Ritter, F. E., Yeh, M. K.-C., Yan, Y., Siu, K.-C., Oleynikov, D. (2020). *Effects of varied surgical simulation training schedules on motor-skill acquisition, Surgical Innovation. 27*(1). 68-80.
https://doi.org/10.1177/1553350619881591

Effects of Varied Surgical Simulation Training Schedules on Motor-skill Acquisition

Frank E. Ritter, Ph.D. C. Psychol. (frank.ritter@psu.edu)
College of Information Science and Technology, The Pennsylvania State University
University Park, PA 16802, USA, (814) 865-4453

Martin K.-C. Yeh, Ph.D. (martin.yeh@psu.edu)
College of Information Science and Technology, The Pennsylvania State University, Brandywine
Media, PA 19063, USA, (610) 892-1351

Yu Yan, Ph.D. (yzy122@psu.edu)
Learning, Design, and Technology, The Pennsylvania State University
University Park, PA 16802, USA, (814) 777-2416

Ka-Chun Siu, Ph.D. (kcsiu@unmc.edu)
College of Allied Health Professions, University of Nebraska Medical Center
Omaha, NE 68198, USA, (402) 559-8464

Dmitry Oleynikov, M.D. (doleynik@unmc.edu)
Department of Surgery, University of Nebraska Medical Center
Omaha, NE 68198, USA, (402) 559-4581

This research was supported by ONR grants #W911QY-07-01-0004, #N00014-10-1-0401, and N00014-15-1-2275.

Keywords: surgical education, ergonomics and/or human factors study, simulation, Laparoscopic surgery, training schedules, practice effect

Date: 9/15/2019

effects_of_varied_training-SRI-13-0155-v36.doc

Abstract

There have been many studies to evaluate the effect of training schedules on retention; however, these usually compare only two drastically different schedules, massed and distributed, and they have tended to look at declarative knowledge tasks. This study examined learning on a laparoscopic surgery simulator using a set of procedural or perceptual-motor tasks with some declarative elements. The study used distributed, massed, and two hybrid-training schedules that are neither distributed nor massed. To evaluate the training schedules, twenty-three participants with no previous laparoscopic experience were recruited and randomly assigned to one of the four training schedules. They performed three laparoscopic training tasks in eight 30-minute learning sessions. We compared how task time decreased with each schedule in a between-participants design. We found participants in all groups demonstrated a decrease in task completion time as the number of training sessions increased; however, there were no statistically significant differences in participants' improvement on task completion time between the four different training schedule groups, which suggested that time on task is more important for learning these tasks than the training schedule.

1 Introduction

Practice and skill acquisition mechanisms have been classical areas of study in cognitive psychology. The ubiquitous law of "practice makes perfect" has often been investigated[1]. One aspect of learning that has been examined is training schedules. It has been shown that different training schedules can affect the rate of skill acquisition [2,3]. A useful application for studying learning is laparoscopic surgery, because training is necessary for medical students starting laparoscopic surgery[4]; it would be useful to know the best training schedule for them and whether different types of surgical skills interact with training schedules[5,6].

There can be several kinds of training schedules. Two types of training have typically been examined. The first is massed, that is, the training happens all at once. The second is distributed, the training is spread out as much as possible. There are also combinations of the two, hybrids, such as schedules that had several days of spread out practice and a mass of training on one of the days.

Previous studies have shown conflicting results about which training schedule is better. Verdaasdonk et al.[7] examined massed and distributed training schedules for laparoscopic surgery skills and found that while distributed training is more effective than massed training in the course of one day, it is unclear whether distributed training schedules are more effective than massed training schedules over several days. Traditionally, for example, Moulton et al.[8] argued that a more distributed (e.g., one hour a day, 4 weeks in a row) schedule leads to better performance than a massed schedule (e.g., 4 hours in a row on one day). In their randomized controlled study, the participants' tasks were to remember the related knowledge that were demonstrated in two videos and to perform three different, more complex surgery tasks. However, Paik[3,9] has shown that a more massed training schedule led to improved performance in some tasks, and the distributed group was unable to learn a perceptual-motor task. Kim et al.'s[10] analysis suggest a schedule with some massed practice may be very useful for learning—the massed practice allows declarative learning about procedures to be strong enough to be proceduralized, effectively moving knowledge from declarative to procedural memory. There also exist far more than simply massed and distributed schedules, and few hybrids[2,3,9] (i.e., schedules consisting of a mix of distributed and massed practice) have been examined.

In addition to examining only extremes in training schedules for skill acquisition, many studies have focused on declarative knowledge learning, but not procedural learning. For example, Huckin et al.[11] reported several studies focused on declarative knowledge acquisition in the form of vocabulary learning. Studies on learning have also typically examined declarative knowledge acquisition[12,13].

Kim et al.[10] presented a theory that proposes a clear progression between declarative and procedural task knowledge (Figure 1). Their theory predicts that task knowledge progresses from a declarative stage, to a mixed stage. In the mixed stage periods of high practice after declarative learning has taken place can lead to improved procedural learning because the declarative memory is active enough to be proceduralized, and there is practice to proceduralize it. The final stage is fully proceduralized knowledge. On the other hand, if task practice is too distributed, the declarative memories may decay, and be less able to be proceduralized.

