Attainment and retention of force moderation following laparoscopic resection training with visual force feedback

Rafael Hernandez[1] · Arzu Onar-Thomas[2] · Francesco Travascio[1] · Shihab Asfour[3]

Abstract

Background Laparoscopic training with visual force feedback can lead to immediate improvements in force moderation. However, the long-term retention of this kind of learning and its potential decay are yet unclear.

Methods A laparoscopic resection task and force sensing apparatus were designed to assess the benefits of visual force feedback training. Twenty-two male university students with no previous experience in laparoscopy underwent relevant FLS proficiency training. Participants were randomly assigned to either a control or treatment group. Both groups trained on the task for 2 weeks as follows: initial baseline, sixteen training trials, and post-test immediately after. The treatment group had visual force feedback during training, whereas the control group did not. Participants then performed four weekly test trials to assess long-term retention of training. Outcomes recorded were maximum pulling and pushing forces, completion time, and rated task difficulty.

Results Extreme maximum pulling force values were tapered throughout both the training and retention periods. Average maximum pushing forces were significantly lowered towards the end of training and during retention period. No significant decay of applied force learning was found during the 4-week retention period. Completion time and rated task difficulty were higher during training, but results indicate that the difference eventually fades during the retention period. Significant differences in aptitude across participants were found.

Conclusions Visual force feedback training improves on certain aspects of force moderation in a laparoscopic resection task. Results suggest that with enough training there is no significant decay of learning within the first month of the retention period. It is essential to account for differences in aptitude between individuals in this type of longitudinal research. This study shows how an inexpensive force measuring system can be used with an FLS Trainer System after some retrofitting. Surgical instructors can develop their own tasks and adjust force feedback levels accordingly.

Visual force feedback enhances laparoscopic training of sensitive tasks. Trainees learn to moderate their applied forces with the help of on-screen cues. But the expected outcome of this practice is that it will foster an ability that extends beyond training. In effect, upon immediate culmination of such training, a favorable reduction of applied forces can be observed in delicate tasks [1–3]. In the long term, however, this learned skill is subject to decay in as quickly as a month without sufficient reinforcement [4]. It is unclear how this decay progresses nor its response to substantial training. Also, it is unclear how visual force feedback can benefit the training of other complex tasks relevant to laparoscopy.

✉ Shihab Asfour
sasfour@miami.edu

1 Industrial Engineering Department, University of Miami, 1251 Memorial Drive, McArthur Engineering Building, Coral Gables, FL 33146, USA

2 St. Jude Children's Research Hospital Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Mail Stop 768, Memphis, TN 38105, USA

3 College of Engineering, University of Miami, 1251 Memorial Drive, 268 McArthur Engineering Building, Coral Gables, FL 33146, USA

The application of visual force feedback to laparoscopic training is due in large part to the work of Horeman and Delft University [1, 2]. The benefits already observed in open suturing [4–6] and knot tying [7] were shown to transfer over to laparoscopic tasks, such as needle driving [1] and tissue manipulation with graspers [2, 3]. Improvements due to force feedback are based on the successful moderation of applied forces toward an idealized goal, which is usually the average performance of expert surgeons. Although these standards are task-specific and arguably subjective, they have a didactic application in promoting awareness of tissue handling.

Trainees naturally respond to the cues provided by visual force feedback and moderate their applied forces during a task. Therefore, immediate improvements are expected with this kind of training, but the type of visual feedback and the subsequent retention of skill are factors that merit consideration. In a study by Smit and Spruit [3], a group with continuous force feedback on-screen applied significantly lower forces immediately after training compared to a group where feedback appeared only when forces exceeded a threshold. Yet as noted by the authors, continuous visual feedback overlaid on the camera feed is not preferable since it can be distracting. Furthermore, the improvements observed in the study appeared to decay in a follow-up test a week later. Similarly, in another study by Horeman et al. but with open suturing, after one training session of visual force feedback, there were immediate improvements that later seemed to decay substantially within a month [4].

The objective of this research is to explore both the attainment and retention of laparoscopic force moderation as a result of training with visual force feedback. To this end, the research goals include: (1) the estimation of its effect on applied force, completion time, and rated difficulty during training and (2) in a 4-week period thereafter, and (3) if the benefits found in previous studies with this kind of training will transfer over to a laparoscopic resection task. We expect lower applied forces throughout training, yet with a decay in performance throughout the retention period. We also expect greater task completion times and rated difficulty during both the training and retention periods.

Materials and methods

Participants

The procedure, methods, and consent forms used in this study were approved by the Internal Review Board of the University of Miami. The participants were twenty-two male university students (mean $\pm$ 1SD = 27 $\pm$ 3 years old) with no previous experience in laparoscopy. Gender was controlled in the study since we did not have adequate statistical power to account for its potential effect on the results. All participants underwent FLS proficiency training [8] as outlined in Table 1 for the Peg Transfer and Precision Cutting tasks before starting the experiment (mean $\pm$ 1SD = 4 $\pm$ 2 h for completion). However, three participants dropped out during this time, and thus the number of subjects who participated in this study was 19.

Apparatus

The experiment was conducted with an FLS Trainer System (VTI Medical, Waltham, MA, USA), retrofitted with a five megapixel USB camera (Ailipu Technology Co., Guangdong, China) to allow for computer connectivity. The live camera feed was then displayed on a 23″ monitor (Dell, Round Rock, TX, USA) mounted on a tripod. The height of both the operating table and tripod were adjusted to ensure individual user comfort.

The force measuring system relies on a cantilever 5 kg load cell (UXCell, Hong Kong, China) placed underneath the resection model. Applied forces cause the load cell to bend, altering the electrical resistance of its strain gages. This principle allows for the direct measurement of pulling and pushing forces, once the load cell is calibrated with precision weights. Since the load cell can only measure in one direction of bending, it was pre-tightened so that it could measure changes in both.

The load cell was connected to a computer via a printed circuit board mounted on an Arduino Uno. During data acquisition, the load cell signal was amplified by the circuit and then sent to the computer via USB by the Arduino at a rate of 32 Hz. The software used to calibrate the load cell and process the signal was *Processing 3* programming language [9] with *papaya* [10] and *controlP5* [11] libraries. The load cell measurable force range was set to 15 N for pulling and 10 N for pushing, with an accuracy of 0.05 N.

