

1b). Eye tracking data was collected from the participants performing the suturing exercise on both the avian tissue and on the dVSS simulation when in the novice and proficient phases.

Task performance

Video of all of the tasks that each participant performed was acquired. Each participant's performance was scored using the GEARS score and the NPMTS. Two independent surgeons experienced in using the aforementioned scoring systems carried this out. An average of both scores was taken to generate one solid score for each participant, which was subsequently used in analysis.

Analysis

Raw data files extracted from the BeGaze Software (version 3.5, SensoMotoric Instruments, Germany) were processed using bespoke MATLAB based functions. X-Y Coordinates for successive samples were subtracted to estimate linear displacement for the eye movement. Fixations were identified when the displacement fell below a certain radius threshold and stayed below the threshold for at least a minimum duration (100ms). Fixations, blinks, saccades were computed to formulate a timeline. Data files from all subjects were collated into sessions as i) novice 'real tissue', ii) novice simulator, iii) proficient simulator and iv) proficient 'real tissue'. A database of eye tracking parameters was then computed for each session for each subject.

Pupilometrics and entropy (measure of predictability of pupil change) parameters were computed based on fixations and fixation times(27). Subjects were excluded from analysis due to errors in capturing data (inaccurate calibration, incomplete task recording), using the eye tracking device.

Statistical analysis was performed using PRISM 7.0a (GraphPad software, La Jolla, Ca, USA) employing non-parametric tests. Comparison of eye tracking data between the four study groups; novice real tissue (NRT), proficient real tissue (PRT), novice simulator (NS), proficient simulator (PS) was undertaken using the Wilcoxon matched pairs signed rank test. Correlation between GEARS and NPMTS and eye tracking data was undertaken to using nonparametric Spearman rank correlation coefficient. Two tailed p-value was calculated and deemed significant when the value was reported as <0.05.

Results

Demographic data

The participant's median age was 32 (IQR 4) years old; 19 participants were right hand dominant, 2 were left hand dominant. Participants were on average 7 (IQR 3) years post qualification, had performed 50 (IQR 95) laparoscopic cases as primary surgeon and 100 (IQR 150) cases as a secondary surgeon. None of the participants had performed any cases as primary surgeon on the da Vinci robot (table 1).

Task performance

Agreement between the two surgeons, independently assessing performance scores was good (Cronbach's α; NPMTS = 0.99, GEARS = 0.98). There was a significant difference (p=0.0016) between the GEARS scores of NRT and PRT (17, IQR 2.75 vs. 23.5, IQR 5.75). No difference was found between NPMTS (8, IQR 3 vs. 7.5, 3.75). No significant differences were seen between NS and PS in either GEARS score (18, IQR 8 vs. 18, IQR 5) or NPMTS (31.5, IQR 23 vs. 26, IQR 15).

Eye tracking data

Comparison of 4 study groups

The median and IQR of all 4 groups can be seen in table 2. There was a significant difference in the number of blinks per minute between the NRT and PRT (p = 0.0034, NRT 20.6 vs. PRT 27.4). Between NS and PS there was a significant difference between rate of pupil change of the right eye (p = 0.0419, NS 0.0041 vs. PS 0.0065). No significant differences were seen between NRT and NS. Comparison of PRT and PS demonstrated a significant difference in the number of blinks per minute (p = 0.0023 PRT 27.4 vs. PS 25.28) (Table 3)

Correlation with GEARS and NPMTS

Novice and proficient real tissue

In the NRT group there was significant correlation between NPMTS and average fixation duration (p = 0.02, R = -0.54), total number of blinks (p = 0.012, R = 0.57) and number of blinks per minute (p = 0.03, R= 0.52) (Table 4).

In the PRT group there was significant correlation between GEARS and right pupil size entropy (p = 0.03, R = -0.50) (Table 4)

Novice and proficient simulator

The NS group showed statistically significant correlations between NPMTS in left pupil average size (p = 0.006, R = 0.62), both pupils average size (p = 0.042, R = 0.484), left eye rate of pupil change (p = 0.011, R = -0.58), right eye rate of pupil change (p = 0.014, R = -0.57), left pupil size entropy (p = 0.03, R = 0.50), total number of blinks ($p = <0.001$, $R = 0.75$)(figure 2) and between GEARS and total number of blinks ($p = 0.045$, $R = -0.48$) (Table 4).