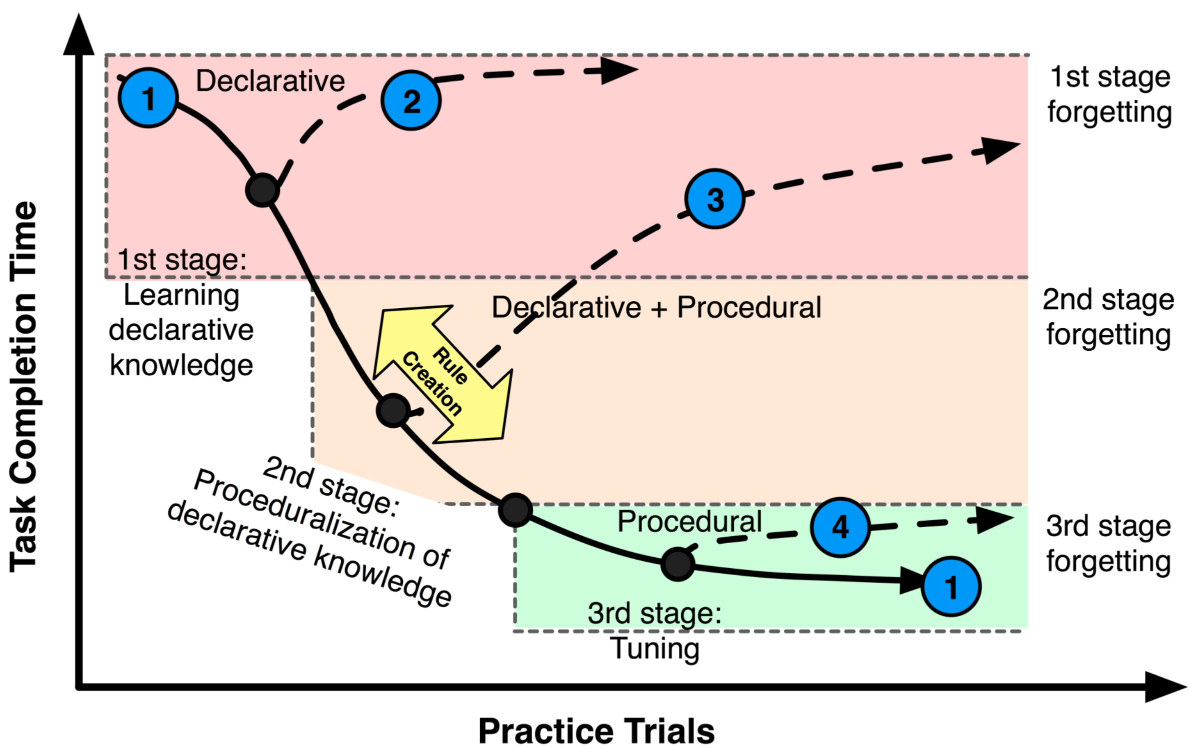


Figure 1. The theory of learning and retention by Kim et al.[10].

[High quality PDF is available.]

The ACT-R cognitive architecture[1] has been used to model many cognitive tasks including skill acquisition. It proposes a fixed set of mechanisms that give rise to cognition, including perception, action, procedural memory, declarative memory, and formulas to compute how they are related, how they get faster with practice, and how they decay. Declarative and procedural knowledge learning are driven by different mechanisms[14-16]. Within the ACT-R architecture, procedural knowledge is represented as "productions" that can be executed, and are composed of information that represents declarative knowledge. Performing a task can lead to moving task knowledge from a declarative representation to a procedural representation.

Paik[3,9] examined a wide range of training schedules, ranging from purely massed to purely distributed. He used the ACT-R declarative learning equations to predict retention and chose two non-massed-non-distributed learning schedules that he called hybrid-massed and hybrid-distributed to test. The hybrid schedules were particularly promising, and later shown to help learn a perceptual-motor skill better than the massed or distributed schedules[3].

Based on Kim et al.'s[10] theory and Paik's[3,9] results, we hypothesized that the participants in massed or hybrid-massed groups would have a greater decrease in task completion time than groups given a distributed or hybrid-distributed training schedule. Unlike previous studies that tend to examine declarative knowledge learning, we study a set of laparoscopic tasks that include procedural and perceptual-motor learning, as these additional knowledge types may particularly be helped by some massed practice to get procedural knowledge created.

2 Method

The purpose of this study was to examine the effects of varied training schedules on tasks used to train laparoscopic surgery. To this end, we compared four different training schedules (distributed, massed, hybrid-distributed, and hybrid-massed) where participants completed three tasks using a laparoscopic surgery simulator.

2.1 Participants

Participants were recruited at the Pennsylvania State University's University Park (main) campus. Twenty six participated in the study and 23 were analyzed (three could not complete the study, which is explained below). Among those 23 participants, there were 14 undergrad

---

[1] ACT-R is basically a name, but has been expanded into Atomic Components of Thought-Rationale[13].

students, eight graduate students, and one staff member. The students were in Informatics (11), Education (2), Health (2), Business (2), Communication (2), Liberal Arts (2) Engineering (1), Science (1), Health and Human Development (1), and Nursing (1). They were 9 (40%) female and 14 (60%) male, with a mean age of 23 (ranging from 18 to 34) years. The majority 19 (83%) were right-handed, and four (17%) were left-handed. Less than half (44%) of the participants were video game players based on self-report. None reported working with any tools similar to the one used in this study and none had training in laparoscopic surgery, including the two students from the Colleges of Nursing and of Health and Human Development). They were paid $7 per 30-minute session.

The participants were randomly assigned to four different training schedule groups using an assignment table. In the assignment table, the participant ID was ordered from 1 to 26, and the schedule ID (1,2,3, and 4) was assigned randomly in blocks of four. Participants were assigned to participant ID/Schedule ID in the order of recruitment. The first part of the assignment table is shown in Table 1.

Table 1. The assignment table for the random assignment

Participant ID	Training Schedule ID
1	4
2	1
3	3
4	2
5	1
6	4
7	2
8	3
9	2
10	4
11	1
12	3
…	…

2.2 Materials and Apparatus

We expanded upon a laparoscopic surgery simulator design of Alfa-Wali and Antoniou[17] (that is also similar to the MISTELS system[18]). Instead of using a cell phone as a light and recording source, we used a USB camera and additional LED light source. Additionally, the box (11” x 17” x 12”) was crafted out of wood instead of cardboard for extended use, shown in Figure 2. The box allowed the participants to use two AutoSuture Endo DissectTM graspers (Covidien Corp., Mansfield, MA) to manipulate objects while viewing the inside of the box through an Apple iSight USB camera (Apple Inc., Cupertino, CA; 640x480 resolution) connected to a desktop computer with a 20-inch display (at 1680 x 1050 resolution). The participants were able to view the interior of the box while having their movements inside the box video recorded.

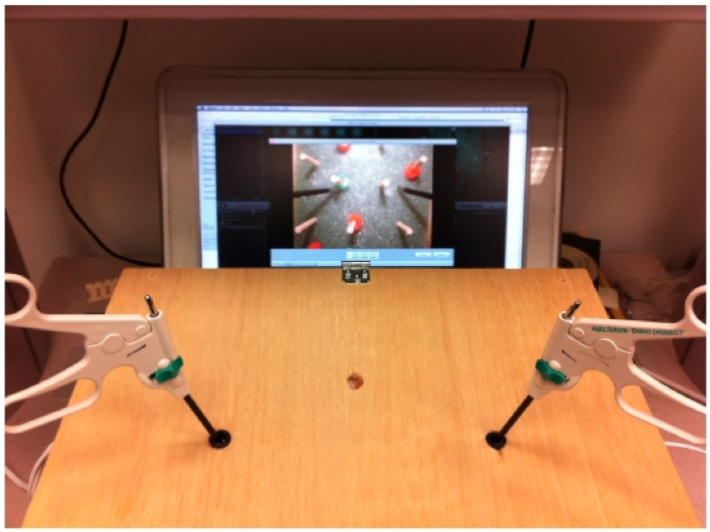


Figure 2. The laparoscopic surgery simulator box. The two graspers pass through the lid of the box; the interior of the box are displayed on the monitor, in this case cylinder positioning.