Resection task

The resection task was modeled as the latter part of a cholecystectomy due to the popularity of this procedure and since it requires both precision cutting and tissue manipulation. The gallbladder was made of soft mold material (Alumilite Corp., Kalamazoo, MI, USA) filled with cotton to a size similar to that of real life [12]. The user is required to remove the gallbladder from the base, which acts as the liver, by first clipping the ducts and the outer connective tissue with a 5 mm laparoscopic scissor (Ethicon, Somerville, NJ, USA) using their right hand, as shown in steps 1–3 in Fig. 1. Finally, the gallbladder is removed from its Velcro attachment to the base with a

Table 1 Required FLS skill proficiency for the two basic tasks [8]

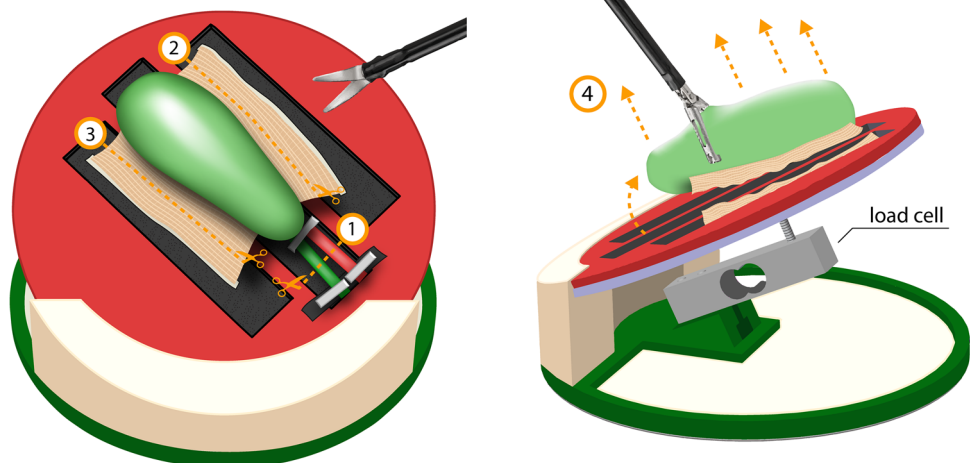| Task	Completion time (s)	Requirements	Repetition
Peg transfer	48	No drops outside field of view	Two consecutive + ten nonconsecutive
Precision cutting	98	All cuts between two circles of the training gauze	Two consecutive

5 mm babcock grasper (Ethicon, Somerville, NJ, USA) using their left hand (step 4). The user is instructed to focus on efficiency as well as gentle manipulation during resection.

Visual force feedback

A display was created using the *Processing 3* software. The pulling and pushing forces exerted on the base were shown in the form of a circle that moves from left to right, respectively, as shown in Fig. 2. The force feedback display is overlaid on the live camera feed on the top left of the screen, without obstructing the user's view of the task.

The endpoints of the display indicate the maximum desirable force and were based on the results of a pilot study on six expert laparoscopic surgeons (Table 2) who performed the resection task. Descriptive statistics of this data are included in Table 3. The maximum pulling and pushing force for the surgeons were 3.22 N (mean = 2.25 N, SD = 0.86) and 5.92 N (mean = 5.40 N, SD = 0.39 N), respectively, yet these values were used as a warning in the display, shown in yellow. The maximum values for the feedback system, shown in red, correspond to a standard deviation above the surgeon maximums (4.08 and 6.31 N). Once the user reaches any of these forces, there is a corresponding color flash over the circle to bring attention to the display.

Outcomes measured

- *Maximum pulling force* The maximum pulling force measured by the load cell underneath the base throughout the task. Measured in newtons.
- *Maximum pushing force* The maximum pushing force measured by the load cell underneath the base throughout the task. Measured in newtons.
- *Completion time* Time measured from first movement of graspers in view to the removal of the gallbladder from the base. Measured in seconds.
- *Task difficulty* Participant rating of difficulty once task is completed using the following scale: 1 = very easy, 2 = easy, 3 = neutral, 4 = hard, and 5 = very hard.

Protocol

In order to assess the benefits of visual force feedback training, participants were assigned to either a control or treatment group. Group allocation was done in alternating fashion based on the order in which participants completed the FLS proficiency training. For example, the first participant to finish proficiency training was assigned to the control group, so the next participant was assigned to the treatment group, and so on.

All participants watched a video of an actual laparoscopic cholecystectomy for them to gauge the amount of

Fig. 1 Resection task steps: cut (*1*) ducts and (*2–3*) connective tissue with laparoscopic scissor, and then (*4*) remove gallbladder from Velcro attachment with babcock grasper

Fig. 2 Visual force feedback system design: ball slides left/right when pulling/pushing, and warnings appear when specific force values are reached
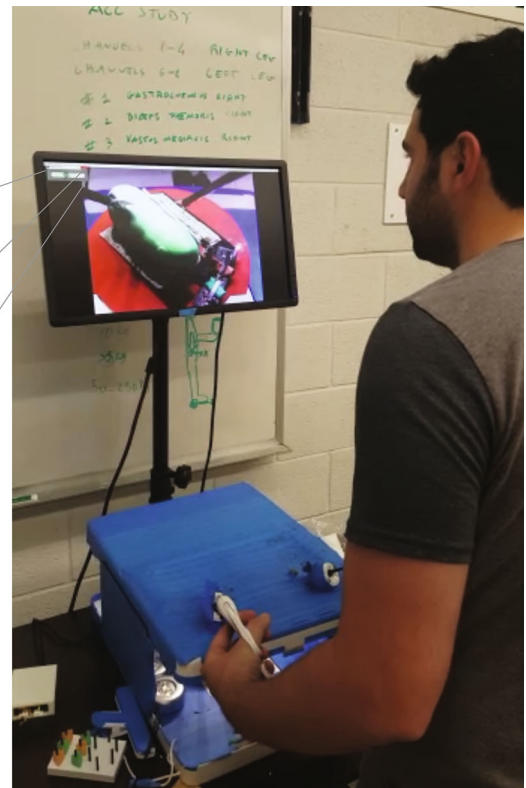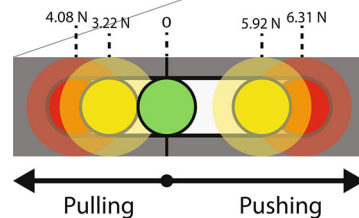