In the PS group significant correlation was seen in both the NPMTS and GEARS with saccade frequency (NPMTS $p = 0.04$, $R = -0.48$) (GEARS $p = 0.04$, $R = 0.47$) average fixation duration (NPMTS $p = 0.02$, $R = -0.64$) (GEARS $p = 0.026$, $R = 0.62$), right eye rate of pupil change (NPMTS $p = 0.016$, $R = -0.66$) (GEARS $p = 0.001$, $R = 0.82$) and left pupil size entropy (NPMTS $p = 0.05$, $R = 0.55$) (GEARS $p = 0.036$, $R = -0.59$) (figure 2). Significant differences were seen between NPMTS and right pupil maximum size ($p = 0.026$, $R=0.62$) and the left pupil average size ($p = 0.05$, $R = 0.56$) (Table 4).

Discussion

This paper presents the first eye-tracking study to compare eye metrics of novice robotic surgeons trained to proficiency when performing tasks in simulated and real tissue setting. The study sought to investigate the visual and cognitive behaviour of the surgeon during the learning process. The findings provide important evidence that pupilometrics and can be used to measure surgeon's development from a novice to a proficient operator in the robotic environment. In addition, the pupilometrics data strongly correlates with the objective measures of robotic performance, GEARS and NPMTS. The second objective of the study was to compare the differences in gaze behaviour between the simulated and real tissue environments in the robotic setting. The lack of significant differences between the surgeons behaviour in the two environments suggest that the surgeon behaviour is similar in both environments. This promotes the fidelity of the simulated environment helping to prove its value in training and transfer of skills. Finally, this study shows the value of the FRS training programme in its ability to increase the performance of the surgeon using an objective scoring system.

When examining the correlation between pupilometrics and the NPMTS and GEARS, a strong correlation is found in novice and proficient participants in terms of pupil size, rate of diameter change and entropy. In participants who performed worst (higher numeric psychomotor metric test scores) the average and maximum pupil size was larger, the rate of pupil change was faster and the entropy increased. These findings are supported by Hess and Polt (21) who found that mean pupil dilation was a function of the level of difficulty. More recent research has shown that the rate of change of the pupil diameter matches well to the change of precision requirement and difficulty of a surgical task and the surgeon’s mental workload (28). Previous studies have shown there is an increase in entropy in less experienced operators indicating a higher cognitive demand and reflects that the surgeon is processing more information which makes their pupil change less predictable (23). This is an important finding as it has validated the use of a constant psychophysiologic measurement in the robotic environment avoiding the subjective bias or interruption of other workload measures (29).

There was only one significant difference when comparing surgeons who performed tasks in both the simulated and real tissue environments as both a novice and proficient. This was found to be in the number of blinks per minute, which was higher in the proficient real tissue group than simulator group. Previous studies have shown (20) that a more infrequent blink frequency is associated with increase frustration and workload. This may suggest that after training in a high fidelity simulation, surgeons found it less demanding subsequently performing in the real environment. However, apart from this difference, the overall lack of significant differences in the participant’s behaviour between the two environments would suggest the surgeon behaves similarly in both. This finding can provide both researchers and those involved with training confidence that the response and behaviour seen in the surgeon in the simulated environment should be very similar to that seen in real tissue environment. These findings differ to those of Vine et al. (19) who found that when performing a procedure on real tissue the surgeons displayed more fixations of a shorter duration. However, in this study they acknowledge that as the real procedure was performed in an operating theatre the additional complexity of this environment may have influenced the difference in outcomes.

Previous studies comparing the visual behaviour between novices and experts (23)(16) have found differences between the two in terms of higher fixation rates. This was not seen in this study. The absence of this finding may be due to the fact that although surgeons were deemed to be proficient, they still had not reached the level of expert and as such not undergone the accompanying changes in visual behaviour. Statistically significant differences were seen with more blinks per minute in the real tissue groups and the slower rate of pupil diameter change in the simulator group in proficient operators. Both of these suggest an alteration in cognitive ability and workload which one would expect to see as the surgeon becomes more experienced (28). These findings of a stepwise changes in visual behaviour from novice to expert are valuable in both assessing surgeons and identifying where the surgeon is in their learning curve (30)(31).