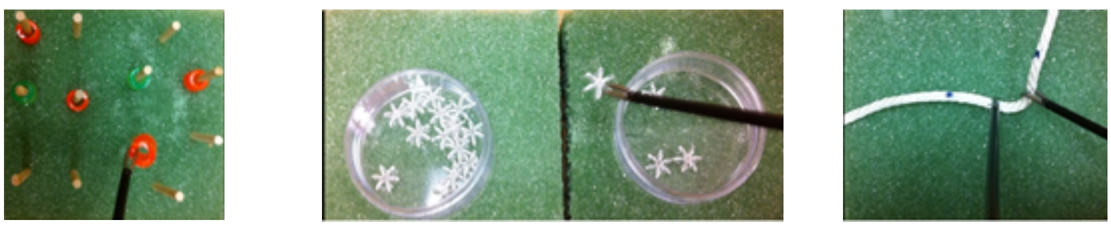
2.3 Training Tasks

During the experiment, the participants all used the laparoscopic surgery simulator to perform three different tasks in eight sessions in a separate, quiet room. The materials for the training tasks, shown in Figure 3 are based upon those used by Garcia-Ruiz, Gagner, Miller, Steiner, and Hahn[19]. The task materials were made from floral foam, wooden dowels, rubber pencil grips, a one-foot length of 1/4 inch line, fish eye hooks, beads, and small plastic receptacles.

The first task, cylinder positioning (Figure 3a) involved the participants moving 6 cylinder-shaped rubber objects from one set of pegs to another using the graspers. If a piece was dropped, the participants had to open the simulator, placed the piece in its original position by hand, close the simulator, and continued on the task.

The second task, bead drop (Figure 3b), involved the participants moving beads between two small dishes. The task was completed using only the participant's dominant hand on a grasper. Each bead was transferred, one at a time, from a receptacle containing 17 beads to an empty receptacle, until 10 beads had been moved, and then put those 10 beads back to their original receptacle. If any bead was dropped outside a receptacle, the participants could pick up the bead with a grasper and continue on the task, or use a new bead to continue.

The third task, rope grasp (Figure 3c), involved the participants grasping a loosely fastened piece of rope at seven marked positions, alternating between their dominant and non-dominant hand. Beginning at the right end of the rope, the rope was grasped on the marked positions by alternating hands until the participant reached the end, and then continued back in the same manner to the original starting position. Participants were not allowed to release the rope with both graspers, but rather had to have at least one grasper on the rope at all times. If the rope was dropped, the participant restarted the task from where they dropped the rope.



a. Cylinder positioning b. Bead drop c. Rope grasp

Figure 3. Task materials for use in the laparoscopic simulator. From left to right: (a) cylinder positioning, (b) bead drop, and (c) rope grasp.

2.4 Training Schedules

Participants were randomly assigned to one of the four groups with each group using a different training schedule (taken from Paik & Ritter[3], shown in Table 2) of eight 30-minute sessions. One schedule was distributed (8 consecutive days @ 30 minutes/day), one was massed (2 consecutive days @ 2 hours/day), one was a hybrid-massed (first and third days @ 1 hour/day, second day @ 1.5 hours, and the fourth day @ 30 minutes), and one was a hybrid-distributed schedule (first, third, fourth, fifth, and seventh days @ 30 minutes/day, eighth day @ 1.5 hours). Participants could choose a different daily time slot provided according to the schedule of the group they were assigned. However, if a participant's available time could not match any time slot for the assigned group they were assigned, we dropped the participant.

Table 2. The Four Training Schedules with Eight 30-Minute Sessions Each

	Schedule	Monday	Tuesday	Wednesday	Thursday
1st Week	Distributed (D)	D1	D2	D3	D4
	Hybrid-D (HD)	HD1		HD2	HD3
	Hybrid-M (HM)				
	Massed (M)				
2nd Week	Distributed (D)	D5	D6	D7	D8
	Hybrid-D (DH)	HD 4		HD5	HD6 HD7 HD8
	Hybrid-M (HM)	HM1 HM2	HM3 HM4 HM5	HM6 HM7	HM 8
	Massed (M)			M1 M2 M3 M4	M5 M6 M7 M8

2.5 Procedure

All participants completed the informed consent approved by the institutional review board before the experiment began. They then were given all three tasks to complete using the apparatus shown in Figure 2. In each session, participants completed the three tasks in the order shown in Figure 3 for 10 minutes per task per session. The number of times each task was completed varied, i.e., some participants completed the task more times than others in 10 minutes. Their performance was video recorded. We also collected participants' demographic data including gender, age, dominant hand, and video game experience.

Regarding the game playing experiences, we asked the participants three questions: "Are you a video game player (Y/N)?", "How many years have you played video games?", and "How many hours do you play video games every week?" The second and the third questions were hidden from those who answered "no" to the first questions.

2.6 Data Analysis

Of the original the 26 participants, one participant was dropped in the middle of the study because the device broke during the session. Two participants' data were not included in the data analysis because one was too slow to complete one repetition in the first session for task 1 and task 2 (the time to complete the task cannot be computed in this case), and the other started but could not complete the experiment due to an unrelated health issue (flu). As a result, 23 participants were coded and used in the data analysis, resulting in 6 in three groups and 5 in one group. Three coders watched and coded each video for task start and end time. The coders first coded one participant's data separately after the start and end time of each task was defined. The multiple correlation between three coders for all tasks times for this single participant was greater than .999[2]. Then the remaining 22 participants' data were coded separately by one of the three coders. Participant errors were not separately analyzed, as the dependent variable of interest is task completion time, and errors in these tasks had to be corrected before proceeding.

3 Results

We first reported a comparison of the training schedules. We then examined individual groups by several factors.

3.1 Comparison of Groups

We examined whether the four resulting groups were similar in performance on their first session. From Table 3 we can see that each group was similar in terms of gender, handedness, and age. In addition, we conducted several ANOVA tests on the first learning session for the four groups, and the results indicated that the participants from the four groups had no significant differences when completing their first training session in Task 1 [$F(3,19) = 0.18$, $p = .91$], Task 2 [$F(3,19) = 0.91$, $p = .453$], or Task 3 [$F(3,19) = 1.29$, $p = .305$]. This also suggested that the four result groups were comparable.

Table 3. Descriptive data for the four groups.

Groups		Gender		Handed		Age	
	N	Male	Female	Right	Left	Mean	SD
Distributed (D)	6	4	2	5	1	21.7	6.1
Hybrid-D (HD)	5	3	2	4	1	25.4	4.1
Hybrid-M (HM)	6	3	3	6	0	22.2	4.9
Massed (M)	6	4	2	4	2	25.2	5.5

The average task completion time in each session for each task was examined for the learning effect of the training schedule. Table 4 shows the descriptive statistics for each task, and Figure 4 shows the completion time over time (8 sessions) in each task for the four training schedule groups. Figure 5 shows the log-log plot in each task for the four training schedule groups. Visually, these schedules do not appear to have much effect on the learning rate. So, we examined inferentially each task separately.

[2] The multiple correlation was calculated using: http://www.realstatistics.com/correlation/multiple-correlation/

Table 4. Descriptive Data of Task Completion Time (in Seconds) of First Session, Last Session, and Improvement (First Session – Last Session) for the Three Tasks in the Four Training Groups.