Table 2 Surgeon information

Surgeon	Years of laparoscopic experience
1	>20
2	>10
3	>20
4	9
5	>20
6	17

Table 3 Descriptive statistics for surgeon data

Statistic	Maximum pulling force (N)	Maximum pushing force (N)	Completion time (s)
N	6	5	6
Max	3.22	5.92	240
Mean	2.24	5.40	172
Std dev	0.86	0.39	40
Median	2.22	5.32	169
Min	0.85	5.01	128

forces applied during the procedure. They then performed the resection task to provide an initial baseline of performance. Both groups performed the resection task sixteen more times. However, the treatment group had visual force feedback throughout, whereas the control group did not. Both groups then performed a post-test without feedback immediately after training. All of the eighteen trials were spread within two weeks to avoid fatigue effects, grouped in six sessions of three trials. Participants then performed four weekly test trials to assess long-term retention of training (see Fig. 3).

Data analysis

The data were analyzed descriptively via plots and summary statistics. Pairwise t tests were run using the R programming language [13] to estimate changes in average performance within Groups across time. A p value <0.05 was considered significant.

A longitudinal analysis was performed on the data. Linear mixed-effect regression models were run using R and the *nlme* library [14] on maximum pulling force, maximum pushing force, and completion time outcomes. Ordinal mixed-effect regression was run on Task Difficulty using R and the *ordinal* library [15]. The fixed effects were Trial/Week, Group (feedback vs. no feedback), and the interaction effect between Group and Trial/Week. Random slope and intercept for the Trial/Week variable was incorporated into the model to allow for better estimation of participant-specific profiles. The data for the training and retention periods were modeled separately.

Fig. 3 Overview of experimental design (shaded region represents use of visual force feedback)
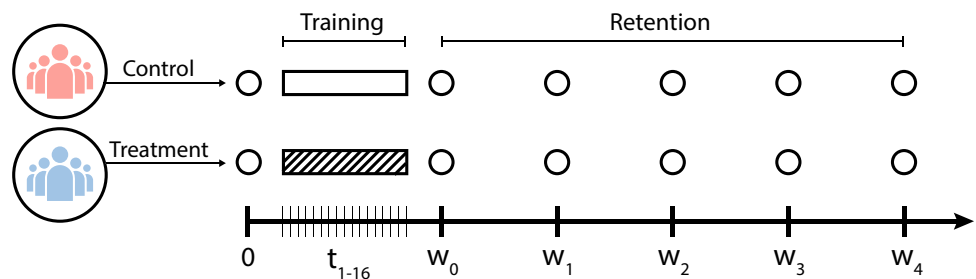
Group levels were dummy coded with "0" for the control group and "1" for the treatment group. An additional within-Group variance estimation was included in the model (when significant) to calculate relative proportion of variance between the two groups. 2nd order polynomial terms for the trial and week were used when appropriate to control for nonlinearity in the data. Data transformations were applied to the outcomes in order to help reduce skewness and normalize the residuals in the models.

Results

Baseline performance values along with the time to achieve FLS proficiency are provided in the scatterplots of Fig. 4. It is apparent that both groups had similar spread of performance values and that there was no correlation between baseline performance and time to achieve FLS proficiency.

Descriptive statistics for the measured outcomes at key time points are shown in Table 4. The scatterplots given in Figs. 5, 6, and 7 provide visualization of the raw data as well as the estimated fixed effect model for the training and retention periods. Reference lines for both the feedback maximum value and surgeon means from the pilot study are overlaid on the plots. The pairwise t test results are shown in Table 5.

Overall, the inclusion of Participant as a random effect in all of the models accounted for the majority of the variability across time, suggesting large variability across the forces exerted among participants. Significant results for each outcome are summarized in the following subsections.

Maximum pulling force

As shown in the pairwise t test results in Table 3, a significant decrease of average maximum pulling force of
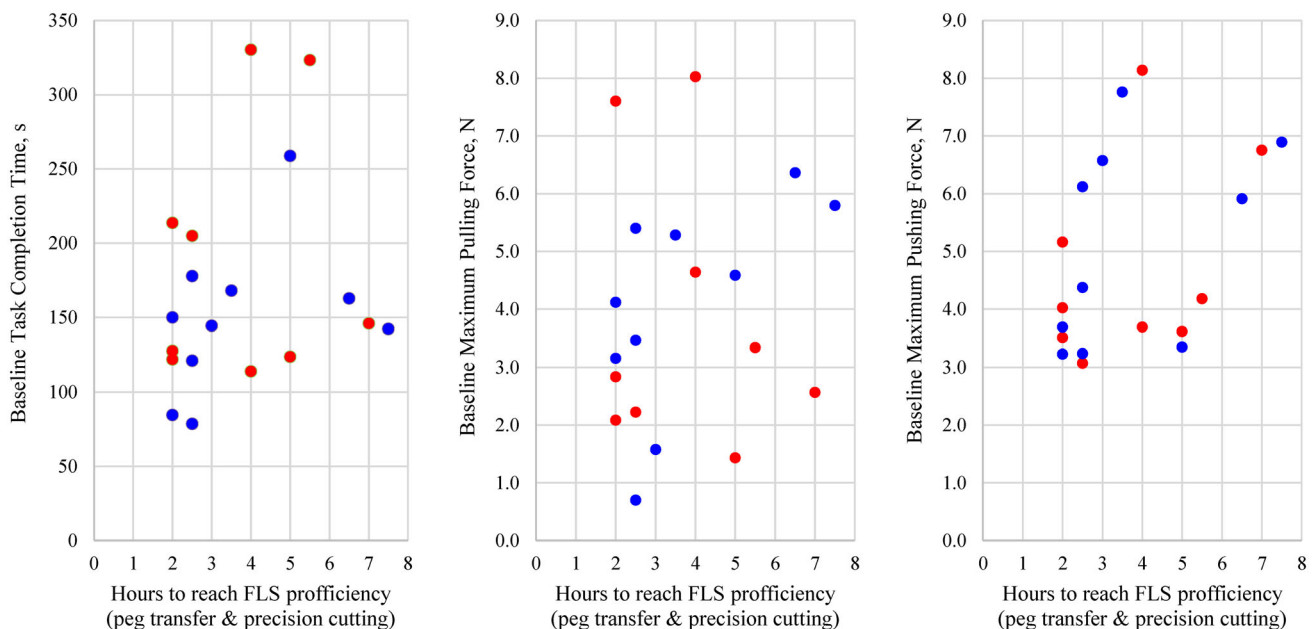