This study has several strengths. Firstly it uses a robust study design and approved training curriculum. The study uses two separate objective scoring systems, and shows strong correlation between them. Participants were able to perform the exercise on both the real tissue and simulator using the same console minimising the differences between the groups. However participants numbers in this study were limited with participants only performing one task, as such future studies should look at performing tasks of different complexities(32) to understand further the relationship in the robotic environment. The study only captured eye metrics when the participant was novice or proficient. Future studies should capture eye metrics through the training to observe the learning curve. This study did not record any subjective measures of the participant’s workload to compare to the objective eye tracking data but the relationship between workload score, cognitive function and performance are well established. One criticism that could be levelled at this study is the use of a mobile eye-tracking device for a screen-based task, however due to the set up of the da Vinci console no alternative commercial screen-based eye tracker is currently available. Additionally although the glasses are lightweight, they might have been an unfamiliar addition to the surgeon, which may have caused a distraction leading to an altered performance. The purpose of this study was to explore the potential value of eye tracking in the simulated and real tissue of robotic environment. Results from the study would suggest that it does have a role in assessing training. However it is likely that the method used in this study to capture the data are too cumbersome in their present state and would need refining to allow integration into training assessment and surgical workflow. Finally for some participants the eye tracking data recorded was not of sufficient quality and to ensure integrity of the analysis was excluded. This had the disadvantage of reducing the numbers involved with the study.

Conclusion

This study provides longitudinal assessment of a novice surgeon's training using the simulator platforms and demonstrates the potential value in using eye tracking as an objective physiology measure in robotic surgery training. The results contribute towards validating the use of simulator based training for complex robotic procedures on patients, as there are non-significant changes in visual parameters between simulator and real tissue but a significant improvement in visual behaviour of a novice untrained robotic surgeon to a proficient trained surgeon. The study provides a robust evidence for developing programmes for the use of simulator technology for surgical skills training.

Contributorship

The study design was conceived and developed by JD, HS, AD and EM. JD and HS executed the experiment and collected eye tracking data. Data pre-processing was conducted by HS and JD. GEARS and numeric psychomotor metric test scoring was performed by JD and IO. Data interpretation was performed by HS, JD in consultation with PP, EM, and AD. The manuscript was drafted by JD, HS, EM, PP and AD. Critical editing of the manuscript was performed by PP, EM and AD.

Conflicts of interests and disclosures

James Dilley, Harsimrat Singh, Philip Pratt, Ismail Omar, Ara Darzi, Erik Mayer have no conflicts of interests or financial ties to disclose

References

1. Wright JD, Ananth C V, Lewin SN, Burke WM, Lu Y, Herzog TJ, et al. Robotically Assisted vs Laparoscopic. JAMA. 2013;309(7):689–98.
2. McCullough TC, Barret E, Cathelineau X, Rozet F, Galiano M, Vallancien G. Role of robotics for prostate cancer. Curr Opin Urol. 2009;19(1).
3. Gaba D. The future vision of simulation in health care. Qual Saf Heal Care. 2004;13(Suppl)(1):2–10.
4. Moglia A, Ferrari V, Morelli L, Ferrari M, Mosca F, Cuschieri A. A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery. Eur Urol. 2016;69(6):1065–80.
5. Abboudi H, Khan MS, Aboumarzouk O, Guru KA, Challacombe B, Dasgupta P, et al. Current status of validation for robotic surgery simulators a systematic review. BJU Int. 2013;111(2):194–205.
6. Murphy D, Sundaram C. Comparative assessment of three standardized robotic surgery training methods. BJU Int. 2013;112(6):864–71.
7. Fisher RA, Dasgupta P, Mottrie A, Volpe A, Khan MS, Challacombe B, et al. An over-view of robot assisted surgery curricula and the status of their validation. Int J Surg. 2015;13:115–23.
8. Smith R, Patel V, Satava R. Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. Int J Med Robot. 2014;10:379–84.
9. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic

skills. Am J Surg. 2005;190(1):107–13.

10. Goh AC, Goldfarb DW, Sander JC, Miles BJ, Dunkin BJ. Global evaluative assessment of robotic skills: Validation of a clinical assessment tool to measure robotic surgical skills. J Urol. 2012;187(1):247–52.