			First learning session		Last learning session		Improvement	
		N	Mean	SD	Mean	SD	Mean	SD
Task 1:	D	6	295.0	145.2	118.6	40.6	176.3	138.2
Cylinder	M	5	282.1	155.5	92.0	12.5	190.2	143.8
positioning	HM	6	253.6	62.8	103.1	20.4	150.5	61.0
	HD	6	299.7	103.5	104.6	25.5	195.1	96.0
Task 2:	D	6	251.2	89.8	115.4	27.9	135.8	68.6
Bead	M	5	193.0	51.6	144.2	86.0	48.8	73.4
drop	HM	6	215.3	65.2	115.9	28.4	99.5	45.6
	HD	6	239.4	29.2	132.7	36.8	106.7	33.9
Task 3:	D	6	173.6	69.0	56.7	19.2	116.8	50.4
Rope	M	5	187.1	103.7	75.6	27.7	111.5	82.9
grasp	HM	6	125.9	29.1	55.3	18.5	70.6	13.0
	HD	6	193.2	43.6	75.4	18.1	117.9	38.4

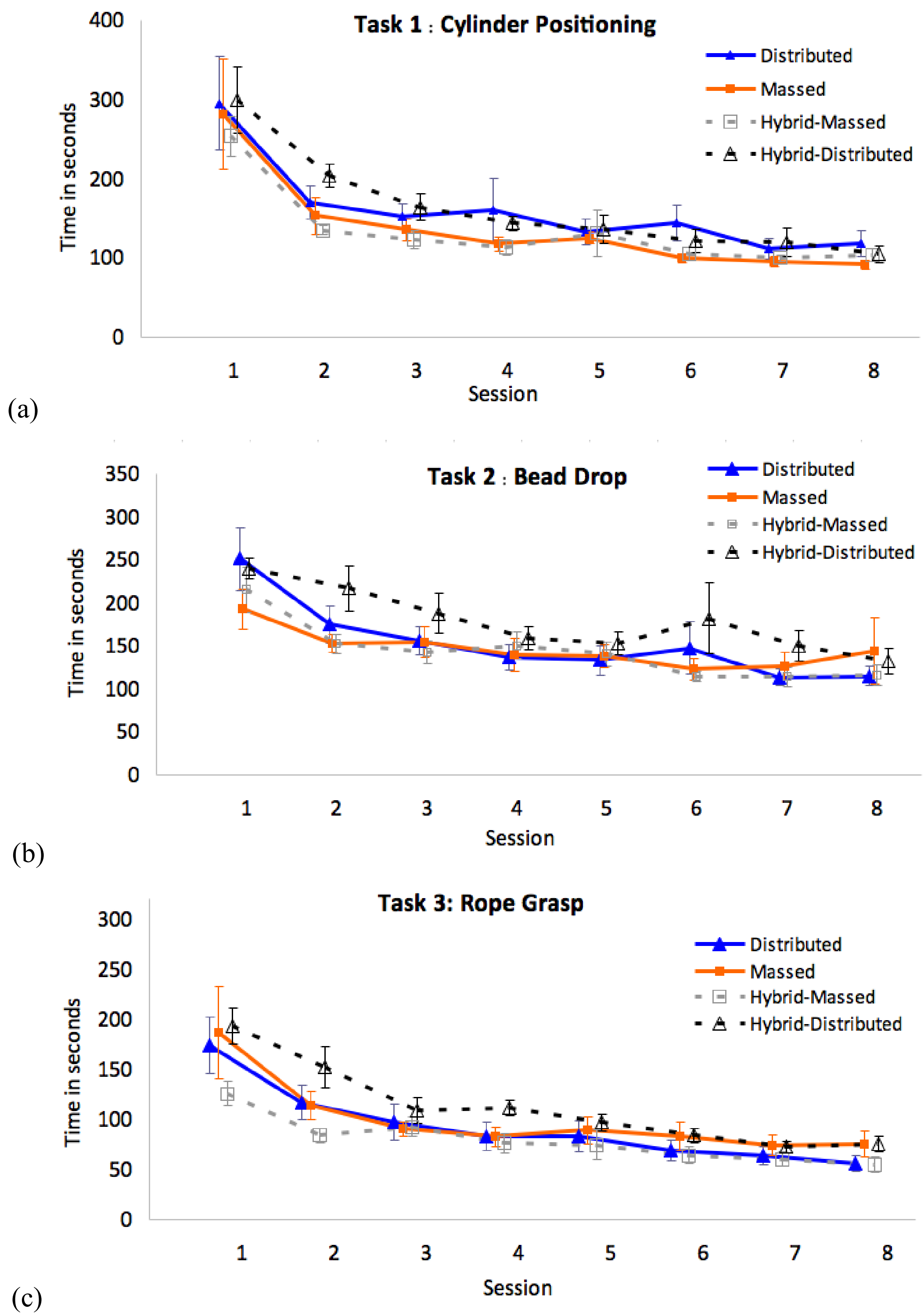
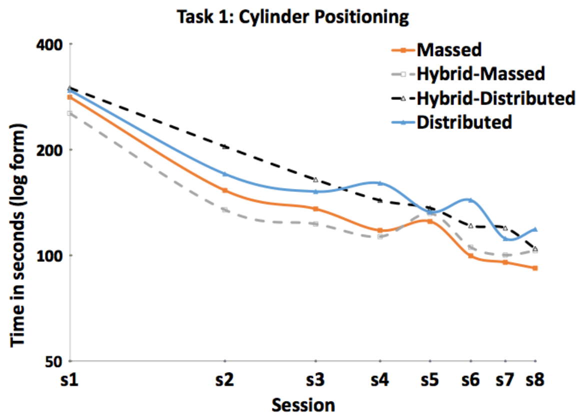
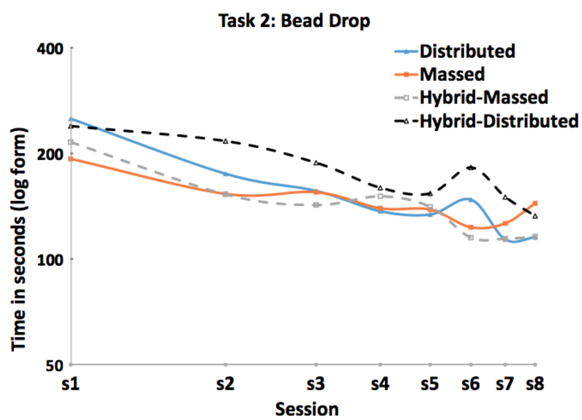


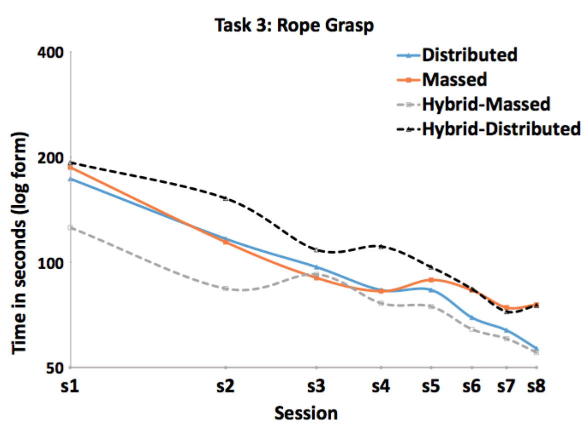
Figure 4. Task completion time on the four training schedules for (a) cylinder positioning, (b) bead drop, and (c) rope grasp. Error bars show the standard error. Series are shifted horizontally to disambiguate lines.