Fig. 4 Scatter point graphs of baseline performance values vs time to reach FLS proficiency for peg transfer and precision cutting (*red* control, *blue* treatment) (Color figure online)

Table 4 Descriptive statistics (0 = control group, 1 = treatment group)

| Statistic	Group	Maximum pulling force (N)				Maximum pushing force (N)				Completion time (s)			
		T_0	T_{16}	W_0	W_4	T_0	T_{16}	W_0	W_4	T_0	T_{16}	W_0	W_4
N	0	9	9	9	9	9	9	9	9	9	9	9	9
	1	10	10	10	9	10	10	10	9	10	10	10	9
Max	0	8.03	4.85	5.60	9.93	8.14	8.28	8.06	9.34	330	78	98	100
	1	6.36	3.63	4.09	4.18	7.76	5.52	6.99	7.15	259	142	126	111
Quantile75	0	6.13	3.52	4.18	3.90	5.96	6.21	6.30	7.50	269	74	74	86
	1	5.50	2.93	3.74	3.46	6.65	4.32	5.66	4.85	171	105	112	105
Mean	0	3.86	2.88	3.18	3.27	4.68	5.54	5.10	5.86	190	64	68	75
	1	4.05	2.64	2.98	2.75	5.11	4.12	4.12	4.31	149	92	95	89
Std dev	0	2.42	1.05	1.35	2.71	1.71	1.29	1.64	2.03	86	9	14	16
	1	1.84	0.56	0.85	0.90	1.72	0.66	1.68	1.32	51	26	20	22
Median	0	2.84	2.99	2.96	2.56	4.02	5.30	5.32	5.74	146	60	68	79
	1	4.36	2.68	3.01	2.49	5.14	3.93	3.95	4.21	147	87	96	97
Quantile25	0	2.16	1.89	2.09	1.57	3.56	4.64	3.65	4.08	123	56	57	62
	1	2.77	2.39	2.48	2.06	3.32	3.71	2.79	3.40	112	75	78	65
Min	0	1.43	1.53	1.41	0.87	3.07	4.08	2.93	2.97	114	53	51	51
	1	0.70	1.57	1.22	1.31	3.23	3.53	1.72	2.43	79	54	63	52
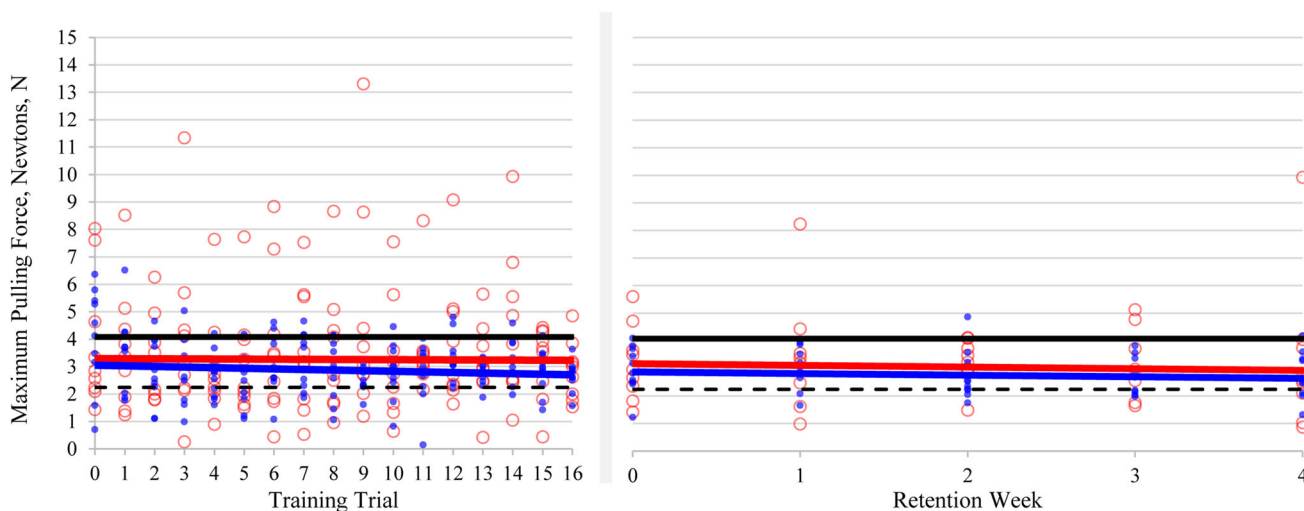
Fig. 5 Maximum pulling force scatterplot and fitted mixed-model curves (*red* control group, *blue* treatment group, *black* feedback max value, *dashed* surgeon mean) (Color figure online)

−1.40 N ($t(9) = -2.60$, $p = 0.0289$) was found for the treatment group between the first and last training trial.