11. Chowriappa AJ, Shi Y, Raza SJ, Ahmed K, Stegemann A, Wilding G, et al. Development and validation of a composite scoring system for robot-assisted surgical trainingdthe Robotic Skills Assessment Score. J Surg Res. 2013;185(2):561–9.

12. Satava RM, Stefanidis D, Levy JS, Smith R, Martin JR, Monfared S, et al. Proving the Effectiveness of the Fundamentals of Robotic Surgery (FRS) Skills Curriculum. Ann Surg. 2019;epub(epub):1.

13. Gallagher AG, O'Sullivan GC, Leonard G, Bunting BP, Mcglade KJ. Objective structured assessment of technical skills and checklist scales reliability compared for high stakes assessments. ANZ J Surg. 2014;84(7–8):568–73.

14. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang G-Z, Darzi A. Eye tracking for skills assessment and training: a systematic review. J Surg Res. 2014 Sep;191(1):169–78.

15. Ashraf H, Sodergren MH, Merali N, Mylonas G, Singh H, Darzi A. Eye-tracking technology in medical education: A systematic review. Med Teach. 2017;0(0):1–8.

16. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. Ann Surg.

2010;252(1):177–82.

17. Kocak E, Ober J, Berme N, Melvin WS. Eye Motion Parameters Correlate with Level of Experience in Video-Assisted Surgery: Objective Testing of Three Tasks. J Laparoendosc Adv Surg Tech. 2005;15(6):575–80.

18. Harvey A, Vickers JN, Snelgrove R, Scott MF, Morrison S. Expert surgeon's quiet eye and slowing down: expertise differences in performance and quiet eye duration during identification and dissection of the recurrent laryngeal nerve. Am J Surg. 2014 Feb;207(2):187–93.

19. Vine SJ, McGrath JS, Bright E, Dutton T, Clark J, Wilson MR. Assessing visual control during simulated and live operations: gathering evidence for the content validity of simulation using eye movement metrics. Surg Endosc. 2014 Jun;28(6):1788–93.

20. Zheng B, Jiang X, Tien G, Meneghetti A, Panton ONM, Atkins MS. Workload assessment of surgeons: Correlation between NASA TLX and blinks. Surg Endosc Other Interv Tech. 2012;26(10):2746–50.

21. Hess E., Polt J. Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Sci ,. 1964;143:1190-1192.

22. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. Psychon Bull Rev. 2018;1–11.

23. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang GZ, Darzi A. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. Surg Endosc. 2014;405–13.

24. Moore LJ, Wilson MR, McGrath JS, Waine E, Masters RSW, Vine SJ. Surgeons' display reduced mental effort and workload while performing robotically assisted surgical tasks, when compared to conventional laparoscopy. Surg Endosc. 2014;

25. Hagen ME, Meehan JJ, Inan I, Morel P. Visual clues act as a substitute for haptic feedback in robotic surgery. Surg Endosc Other Interv Tech. 2008;22(6):1505–8.

26. Fundamentals of robotic surgery. http://frsurgery.org/fifteen-institutions-selected-to-participate-in-the-validation-study-for-fundamentals-of-robotic-surgery/No Title. 2015.

27. Moddemeijer R. On Estimation of Entropy and Mutual Information of Continuous Distributions. Signal Processing. 1989;16(3):233–46.

28. Jiang X, Zheng B, Tien G, Atkins MS. Pupil response to precision in surgical task execution. Stud Health Technol Inform. 2013;184(September 2017):210–4.

29. Carswell CM, Clarke D, Seales WB. Assessing Mental Workload During Laparoscopic Surgery. Surg Innov. 2005;12(1):80–90.

30. Andolfi C, Umanskiy K. Mastering Robotic Surgery: Where Does the Learning Curve Lead Us? J Laparoendosc Adv Surg Tech. 2017;27(5):470–4.

31. Yohannes P, Rotariu P, Pinto P, Smith AD, Lee BR. Comparison of robotic versus laparoscopic skills: is there a difference in the learning curve? Urology. 2002;60(1):39–45; discussion 45.

32. Zheng B, Jiang X, Atkins MS. Detection of Changes in Surgical Difficulty: Evidence From Pupil Responses. Surg Innov. 2015;22(6):1–7.