(a)



(b)

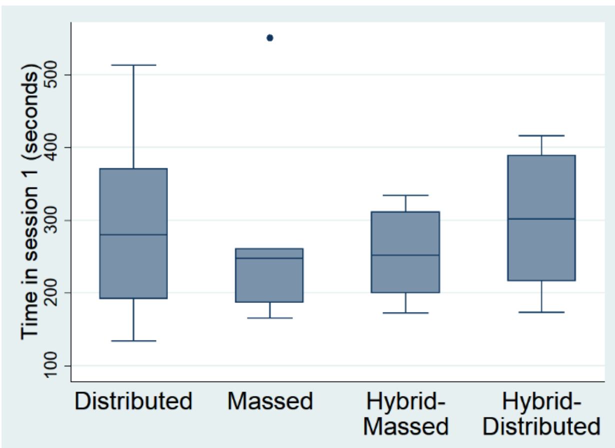


(c)

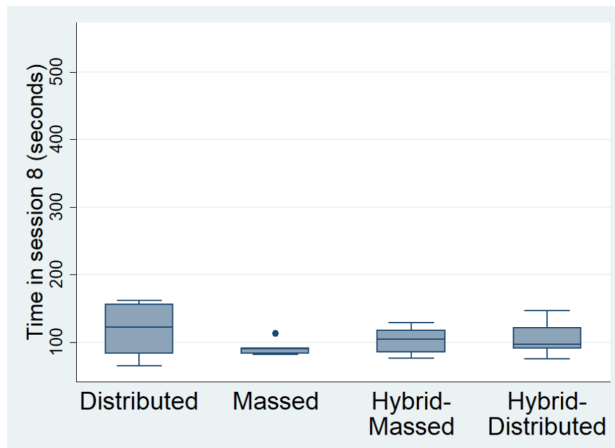
Figure 5. Log-log plot of task completion time on the four training schedules for (a) cylinder positioning, (b) bead drop, and (c) rope grasp.

Task 1: Cylinder positioning

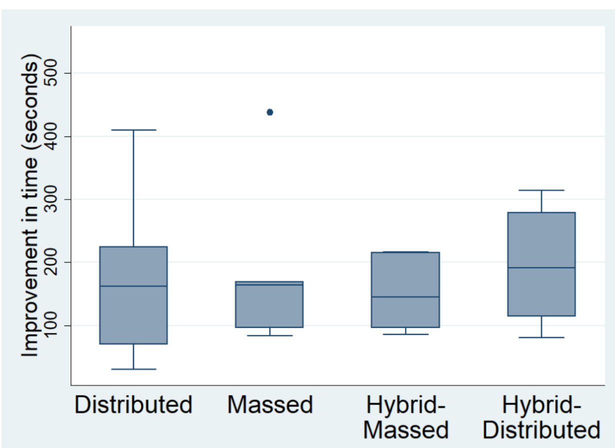
The first task was positioning a cylinder using the dominant hand. Figure 6 shows the box plots of task completion time in the first session, last session (session 8), and the improvement between the two sessions for each group. Figure 6 shows that the distributed group had the longest completion time, and the massed group had the shortest task completion time in the last session. Except for one outlier, the massed group had the smallest standard deviation in completion time in the last session, whereas the distributed group had the largest SDs.



(a) Task time in session 1 of the cylinder positioning task



(b) Task time in session 8 of the cylinder positioning task



(c) Improvement in time from session 1 to session 8 of the cylinder positioning task

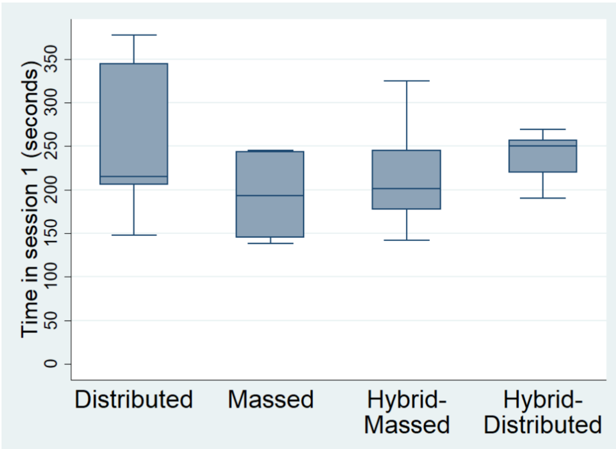
Figure 6. The box plot of (a) session 1, (b) session 8, and (c) the improvement for the cylinder positioning task.

We performed a series of ANOVA analyses on the task completion time for: (a) the first learning session, (b) the last learning session, and (c) the improvement between the first and the last sessions of the four groups. The results show that there was no significant difference on task completion time in the first session, $F(3,19) = 0.18$, $p = .91$, in the last session, $F(3,19) = 0.89$, $p = .465$, and the improvement, $F(3,19) = 0.18$, $p = .906$. These results suggest that each group started with similar task completion times when participants were first exposed to the tasks and learned at the similar rate from the first session to last session.

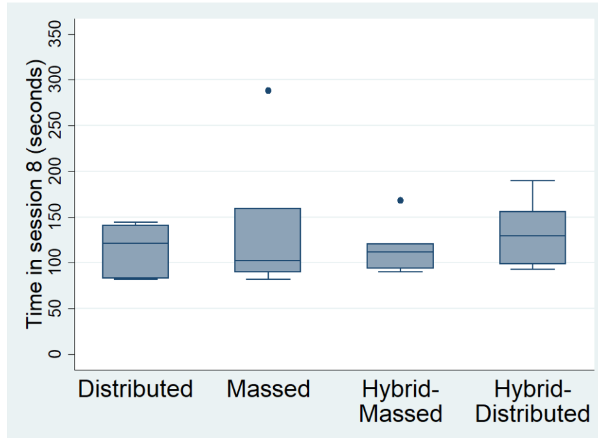
Task 2: Bead drop

The second task was moving beads between two containers using the dominant hand. Figure 7 shows the box plots of task completion time in the first session, last session, and the improvement from the first session to the last session of each training group. As we can see from Figure 7, the four groups achieved similar task completion times in the last session. Distributed and hybrid-distributed groups started with longer completion times than the massed and hybrid-massed groups in session 1 (Figure 7a), and ended with a similar pattern in session 8 (Figure 7b). Regarding the improved task completion time, the massed group had outliers, and the hybrid groups (both hybrid-massed and hybrid distributed group) had smaller standard deviations than the distributed group.