None of the terms in the training and retention models (Eqs. 1, 2) were significant. Consequently, the fitted model curves on Fig. 5 for both the training and retention period are almost identical, suggesting that there was no difference in the maximum pulling force between the two groups during the retention or training periods. Furthermore, there was no suggestion that either group's performance improved over time. However, the variance of the treatment group was lower throughout the experiment. The treatment group had 76% (CI 65–90%) of the variance of the control group during the training period and 57% (CI 40–81%) during the retention period.

$$\text{Max pulling force (Training)} = -0.004 \times \text{Trial} - 0.24 \\ \times \text{Group} - 0.018 \\ \times \text{Group} \times \text{Trial} + 3.29 \tag{1}$$

$$\text{Max pulling force (Retention)} \\ = -0.061 \times \text{Week} - 0.30 \times \text{Group} + 0.003 \times \text{Group} \\ \times \text{Week} + 3.17 \tag{2}$$
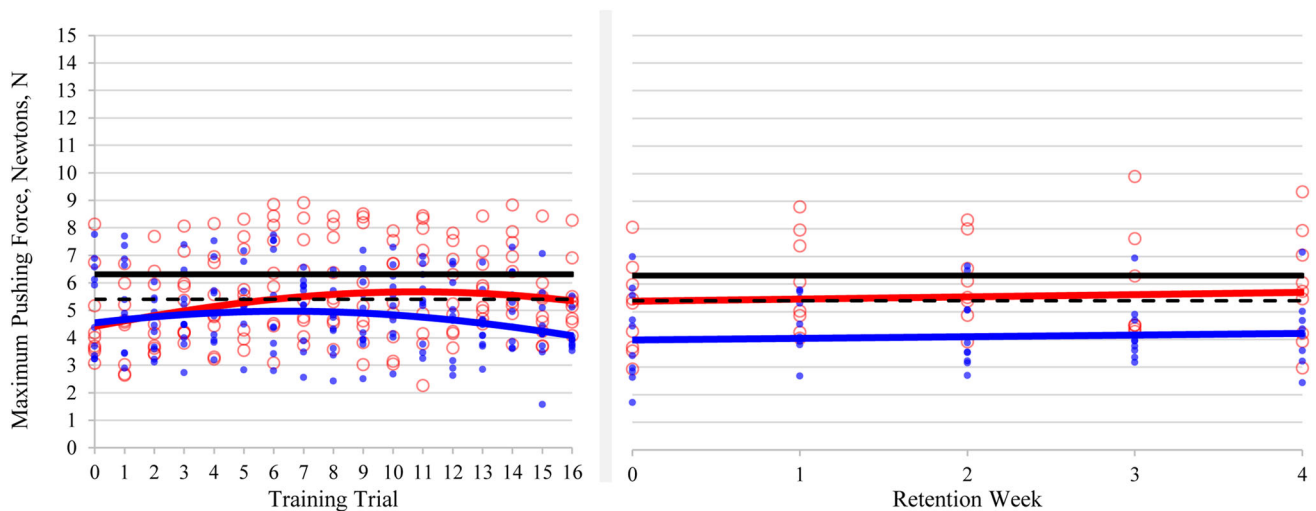
Fig. 6 Maximum pushing force scatterplot and fitted mixed-model curves (*red* control group, *blue* treatment group, *black* feedback max value, *dashed* surgeon mean) (Color figure online)

Both of the fitted model curves are below the feedback maximum and above the mean surgeon values throughout the experiment (see Fig. 5).

Maximum pushing force

As shown in the pairwise *t* test results in Table 3, there is a significant increase of average maximum pushing force of +0.86 N ($t(8) = 3.23$, $p = 0.0121$) for the control group between the first and last training trial.

For the longitudinal model, since the data suggested nonlinearity, a second order model was fit to the training period data over time with a 1st order interaction of Group by Trial (Eq. 3). The model suggested lower average maximum pushing force for the feedback group over time

(-0.11 N/Trial, $t(300) = -2.15$, $p = 0.0320$), given by the significance of the interaction term. The variance of the treatment group was larger 22% (CI 4–44%) compared to the control group.

$$\begin{aligned}\text{Max pushing force (Training)}\\ = -0.011 \times \text{Trial}^2 + 0.24 \times \text{Trial} + 0.16 \times \text{Group}\\ - 0.11 \times \text{Group} \times \text{Trial} + 4.39.\end{aligned} \tag{3}$$

During the retention period (Eq. 4), the overall mean Maximum Pushing Force of the treatment group was 1.40 N lower than the control group, provided by the significance of the Group term in the model ($t(17) = -3.01$, $p = 0.0078$). The variances of the two groups were not significantly different. A model with an interaction term was also considered but it was not significant.
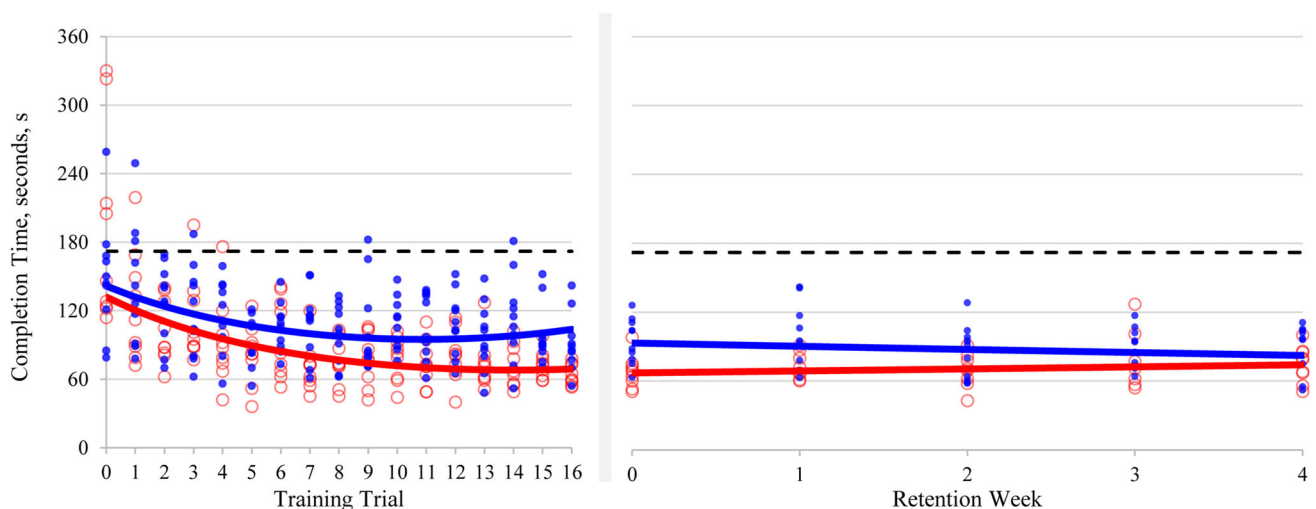

Fig. 7 Completion time scatterplot and fitted mixed-model curves (*red* control group, *blue* treatment group, *dashed* surgeon mean) (Color figure online)

Max pushing force (Retention)
$$= 0.080 \times \text{Week} - 1.40 \times \text{Group} - 0.02 \times \text{Group}$$
$$\times \text{Week} + 5.38. \tag{4}$$

The fitted model curves are below the feedback maximum line (Fig. 6). However, the control group curve appears to lie along the surgeon mean line throughout most of the experiment, whereas the treatment group curve is lower throughout.