Surgeon demographics	
Age (median, IQR)	32, 4
Hand dominance	19 Right, 2 Left
Postgraduate years of training (median, IQR)	7, 3
Number of laparoscopic cases as primary surgeon (median, IQR)	50, 95
Number of laparoscopic cases as secondary surgeon (median, IQR)	100, 150

Table 1 – Demographics of participants

Figure 1a - Photograph showing eye-tracking glasses secured to participants head and Velcro band to secure participants to console. Figure 1b - Participant at da Vinci console with head secured in console using Velcro straps.

Eye tracking data	Novice real tissue (n=18)		Proficient real tissue (n= 19)		Novice simulator (n=19)		Proficient simulator (n=14)	
	Median	IQR	Median	IQR	Median	IQR	Median	IQR
Event duration (seconds)	515.68	193.74	376.88	176.24	577.95	257.07	473.37	361.12
Time to perform task (minutes)	8.59	3.23	6.28	2.94	9.63	4.28	7.89	6.02
Number of fixations	1304.00	786.75	746.00	701.00	1352.00	650.00	1065.00	1161.25
Average fixation duration	0.15	0.04	0.14	0.07	0.15	0.06	0.14	0.05
Maximum fixation duration	1.60	0.91	1.43	1.21	1.46	1.40	1.90	0.66
Fixation frequency (number of fixations per second)	2.57	0.77	2.28	1.33	2.87	1.15	2.66	1.28
Number of saccades (total)	1209	460	808	423	1540	643	1213	1063
Length of saccades (cumulative)	184601	350361	239140	2838343	309253	2427792	631864	2041391
Saccade frequency (number per second)	2.59	0.64	2.22	0.59	2.85	0.37	2.58	0.37
Saccade average velocity	1769	1871	3298	15209	2040	11691	3877	16668
Saccade average durations (seconds)	0.07	0.03	0.09	0.08	0.08	0.08	0.08	0.09
Pupil maximum size (left eye)	6.94	1.81	6.38	1.83	6.64	0.99	7.26	1.09
Pupil max size right eye (mm)	6.51	0.79	6.68	1.20	7.34	1.41	7.15	1.04
Pupil Av size left eye (mm)	3.71	0.88	3.64	0.87	3.48	1.01	3.35	0.55
Pupil Av size right eye (mm)	3.75	0.93	3.65	0.81	3.27	0.96	3.11	0.41
Pupil average size both eyes (mm)	3.62	1.02	3.73	0.99	3.25	0.96	3.26	0.70
Rate of pupil diameter change for left eye (mm/s)	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01
Rate of pupil diameter change right eye (mm/s)	0.01	0.01	0.01	0.00	0.00	0.00	0.01	0.01
Pupil size entropy left eye	4.63	0.81	4.48	0.66	5.13	0.60	4.72	0.73
Pupil size entropy right eye	4.87	0.48	4.58	0.72	4.83	0.64	4.81	0.46
Number of blinks over total time period	188.50	246.75	123.00	176.00	227.00	277.00	158.00	335.75
Number of blinks per minute	20.60	25.89	27.43	25.31	22.70	20.87	25.28	20.08

Numeric psychomotor metric test score	8.00	3.00	7.50	3.75	31.50	23.00	26.00	15.00
GEARS score	17.00	2.75	23.50	5.75	18.00	8.00	18.00	5.00

Table 2- Median and IQR range of all participants eye tracking data across the four groups of the study.

Non parametric, paired, Wilcoxon Test (* indicates statistical significance)				
Comparison groups	Novice vs Proficient Real tissue p-value	Novice vs Proficient Simulator p-value	Novice Real tissue vs Novice simulator p-value	Proficient real tissue vs Proficient simulator p-value
Event duration (seconds)	*0.0066**	0.3575	n/a	n/a
Number of fixations	*0.0066**	0.1888	n/a	n/a
Average fixation duration	0.6095	0.5416	0.4171	0.4631
Maximum fixation duration	0.6095	0.5416	0.7421	0.2676
Fixation frequency (number of fixations per second)	0.0987	0.7148	0.7019	0.5016
Number of saccades (total)	*0.0023**	0.6698	n/a	n/a
Length of saccades (cumulative)	0.8317	0.583	0.1815	0.9515
Saccade frequency (number per second)	0.0987	0.8552	0.1964	0.0906
Saccade (maximum velocity/ pixels per second)	0.0898	0.7609	0.1415	0.9032
Saccade average velocity	0.5798	0.6698	0.2462	0.6257
Saccade average durations (seconds)	0.2837	>0.9999	0.9661	0.4631
Pupil maximum size (left eye)	0.6397	0.2412	0.4951	0.7148
Pupil maximum size right eye (mm)	0.8403	0.7609	0.3289	0.5516
Pupil average size left (mm)	0.865	0.7609	0.7019	0.2412
Pupil average size right eye (mm)	0.766	0.9515	0.2121	0.0785
Pupil average size both eyes (mm)	0.766	0.6698	0.2645	0.2166
Rate of pupil diameter change for left eye (mm/s)	0.3927	0.4263	0.0539	0.1726
Rate of pupil diameter change right eye (mm/s)	0.4423	*0.0419**	0.4423	0.2958
Pupil size entropy left eye	0.1674	0.1937	0.154	0.3258
Pupil size entropy right eye	0.0987	0.5016	0.4423	0.1937
Number of blinks over total time period	0.3927	>0.9999	0.4951	0.6367
Number of blinks per minute	*0.0034**	0.9032	0.8986	*0.0023**