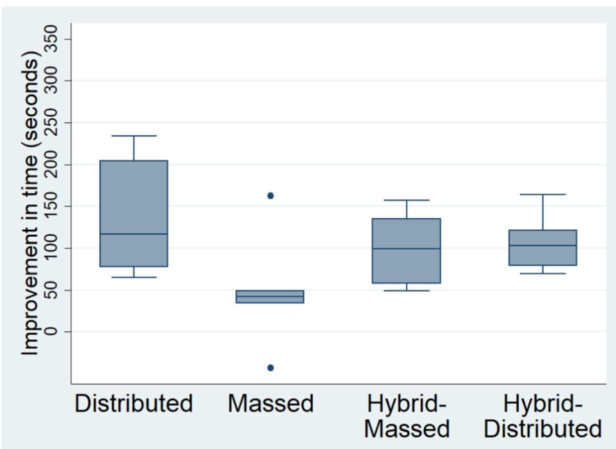
We performed a series of ANOVA analyses on task completion time for the first learning session, last learning session, and the improvement between the four schedule groups. The results show that there was no significant difference on task completion time in the first session, $F(3,19) = 0.91$, $p = .453$, the last session, $F(3,19) = 0.46$, $p = .713$, and the improvement, $F(3,19) = 2.18$, $p = .124$. These results demonstrate that each group started with similar task completion time when participants were exposed to the tasks for the first time and learned at a similar rate from the first session to the last session.



(a) Task time in session 1 of the bead drop task



(b) Task time in session 8 of the bead drop task

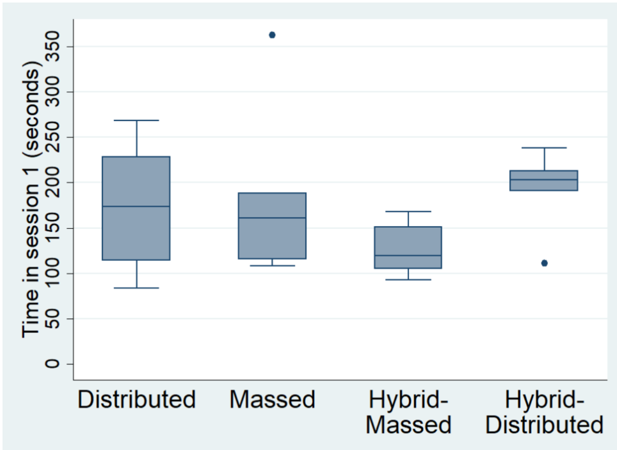


(c) Improvement in time from session 1 to session 8 of the bead drop task

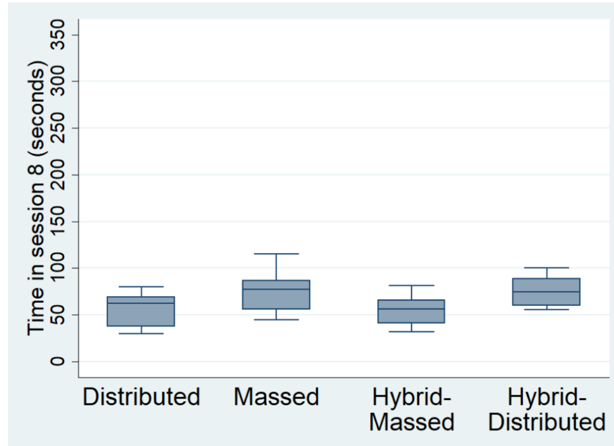
Figure 7. The box plots of (a) session 1, (b) session 8, and (c) the improvement for the bead drop task.

Task 3: Rope grasp

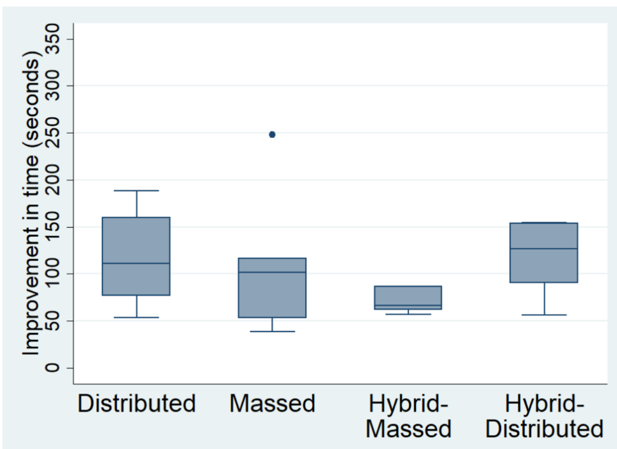
The third task was grasping a rope using both grippers alternately. Figure 8 shows the box plots of task completion time in the first session, last session, and the improvement from the first lesson to the last session of each training groups. As we can see from Figure 8, the distributed and hybrid-massed groups showed a shorter completion time, whereas the massed and hybrid-distributed groups showed a longer task completion time at the last session (Figure 8b).



(a) Task time in session 1 of the rope grasp task



(b) Task time in session 8 of the rope grasp task



(c) Improvement in time from session 1 to session 8 of the rope grasp task

Figure 8. The box plot of (a) session 1, (b) session 8, and (c) the improvement for the rope grasp task.

We performed a series of ANOVA analysis on task completion time in the first learning session, the last learning session, and the improvement of four groups. The results show that there was no significant difference on task completion time in the first session, $F(3,19) = 1.29$, $p = .305$, the last session, $F(3,19) = 1.68$, $p = .206$, or the improvement, $F(3,19) = 1.19$, $p = .339$. Each group started with similar task completion time when participants were first exposed to the tasks and learned at a similar rate from the first session to the last session.

We also analyzed participants' task completion time between sessions using a two-way ANOVA with repeated measures. The between-group variable is schedule (4 schedules), and the repeated within group variable is training session (8 sessions). Task completion time is the dependent variable. Separate ANOVA tests were repeated for the three tasks. The results show that there was no significant effect of the training schedules on task completion time (task 1: $F(3,180) = .87$; task 2: $F(3,180) = .91$; task 3: $F(3,180) = 1.41$; all $p > .05$). However, a main effect of the session was found in all tasks (task 1: $F(7,176) = 39.46$; task 2: $F(7,176) = 8.08$; task 3: $F(7,176) = 54.94$; all $p < .05$), that is, learning occurred. Tukey's HSD method[20] was used to examine the pairwise differences on task completion time between sessions. The results indicate there was a significant improvement on task time from session 1 to session 8 for all of the tasks for all training schedules.

3.2 Comparison of Individuals

We also analyzed data based on the participants' gender and gaming experiences. Because we do not have enough participants in each training schedule group, a two-way ANOVA will not produce reliable analysis. Instead, in this section, we present the descriptive results for different groups and the statistical tests by combing all groups.

Gender

There were 14 male and 9 female participants in total. Male and female participants were relatively balanced in each training schedule group, with 3 or 4 males and 2 or 3 females in each group. We compared male and female performance during the first and last learning session regardless of their training schedules.

Table 5 reports the mean and standard deviation of task completion time in each task broken down by gender. As we can see from Table 5, male participants started with a shorter completion time in the first session for all three tasks, though t-tests show no statistically reliable differences (all $p > .05$). As the training time increases, the differences between male and female participants become less, except in task 2 (although this was not reliable, $t(21) = 1.27$, $p > .05$). In task 1, female participants outperformed male participants in the last session. Two sample t-tests (with unequal variances) showed the differences between gender were significant ($t(21) = 2.4$, $p = .026$) in the last session of task 1.

Table 5. Descriptive Data of Task Completion Time for Each Task Broken Down by Gender.