Completion time

The mean values for both groups decreased significantly over time suggesting a learning effect, with the control group having the greatest change (see Table 5). By the end of the experiment, the control group mean decreased by -115 s ($t(8) = -4.33$, $p = 0.0025$), whereas the treatment group's decreased by -60 s ($t(8) = -3.34$, $p = 0.0103$).

The completion times during the training period displayed a nonlinear pattern and were modeled via a 2nd order polynomial (Eq. 5). The interaction term was significant, indicating that the average completion time for the treatment group was higher relative to the control group over time (1.09 s/Trial, $t(300) = 2.11$, $p = 0.0356$). The variances of the two groups were not significantly different.

$$\text{Completion time (Training)} = 0.366 \times \text{Trial}^2 - 9.51$$
$$\times \text{Trial} + 11 \times \text{Group}$$
$$+ 1.09 \times \text{Group} \times \text{Trial}$$
$$+ 129. \tag{5}$$

During the retention period (Eq. 6), the initial completion time for the feedback group was higher compared to the control group ($t(17) = 3.31$ $p = 0.0042$). However, there is also a significant and negative interaction term ($t(67) = -2.06$, $p = 0.0433$), indicating that the treatment group lowered its completion time on average by 4.5 s per week. No such change was observed in the control group. The variances of the two groups were not significantly different.

$$\text{Completion time (Retention)} = 1.774 \times \text{Week} + 26$$
$$\times \text{Group} - 4.48 \times \text{Group}$$
$$\times \text{Week} + 67. \tag{6}$$

Both fitted model curves lie well below the surgeon mean line throughout the experiment, as shown in Fig. 7. Towards the end of the retention period, both model curves appear to converge.

Task difficulty

The scatterplot and mean lines in Fig. 8 provide a basic representation of the raw data across time without taking into account Participants, which is done with the ordinal mixed effect models.

The interaction term was not significant during the training period and was therefore removed from the model. There was, however, a significant learning curve effect during the training period ($z = -6.99$, pr($>|z|$) < 0.001). Task difficulty was assessed as being greater on average for the treatment group during the training period, given by the significance of the Group term in the model ($z = 3.13$, pr($>|z|$) $= 0.0018$).

The interaction term was also not significant during the retention period and was removed from the model. During the retention period, there were no significant differences in task difficulty ratings between the two groups.

Discussion

Our results confirm that visual force feedback training can improve on certain aspects of force moderation in a laparoscopic resection task. Extreme pulling force values were tapered throughout both the training and retention period. Average maximum pushing forces were significantly lowered towards the end of training and throughout the retention period. No significant decay or change of applied force learning was found during the 4-week retention period. Completion time and rated task difficulty were higher during the training period, but the results suggest that the difference eventually becomes negligible during the retention period.

The significant improvements in force moderation found in the study are supported by the literature, but on different laparoscopic tasks [1–4] as well as those done without laparoscopy [5, 6]. In particular, the average maximum force is known to decrease significantly after training with visual force feedback, as was the case for the applied pushing force. However, we also introduced the measure of relative variance between the two groups as another indicator of force moderation. In effect, we found that the treatment group had significantly lower variance in their maximum applied pulling force, even though the average value between the two groups was essentially the same throughout the experiment. This means that the force feedback in the treatment group resulted in lowering extreme pulling force values compared to the control group.

One of the interesting results of this study is that significant improvements in force moderation were still found despite preconditioning of the participants before the experiment and having them endure prolonged training. For one, the substantial FLS training gave participants previous experience in laparoscopy, which is typically avoided in this kind of study to enhance contrast of

Table 5 Pairwise tests within-group at key time points (0 = control group, 1 = treatment group)

Outcome	Group	Baseline → end of training ($T_0 \rightarrow T_{16}$)						Baseline → immediate retention ($T_0 \rightarrow W_0$)						Baseline → retention week 4 ($T_0 \rightarrow W_4$)						End of training → retention week 4 ($T_{16} \rightarrow W_4$)					
		ΔM	95% CI lwr	uppr	t	df	p	ΔM	95% CI lwr	uppr	t	df	p	ΔM	95% CI lwr	uppr	t	df	p	ΔM	95% CI lwr	uppr	t	df	p
Max pulling force (N)	0	−0.99	−2.22	0.25	−1.84	8	0.1027	−0.68	−1.84	0.48	−1.35	8	0.2136	−0.59	−2.73	1.54	−0.64	8	0.5399	−0.59	−2.73	1.54	−0.64	8	0.5399
	1	−1.40*	−2.62	−0.18	−2.60	9	0.0289	−1.07	−2.47	0.34	−1.72	9	0.1201	−1.28	−2.82	0.25	−1.94	8	0.0890	−1.28	−2.82	0.25	−1.94	8	0.0890
Max pushing force (N)	0	0.86*	0.24	1.47	3.23	8	0.0121	0.42	−1.07	1.92	0.65	8	0.5329	1.18	−0.14	2.49	2.06	8	0.0736	1.18	−0.14	2.49	2.06	8	0.0736
	1	−1.00	−2.35	0.35	−1.67	9	0.1290	−1.00	−2.23	0.24	−1.82	9	0.1025	−0.96	−2.70	0.78	−1.27	8	0.2404	−0.96	−2.70	0.78	−1.27	8	0.2404
Completion time (s)	0	−126*	−192	−60	−4.38	8	0.0235	−122**	−181	−62	−4.71	8	0.0015	−115**	−176	−54	−4.33	8	0.0025	−115**	−176	−54	−4.33	8	0.0025
	1	−57**	−94	−21	−3.57	9	0.0061	−54**	−89	−19	−3.52	9	0.0065	−60*	−102	−19	−3.34	8	0.0103	−60*	−102	−19	−3.34	8	0.0103

* $p < 0.05$, ** $p < 0.01$

performance. There is therefore an indication that improvements in force moderation can still be observed even with those who have previous laparoscopic experience. Furthermore, the training period was longer than what is found in similar literature studies as well, potentially providing more opportunity for the control group to catch up with the treatment group in skill by the end of the training period.