Numeric psychomotor metric test score (NPMTS)	0.4095	0.385
GEARS score	0.0016*	0.8677

Table 3 – Statistical comparisons of eye tracking data between study groups. For tasks compared carried out in real tissue and simulator certain parameters cannot be compared and represented as n/a.

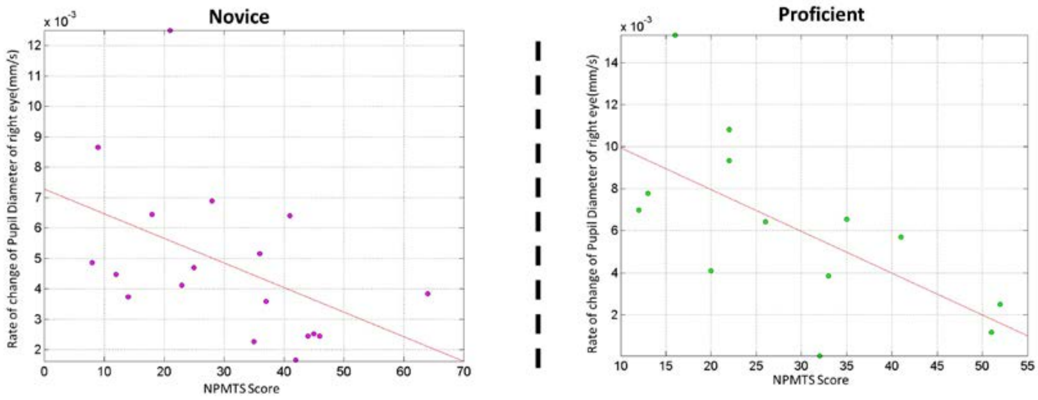
Correlation using Spearman R																
	Correlation novice real tissue				Correlation proficient real tissue				Correlation novice simulator				Correlation proficient simulator			
	p-value *		Correlation R		p-value *		Correlation R		p-value *		Correlation R		p-value *		Correlation R	
Eye tracking data	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS	NPMTS	GEARS
Numeric psychomotor metric test score (NPMTS)		0.93		-0.022		0.049		-0.46		0.040		-0.490		0.088		-0.494
GEARS score	0.93		-0.022		0.049		-0.46		0.04		-0.49		0.088		-0.494	
Event duration (seconds)	0.463	0.964	0.185	-0.012	0.571	*0.034	0.143	-0.501	*<0.001	*0.01	0.86	-0.56	*0.005	*0.006	0.74	-0.73
Time to perform task (minutes)	0.463	0.964	0.185	-0.012	0.571	*0.034	0.143	-0.501	*<0.001	*0.01	0.86	-0.56	*0.005	*0.006	0.74	-0.73
Number of fixations	0.997	0.491	-0.001	-0.174	0.652	0.765	0.114	-0.076	*0.007	*0.03	0.61	-0.53	0.44	0.46	0.23	-0.22
Average fixation duration	*0.02	0.323	-0.540	-0.247	0.423	0.912	-0.201	0.028	0.54	0.49	-0.15	-0.17	*0.02	*0.03	-0.64	0.62
Maximum fixation duration	0.226	0.824	-0.300	-0.057	0.137	0.961	-0.364	0.012	0.78	0.28	-0.07	-0.27	0.46	0.59	-0.22	0.16
Fixation frequency (number of fixations per second)	0.334	1.000	-0.242	0.000	0.765	0.411	0.076	0.207	0.17	0.95	-0.34	0.01	0.13	0.92	-0.45	0.03
Number of saccades (total)	0.524	0.537	0.161	-0.156	0.694	0.494	0.100	-0.172	*0.006	0.11	0.61	-0.39	*0.009	*0.003	0.70	-0.76