			First session		Last session		Improvement	
		N	Mean	SD	Mean	SD	Mean	SD
Task 1:	Female	9	313.2	144.1	91.0	17.7	222.2	130.1
Cylinder Positioning	Male	14	263.0	89.1	114.2	28.7	148.8	80.8
Task 2:	Female	9	231.8	78.2	141.5	62.3	90.4	80.0
Bead drop	Male	14	222.4	54.1	116.5	31.6	105.9	47.9
Task 3:	Female	9	177.9	87.0	68.5	24.2	109.4	68.4
Rope Grasp	Male	14	163.6	52.2	63.3	20.8	100.3	39.0

Gaming experiences

Table 6 shows the task completion time in the first and the last session of all tasks broken down by gaming experience. Only 16 of 23 participants answered this follow-up question. Among those 16 participants, there were 7 video game players and 9 non-players. Seven out of 11 male participants self-identified as video game players. All female participants (N = 6) self-identified as non-players. Those video game players have gaming experience ranging from 2 to 25 years, with 2 to 30 hours per week. As we can see from Table 6, for the cylinder positioning task, gamers started with a lower task completion time than non-players in the first session. The differences for the bead drop task and the rope grasp task were very small. However, t-tests show that none of the differences were significant (all $p > .05$). We also notice that, in the initial learning session, game players' performances have a smaller standard deviation than non-players, especially for the rope grasp task. In that task, game players and non-gamer players started with almost the same task completion time. However, the standard deviation for non-game players was three times as much as the gamers.

Table 6. Descriptive Data of Task Completion Time Broken Down by Gaming Experiences

			First learning session		Last learning session		Improvement	
		N	Mean	SD	Mean	SD	Mean	SD
Task 1: Cylinder positioning	Non-gamer	9	295.8	130.3	103.5	30.6	192.2	116.2
	Gamer	7	262.5	99.8	111.2	32.8	151.3	97.7
Task 2: Bead drop	Non-gamer	9	219.9	60.9	126.8	63.3	93.1	69.0
	Gamer	7	224.8	42.5	116.8	39.2	108.0	45.6
Task 3: Rope grasp	Non-gamer	9	181.9	92.3	65.1	26.9	116.9	68.8
	Gamer	7	180.9	37.6	68.8	22.9	112.1	25.7

4 Discussion and Conclusions

In this study, we explored the effect of different training schedules (massed, distributed, hybrid-massed, and hybrid-distributed) on a set of tasks that have been used for laparoscopic surgery training. Participants were trained for three tasks across eight learning sessions using a laparoscopic surgery simulator.

Our results did not find a training schedule that is more effective when it comes to these motor skill learning tasks. Some reports argue that distributed practice is better than massed practice in medical skill acquisition and retention[8] and other studies provide evidence for the benefits of massed or hybrid-massed practice in supporting the learning of perceptual-motor skills[2,3,21].

Our major finding is that there appears to be no difference in completion time among the four different training schedules for these three tasks. As the tasks used in this study are procedural with large perceptual-motor components, these findings suggest that a distributed learning schedule is equally effective (i.e., not superior nor inferior) to other training schedules for training these perceptual-motor task skills. In fact, the hybrid-massed and hybrid-distributed schedule show promising results for the acquisition of procedural tasks like these because they are equally effective when comparing to the massed and distributed schedules. It appears that time on task is the most important predictor of task time for these tasks.

We believe the differences in performance time over sessions for tasks, in general, may occur due to the different types of knowledge required for the different tasks and different stages of learning the participants were in. In the KRK skill acquisition theory proposed by Kim et al.[10], distributed practice is more useful if the learners are within the declarative learning stage and massed practice is more useful when the learners are about to move to the procedural knowledge learning stage. The tasks in Moulton et al.'s study[8] required a large amount of declarative knowledge (e.g., remembering the knowledge and demonstrations in two videos and performing three different more complex surgery tasks), and, therefore, their participants may have stayed in the declarative learning stage.

In contrast, the task that could not be learned with the distributed training schedule, the inverted pendulum task in Paik and Ritter's[3] study, was a pure perceptual-motor skill task. The inverted pendulum task is relatively straightforward and does not appear to require much declarative knowledge. It was easy for participants to learn the declarative aspects of the task but they appeared to be unable to move to the procedural or perceptual-motor learning stage without some massed learning.

In our study, the tasks required mostly perceptual-motor skills with some declarative knowledge, and participants needed to use different strategies to complete the three different tasks. The participants may have been in the stage between declarative learning and procedural learning, and the tasks may not have been as novel as the inverted pendulum task. Therefore, the choice of training schedule used for other tasks may not always be the same as in the tasks studied in our research, and may vary based on the amounts and types of knowledge being learned. We will need further empirical studies to determine how each type of task benefits from different training schedules. Additional work is also needed looking at more advanced skills, longer training periods, and interventions such as feedback and deliberate practice. The results found in this study encourage more empirical studies exploring this topic on a wider variety of tasks.

We also explored individual differences in learning motor skills with different learning schedules. In terms of gender, male participants usually started with shorter task completion times than female participants. As the training continued, however, the gap decreased and females even outperformed males in one task at the last learning session. However, in a previous study, researchers found that while male medical residents appeared to complete the tasks in less time than female residents using a virtual reality laparoscopy simulator, this difference was better explained by handedness and experience with computer games[22]. Our results are encouraging because they indicate that while this group of participants may start with relatively lower performance than male participants (possibly due to handedness and experience with video games), the completion time gap between males and females appears to decrease after training across all the training schedules.

In our study, we also find participants with video gaming experience performed slightly better than participants with no video gaming experience in the cylinder positioning task and rope grasping task in the first session. These results agree with the previous studies showing that playing video games improves people's initial task time on laparoscopic techniques[23] and leads to less errors[22]. Though the differences are not significant in our study, we find the performance of people with gaming experience are similar to each other. The standard deviation of task completion for gamers was 1/3 of the time compared with people with no gaming experience in the first learning session for task 3, although the means of both groups are almost the same. This can be explained by that game players are more used to controlling 3D objects on a 2D screen by moving their wrist and hand. Therefore, regular game players may start with shorter task completion time in the initial learning stage. However, as the participants practice more, the slight initial advantage brought from gaming experience disappeared. These results agree with a previous study[23] that examined 12 related studies and concluded that video game playing is associated with a better initial performance with some surgical tasks.

One advantage of our method is that we included participants' errors into the task times. For example, in the first task, cylinder positioning, if a mistake was made, the participant had to re-position the cylinder back to its original position before continuing. For the second task, bead drop, if a bead was dropped outside of a receptacle, another bead had to be picked up from the original location before continuing. For the third task, rope grasp, if a grasper was not in the correct location, it had to be corrected. Task errors will increase the task completion time. Therefore, the performance in these tasks account for errors without requiring subjective manual coding of errors.