It is unclear why there was an absence of a learning decay in this study for either the control or feedback group. Perhaps, it is related to the amount of training that both groups received. In effect, Horeman et al. attributed insufficient training to the decay of performance they observed in their study on laparoscopic knot-tying within a month's time [4]. Alternatively, the same study supports the longer task completion time we found for the treatment group. This is reasonable considering the increase of mental workload from using a display during training and the concern about reaching the warning zones. Task difficulty was also significantly higher for the treatment group. Nonetheless, there is indication in our study that the difference in completion time and difficulty between the two groups eventually became negligible during the retention period, perhaps due to an increase in task confidence for the treatment group over time.

There were some other unexpected results in our study as well. For one, we did not find a significant improvement in the average maximum pulling force. Although the extreme values were successfully tapered, we expected significantly lower mean values as suggested by the literature. This might be an indication that the feedback maximum value was set too high or simply that a longer training period is required to observe significant changes for the resection task. On the other hand, we found that the completion time for both groups was significantly lower than the surgeon mean value. Therefore, there was unnecessary rushing to complete the task which might be the result of having undergone the FLS proficiency training right before the experiment, which is time sensitive. We also discovered that including participants as a random factor in the regression models accounted for the majority of the variance, which indicates a significant difference in aptitudes or approach among the participants.

A potential source of discrepancy in the results would be if both groups were not equally balanced in natural tissue handling aptitude. Aside from the assumption that FLS proficiency would equalize skill across participants, the alternating method of randomization depended on the potential correlation between baseline performance and time to reach FLS proficiency to ensure balanced groups. Although we found that both groups had similar spread of baseline performance values, we also found that these values were not correlated with time to reach FLS
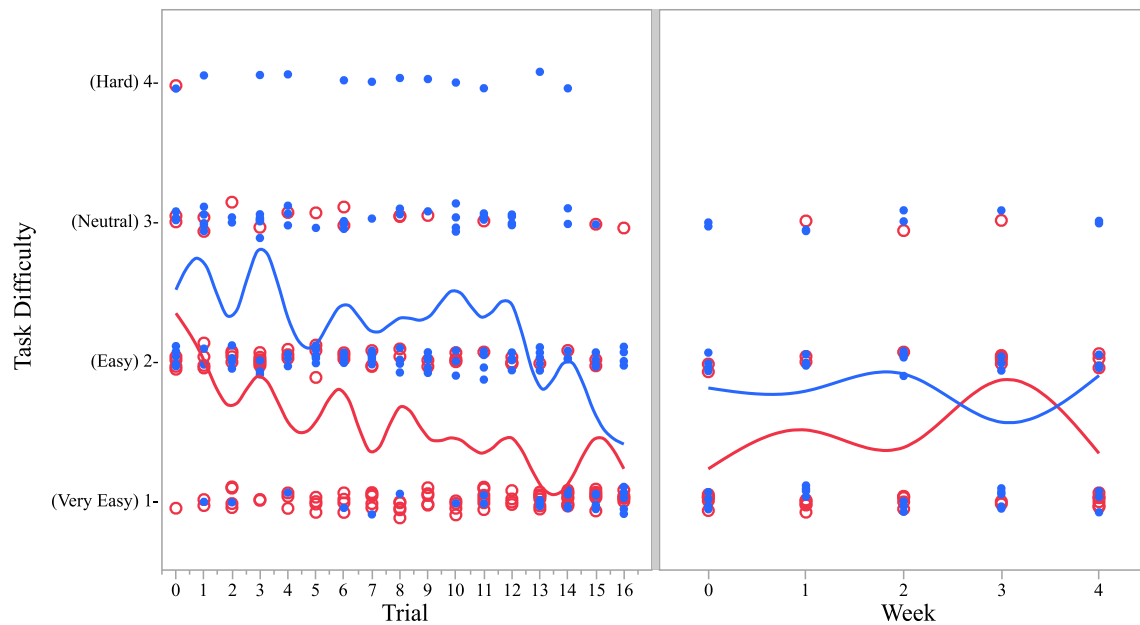
Fig. 8 Scatterplot and connected mean lines for rated task difficulty (*red* control group, *blue* treatment group) (Color figure online)

proficiency. This was an interesting observation, since it seemed reasonable to assume that natural aptitude to reach FLS proficiency would have some correlation to tissue handling performance. Therefore, a pretest based on the intended measurable outcomes might be the most effective method to create balanced groups. However, a participant allocation method that relies on the results of a pretest before starting the experiment might not be practical in an on-going longitudinal study.

Longitudinal studies such as this one present difficulties in terms of participant recruitment, since they require long-term commitment and convenient access to the experiment's location. To this effect, it was not feasible to use medical students. Therefore, we attempted to mitigate the potential differences in performance in drawing from another student population. Namely, our participants went through rigorous FLS proficiency testing involving peg transfer and precision cutting in order to reduce the variability in their level of performance at the start of the experiment. Furthermore, in order to address a potential difference in applied forcefulness, each participant watched a video of an actual laparoscopic cholecystectomy with the intent to illustrate the forces applied during such an operation. Participants were also asked to treat the experimental task as if it were an actual procedure. Given that the control group's mean for both the maximum pulling and, especially, the maximum pushing were very similar to that of the experienced surgeons' mean, there is indication that the potential disadvantage of using non-medical students was mitigated.