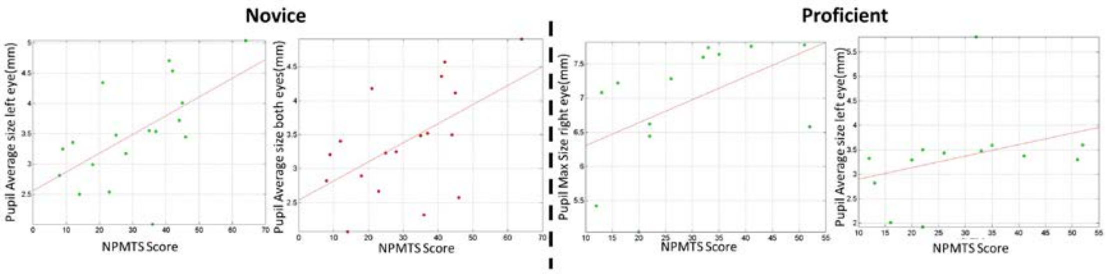
Length of saccades (cumulative)	0.880	0.785	0.038	-0.069	0.691	0.376	0.101	-0.222	0.25	0.67	0.29	0.11	*0.03	0.32	0.59	-0.30
Saccade frequency (number per second)	0.715	0.798	-0.092	0.065	0.771	0.132	-0.074	0.368	*0.04	*0.04	-0.48	0.47	0.77	0.16	0.09	-0.42
Saccade (maximum velocity/ pixels per second)	0.971	0.360	0.009	-0.229	0.753	0.557	0.080	-0.148	0.41	1.00	0.21	0.00	0.09	0.61	0.49	-0.16
Saccade average velocity	0.896	0.833	0.033	-0.053	0.753	0.347	0.080	-0.236	0.89	0.57	0.04	0.14	0.11	0.79	0.46	-0.08
Saccade average durations (seconds)	0.694	0.915	0.100	0.027	0.451	0.068	0.190	-0.440	0.17	0.81	0.34	0.06	0.19	0.49	0.39	-0.21
Pupil maximum size left eye (mm)	0.685	0.793	0.103	0.066	0.105	0.572	0.394	-0.143	0.73	0.94	0.09	-0.02	0.07	0.25	0.53	-0.34
Pupil max size right eye (mm)	0.124	0.646	0.376	0.116	0.125	0.315	0.375	-0.251	0.42	0.97	0.20	-0.01	*0.03	0.54	0.62	-0.19
Pupil average size left eye (mm)	0.438	0.895	0.195	0.033	0.948	0.340	0.017	-0.239	*0.006	0.24	0.62	-0.29	*0.05	0.16	0.56	-0.41
Pupil average size right eye (mm)	0.694	0.776	0.100	0.072	0.565	0.841	0.145	0.051	0.09	0.23	0.41	-0.30	0.65	0.41	-0.14	-0.25
Pupil average size both eyes (mm)	0.605	0.713	0.131	0.093	0.737	0.468	0.085	-0.183	*0.04	0.32	0.48	-0.25	0.39	0.26	0.26	-0.34
Rate of pupil diameter change for left eye (mm/s)	0.513	0.918	-0.165	-0.026	0.889	0.568	0.035	0.144	*0.01	0.24	-0.58	0.29	0.47	0.32	-0.22	0.30
Rate of pupil diameter change right eye(mm/s)	0.251	0.512	-0.285	-0.165	0.756	0.193	-0.079	0.322	*0.01	0.38	-0.57	0.22	*0.02	*<0.001	-0.66	0.82
Pupil size entropy left eye	0.778	0.186	-0.072	0.326	0.113	0.993	-0.387	0.002	*0.03	0.90	0.50	0.03	*0.05	*0.04	0.55	-0.59
Pupil size entropy right eye	0.990	0.593	-0.003	0.135	0.330	*0.036	0.244	-0.497	0.12	0.97	0.38	-0.01	0.91	0.19	-0.04	-0.39
Number of blink over total time period	*0.013	0.974	0.575	-0.008	0.157	0.264	0.348	-0.278	*<0.001	*0.04	0.75	-0.48	0.48	0.38	0.21	-0.27
Blinks per minutes	*0.03	0.689	0.521	0.101	0.167	0.291	0.340	-0.264	0.08	0.28	0.42	-0.27	0.77	0.55	0.09	-0.18