One limitation of this work is that no further tests were performed to test the retention of skill learned. However, according to Kim et al.'s[10,24] theory, the findings should hold provided the learning is proceduralized (because most learning theories predict that proceduralized knowledge is not forgotten). Additionally, a model of the same effect studied here could be created and tested, as this approach has proven useful in the past. In addition, a relatively small sample size might be another limitation. We recruited 26 participants and 23 of them were included in the data analysis, with 5 or 6 people each group. A post-hoc power analysis for ANOVA showed that at least 10 people per schedule are needed to detect a large effect (alpha = .05, and power = .8). Therefore, the current sample size might be another reason for the non-significant results, although we find other effects and there is not much trend across learning sessions. Also, task time is the only outcome measure we used in this paper; other outcome measures such as movement distance, and number of errors could also be considered for future study. Finally, we had 2 participants who could not complete two tasks for the first time in under 10 minutes. This, too, slightly decreased our power. Because it was the first session, their performance should not be caused by the training schedule. It does suggest that longer initial blocks might be considered in the future.

Further research can expand our work to examine other tasks and the effects of varied training schedules and amounts of training time on these additional tasks. Further, the retention of motor-skill with regards to different training schedules can be examined as well. Finally, further research could be conducted to see if the same holds true for refreshing existing skills that are out of practice instead of newly acquired skills.

To conclude, we examined the effects of four training schedules on motor-skill related tasks. The acquisition of such skills is critically important for medical training (e.g., laparoscopic and other forms of surgery), which may be improved or at least are not hurt by using a hybrid schedule. Upon examining American health systems, Kohn et al.[25] presented a four-tiered approach to improving these systems. One of these tiers is raising performance standards, which may be improved simply by evaluating and using current training schedules. Based on these results, it is apparent that training schedules for motor-skill acquisition have been studied as an important factor to consider in the surgical educational community. It appears in this data that time-on-task may be the best predictor of performance time, not training schedule. Additionally, the results can be extended to other areas, such as other tasks that require fine motor-skill movement control (e.g., playing an instrument, working with electronics), to help develop useful training schedules to improve skill acquisition.

Author contributions

Study concept and design: Frank Ritter, Ka-Chun Siu

Acquisition of data: Yu Yan

Analysis and interpretation: Yu Yan, Martin K.-C. Yeh, Frank Ritter

Study supervision and manuscript preparation: Frank Ritter, Martin K.-C. Yeh, Ka-Chun Siu, Yu Yan

Acknowledgments

We thank Joe Sanford for running a pilot version of this study and creating the box, Farnaz Tehranchi for helping run the participants, Dana Chen and Ralph Loyson for helping code the videos, and Cesar Colchado, Stephen Croker, Chris Garrison, and Peter Weyhrauch for comments that helped improve the paper. We thank Geo Surgical LLC for shipping us a replacement grasper quickly.

Funding

This research was supported by ONR grant #W911QY-07-01-0004, #N000141010401 and N00014-15-1-2275.

Declaration of Conflicting Interests

No declared conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

This study was approved by the Penn State IRB, #STUDY00005300, and by the ONR Human Research Protection Official (N000141512275-P00003).

References

1. Newell A, Rosenbloom PS. Mechanisms of skill acquisition and the law of practice. In: Anderson JR, ed. Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum, 1981: 1-51.
2. Smith CD, Scarf D. Spacing repetitions over long timescales: A review and a reconsolidation explanation. *Frontiers in Psychology* 2017;**8**(962)**:**17 pages.
3. Paik J, Ritter FE. Evaluating a range of learning schedules: Hybrid training schedules may be as good or better than distributed practice for some tasks. *Ergonomics* 2016;**59**(2)**:**276-290.
4. Figert PL, Park AE, Witzke DB, Schwartz RW. Transfer of training in acquiring laparoscopic skills. *Journal of the American College of Surgeons* 2001;**193**(5)**:**533-537.
5. Perez RS, Skinner A, Weyhrauch P, et al. Prevention of surgical skill decay. *Military Medicine* 2013; **178(10 Suppl)**(Oct)**:**76-86.
6. Perez RS. Foreword (to the special issue on "Designing and using computer simulations in medical education and training"). *Military Medicine* 2013;**178**(Supplement)**:**iv.
7. Verdaasdonk EGG, Stassen LPS, Van Wijk RPJ, Dankelman J. The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. *Surgical Endoscopy* 2007;**21**(2)**:**214-219.
8. Moulton C-AE, Dubrowski A, MacRae H, Graham B, Grober E, Reznick R. Teaching surgical skills: What kind of practice makes perfect? A randomized, controlled trial. *Annals of Surgery* 2006;**244**(3).
9. Paik J. A novel training paradigm for knowledge and skills acquisition: Hybrid schedules lead to better learning for some but not all tasks: Unpublished PhD thesis, Industrial Engineering, Penn State University, 2011.
10. Kim JW, Ritter FE, Koubek RJ. An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science* 2013;**14**(1)**:**22-37.
11. Huckin T, Haynes M, Coady J. Second language reading and vocabulary learning. Norwood, NJ: Ablex Publishing Corporation, 1993.
12. Spiro RJ, Feltovich PJ, Jacobson MJ, Coulson RL. Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology* 1991;**31**(5)**:**24-33.
13. Hulstijn J. Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research* 2002;**18**(3)**:**193-223.
14. Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. *Psychological Review* 2004;**111**(4)**:**1036-1060.
15. Anderson JR. How can the human mind exist in the physical universe? New York, NY: Oxford University Press, 2007.
16. Ritter FE, Tehranchi F, Oury JD. ACT-R: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science* 2019;**10**(3)**:**Paper e1488.
17. Alfa-Wali M, Antoniou A. Eco-friendly laparoscopic home trainer. *Simulation in Healthcare* 2011;**6**(3)**:**176-179.
18. Fried GM, Feldman LS, Vassiliou MC, et al. Proving the value of simulation in laparoscopic surgery. *Annals of Surgery* 2004;**240**(3)**:**518-528.

19. Garcia-Ruiz A, Gagner M, Miller JH, Steiner CP, Hahn JF. Manual vs robotically assisted laparoscopic surgery in the performance of basic manipulation and suturing tasks. *Archives of Surgery* 1998;**133**(9)**:**957-961.
20. Welkowitz J, Cohen BH, Lea RB. Introductory statistics for the behavioral sciences. 7th ed ed. Hoboken, NJ: John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/pensu/detail.action?docID=826865, 2011.
21. Vearrier LA, Langan J, Shumway-Cook A, Woolacott M. An intensive massed practice approach to retraining balance post-stroke. *Gait and Posture* 2005;**22**(2)**:**154–163.
22. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. *Surgical Endoscopy and Other Interventional Techniques* 2003;**17**(7)**:**1082-1085.
23. Lynch J, Aughwane P, Hammond TM. Video games and surgical ability: A literature review. *Journal of Surgical Education* 2010;**67**(3)**:**184-189.
24. Kim JW, Koubek RJ, Ritter FE. Investigation of procedural skills degradation from different modalities. Proceedings of the 8th International Conference on Cognitive Modeling 2007, Oxford, UK: Taylor & Francis/Psychology Press: 255-260.
25. Kohn LT, Corrigan JM, Donaldson MS, eds. To err is human: Building a safer health system. Washington, DC: National Academies Press, 2000.