Although it is not possible yet to generalize the results of the study to the general population, there is still an

indication of the positive effect of learning with visual force feedback. Nonetheless, factors such as the limited sample size and having controlled for gender do not allow us to observe the full extent of variability in performance resulting from this kind of training. As evident in our study, people react differently to the use of visual force feedback. This has direct implications in the design of effective training curricula, such as in the amount of training necessary to ensure long-term retention of learning.

For future research, we recommend improving the system's capabilities by including the measurement of forces in the transverse direction, as opposed to only those in the axial direction (i.e., pulling and pushing). This was a limitation of the study since trainees also pulled the gallbladder in the transverse direction during resection, which was outside of the sensing capacity of the load cell. We also recommend the inclusion of another metric, other than rated task difficulty, to objectively measure the amount of stress involved in training with visual force feedback. There are examples of this in the literature for laparoscopy, such as measuring changes in heart rate [16, 17] and blood pressure [17]. Furthermore, we suggest extending the retention period even further, since in our study we could not observe enough of a trend to adequately predict how force application might change beyond 4 weeks. It is reasonable to expect some decay, but we did not capture this in our study.

It is important to clarify that this study does not seek to eventually shorten the amount of training needed in an actual surgical setting, nor to influence this part of the curriculum with the use of visual force feedback. Instead, we suggest that preliminary training with visual force

feedback can improve laparoscopic tissue manipulation before trainees operate on live models, in the same way that FLS proficiency does to provide basic skills. Although surgical trainees can learn successful force moderation through eventual practice with lab and clinical operative experience, it is in the interest of a surgical program to best prepare residents before operating on live models. While virtual reality trainers can offer visual feedback based on a prediction of forcefulness, there is limited to no haptic feedback on the graspers to reinforce motor learning, as well as incurring greater costs than the use of an FLS trainer system.

Discovering that there is persistent acquired skill in tissue handling after training with visual force feedback provides the groundwork to then explore its potential influence on a resident's performance in an actual surgical setting. Measuring forces would require a different apparatus that can be introduced inside the body, and other metrics such as the amount of mistakes, stress level, and evaluations by surgical instructors can be used to assess if there are benefits from previous training with visual force feedback.

The results of this study show the benefits of visual force feedback training when applied to a laparoscopic resection task. Trainees are gentler when manipulating tissue, reflected in either lower extreme force values and/or overall average values. There is also evidence that with enough training there is no significant decay of learning within the first month of the retention period. However, accounting for the difference in aptitude between individuals is essential when assessing the benefit of learning with feedback training. Overall, we demonstrate that it is possible to incorporate force feedback capabilities into an FLS trainer, providing a seamless transition from previous FLS proficiency training. Surgical instructors can create tasks that are pertinent to their specialty and then establish objective performance standards of tissue handling which students can later train to replicate.

Compliance with ethical standards

References

1. Horeman T, Rodrigues SP, van den Dobbelsteen JJ, Jansen FW, Dankelman J (2012) Visual force feedback in laparoscopic training. Surg Endosc. doi:10.1007/s00464-011-1861-4
2. Horeman T, van Delft F, Blikkendaal MD, Dankelman J, van den Dobbelsteen JJ, Jansen FW (2014) Learning from visual force feedback in box trainers: tissue manipulation in laparoscopic surgery. Surg Endosc. doi:10.1007/s00464-014-3425-x
3. Smit D, Spruit E, Dankelman J, Tuijthof G, Hamming J, Horeman T (2016) Improving training of laparoscopic tissue manipulation skills using various visual force feedback types. Surg Endosc. doi:10.1007/s00464-016-4972-0
4. Horeman T, Blikkendaal MD, Feng D, van Dijke A, Jansen F, Dankelman J, van den Dobbelsteen JJ (2014) Visual force feedback improves knot-tying security. J Surg Educ. doi:10.1016/j.jsurg.2013.06.021
5. Rafiq A, Tamariz F, Boanca C, Lavrentyev V, Merrell RC (2008) Objective assessment of training surgical skills using simulated tissue interface with real-time feedback. J Surg Educ. doi:10.1016/j.jsurg.2008.05.012
6. Rodrigues SP, Horeman T, Sam P, Dankelman J, van den Dobbelsteen JJ, Jansen FW (2014) Influence of visual force feedback on tissue handling in minimally invasive surgery. Br J Surg. doi:10.1002/bjs.9669
7. Hsu JL, Korndorffer JR Jr, Brown KM (2016) Force feedback vessel ligation simulator in knot-tying proficiency training. Am J Surg. doi:10.1016/j.amjsurg.2015.09.009
8. Ritter S (2006) FLS Technical skills proficiency-based training curriculum. http://www.flsprogram.org/. Accessed 1 Aug 2016
9. Ben Fry and Casey Reas (2016) Processing 3. https://www.processing.org/. Accesssed 3 Nov 2016
10. Adila (2015) Papaya: statistics library for processing. http://www.adilapapaya.com/papayastatistics.html. Accessed 23 Apr 2015
11. Schlegel A (2015) controlP5: a GUI (graphical user interface) library for processing. http://www.sojamo.de/libraries/controlP5/. Accessed 19 Apr 2015
12. Clavien PA, Baillie J (eds) (2006) Diseases of the gallbladder and bile ducts: diagnosis and treatment, 2nd edn. Blackwell Pub, Malden
13. R Core Team (2016) R: a language and environment for statistical computing. https://www.R-project.org/. Accessed 10 Nov 2016
14. Pinheiro JBD, DebRoy S, Sarkar D, R Core Team (2016) nlme: linear and nonlinear mixed effects models. http://CRAN.R-project.org/package=nlme. Accessed 12 Nov 2016
15. Christensen RHB (2015) ordinal—regression models for ordinal data. http://www.cran.r-project.org/package=ordinal/. Accessed 3 Dec 2016
16. Arora S, Sevdalis N, Aggarwal R, Sirimanna P, Darzi A, Kneebone R (2010) Stress impairs psychomotor performance in novice laparoscopic surgeons. Surg Endosc. doi:10.1007/s00464-010-1013-2
17. Andreatta PB, Hillard M, Krain LP (2010) The impact of stress factors in simulation-based laparoscopic training. Surgery. doi:10.1016/j.surg.2009.10.071