* indicates statistical significance

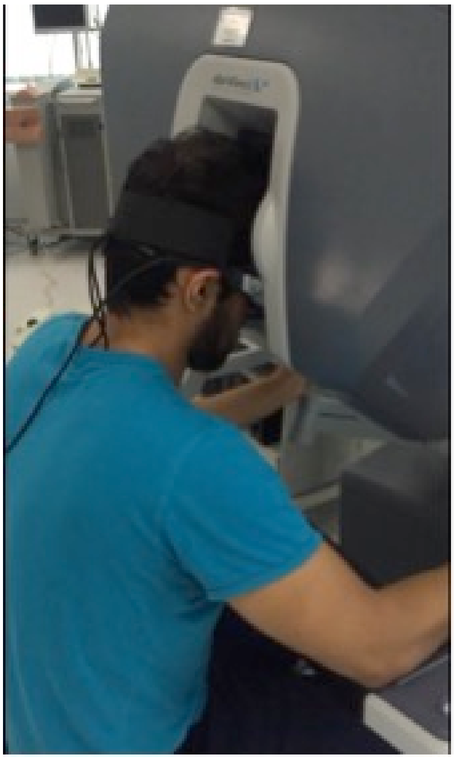
Table 4 – Correlations between Numeric psychomotor metric test score (NPMTS) and GEARS score and eye tracking data in the four study groups.

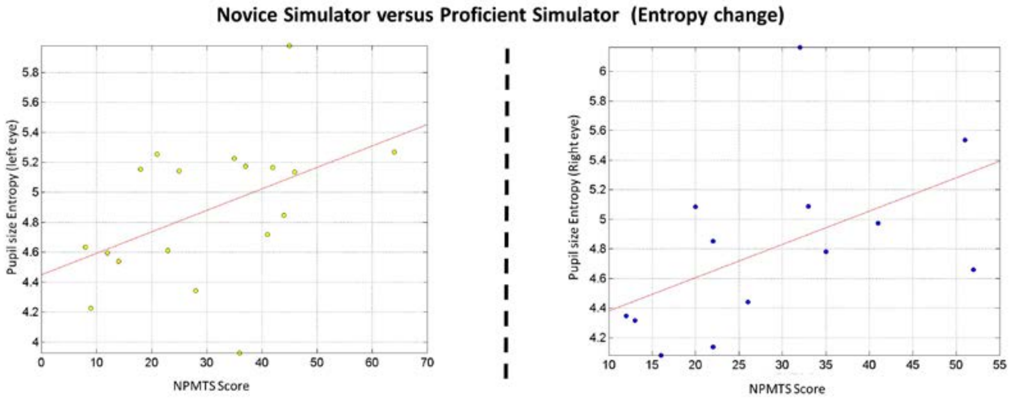
Figure 2. a) Rate of pupil diameter change in right eye NS and PS vs NPMTS; b) Pupil average size in NS and PS groups vs NPMTS; c) Entropy in NS and PS vs NPMTS; d) GEARS score vs rate of pupil change and entropy in PS.


Novice
x 10^-3
Rate of change of Pupil Diameter of right eye(mm/s)
NPMTS Score
Proficient
x 10^-3
Rate of change of Pupil Diameter of right eye(mm/s)
NPMTS Score



Novice
Proficient
Pupil Average size left eye(mm)
Pupil Average size both eyes(mm)
Pupil Max Size right eye(mm)
Pupil Average size left eye(mm)
NPMTS Score
NPMTS Score
NPMTS Score
NPMTS Score





Novice Simulator versus Proficient Simulator (Entropy change)
Pupil size Entropy (left eye)
NPMTS Score
Pupil size Entropy (Right eye)
NPMTS Score

Correlations between GEARS Score and Pupilometrics in Proficient Simulator