Taylor & Francis
Taylor & Francis Group

Appears in: Ergonomics, 59(2), 276-290.

Evaluating a range of learning schedules: hybrid training schedules may be as good as or better than distributed practice for some tasks

Jaehyon Paik[a]* and Frank E. Ritter[b]

[a]*UX Laboratory, LG Electronics, Seoul, Korea;* [b]*College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA*

We investigated theoretically and empirically a range of training schedules on tasks with three knowledge types: declarative, procedural and perceptual-motor. We predicted performance for 6435 potential eight-block training schedules with ACT-R's declarative memory equations. Hybrid training schedules (schedules consisting of distributed and massed practice) were predicted to produce better performance than purely distributed or massed training schedules. The results of an empirical study ($N = 40$) testing four exemplar schedules indicated a more complex picture. There were no statistical differences among the groups in the declarative and procedural tasks. We also found that participants in the hybrid practice groups produced reliably better performance than ones in the distributed practice group for the perceptual-motor task – the results indicate training schedules with some spacing and some intensiveness may lead to better performance, particularly for perceptual-motor tasks, and that tasks with mixed types of knowledge might be better taught with a hybrid schedule.

Practitioner Summary: We explored distributed and massed training schedules as well as hybrids between them with respect to three knowledge types based on theories and an empirical study. The results suggest that industrial and operator training in complex tasks need not and probably should not be done on a distributed training schedule.

Keywords: learning; retention; training schedules; knowledge types

Introduction

Training strategies have been investigated in a wide range of fields because of the importance of training in each of these fields, including education (e.g. Bloom and Shuell 1981), psychology (e.g. Cepeda et al. 2009; Pavlik and Anderson 2005) and ergonomics and human factors (e.g. Kim, Ritter, and Koubek 2013; Moskaliuk, Bertram, and Cress 2013; Pew and Mavor 2007). For example, in ergonomics, training of and in systems is an important design consideration (Pew and Mavor 2007).

Training strategies for declarative learning (e.g. learning vocabulary) and to some extent procedural skills (e.g. learning mathematical skills) have been researched in a variety of ways, for example, second-language vocabulary (Bloom and Shuell 1981), computerised spelling drills (Fishman, Keller, and Atkinson 1968), the spacing effect on memory (Pavlik and Anderson 2005) and computing mathematical permutations (Rohrer and Taylor 2006). (These skills are analogues and components of more complex skills.)

Generally, this literature has compared the performance of two training schedules representing two extremes in the sequencing spectrum: distributed and massed training schedules. In numerous retention tests drawing on declarative and to some extent procedural memory, distributed training schedules have consistently produced better retention than massed training schedules (e.g. Bahrick 1979; Bloom and Shuell 1981; Rohrer and Taylor 2006).

These findings are further supported by studies examining the performance of complex tasks drawing not only on declarative but also perceptual-motor skills. For example, to achieve better performance on a wide range of medical skills, studies have examined the relative effectiveness of massed and distributed schedules for training medical students in laboratory settings (Moulton et al. 2006) as well as virtual reality environments (Gallagher et al. 2005; Mackay et al. 2002). The results of these studies also show better retention for distributed training schedules than for massed schedules.

The almost unanimous results of early studies examining training schedules may have contributed to the examination of only the two poles in the range of learning schedules (completely distributed and completely massed practice). The results of those studies show that a paradox of learning and retention (Schmidt and Bjork 1992), which indicates the best way to get good performance in a knowledge acquisition phase (massed practice), is bad for retaining acquired knowledge during the retention interval. On the other hand, the training schedule that gives poorer performance during the acquisition phase (distributed practice) produces better outcomes in retention tests.

*Corresponding author. Email: hyon23@gmail.com

This paradox of learning and retention may have led to exploring only these two kinds of training schedules. But, our assumption is there could be even better schedules for retaining knowledge. Bahrick et al. (Bahrick et al. 1993) investigated retention performance according to different learning intervals, such as 14, 28 and 56 days, in a vocabulary learning task. The retention tests were held 1, 2, 3 and 5 years later, and the results show that wider spacing between learning sessions gives better retention performance in all retention tests, which indicated that more widely spaced training produced the greater retention performance on a declarative memory task. Cull, Shaughnessy, and Zechmeister (1996) investigated the advantages of expanded spacing over the uniform spacing in cued recall tasks, which was originally investigated by Landauer and Bjork (1978). They explored the expanded spacing schedule (1-5-9, which are the resting periods between training sessions), uniform spacing schedule (5-5-5) and massed schedule (0-0-0), and found that both spacing schedules were superior to the massed schedule and the expanded spacing schedule outperformed the uniform spacing schedule in this task (the numbers stand for the number of intervening items between the testing phases, so it can be assumed the number is the resting period between learning sessions). Recent studies (Cepeda et al. 2009; Cepeda et al. 2008) on declarative knowledge have shown that performance on retention tests relies not only on the degree of spacing between training sessions, but also the spacing between the last training session and the retention session. That means that even if we are trained in the same amount of training sessions in distributed ways, the performance on the retention session varies depending on the resting period length between the training sessions, and the amount of the resting period between the last training session and the retention session. These results support that hybrid schedules might be superior to distributed schedules in learning declarative knowledge.

A recent review is consistent with and summarises the results of the studies that support the efficiency of hybrid schedules (Kim, Ritter, and Koubek 2013). It notes that there are three stages in learning: declarative, declarative to procedural and procedural, and the rate of memory loss when not practising varies within the stages. The initial stage of learning (declarative stage) is more susceptible to memory loss, and the final stage of learning (procedural stage) is more sustainable for learned knowledge. This review argues that periods of massed practice after declarative knowledge is learned may be a more effective way to reach the final stage of learning because the massed practice increases the declarative memory related to procedural tasks enough that it can be retrieved quickly enough to proceduralise. Their review made a case for examining hybrid schedules by citing a study examining the effects of massed practice on stroke patients (Vearrier et al. 2005). Vearrier et al. (2005), however, did not compare distributed and massed training schedules, so they only provide evidence for the relative benefits of massed practice for supporting the retention of perceptual-motor skills.

We can also imagine that massed or somewhat massed training schedules provide better learning than distributed schedules when we are learning tasks that require memories to be active enough to be proceduralised, perhaps in perceptual-motor skills, such as skiing. A hybrid schedule could lead to better learning in a more massed way (e.g. four hours in a row in two days) than in a more distributed way (e.g. one hour per day for eight days).

Schmidt and Bjork (1992) also reviewed the relationship between spacing interval and retention. They show experiment results that random condition outperformed blocked (regular) condition in motor tasks (Shea and Morgan 1979), and spaced learning is better than massed one, but increasing the spacing interval provides better retention (Landauer and Bjork 1978). These findings also support the possibility of a hybrid practice schedule that is superior to the distributed one for retaining knowledge.

There may be schedules better than fully distributed ones capable of supporting even greater retention for certain task types. We can quickly examine the performance profiles of the wide range of possible schedules by using a theory of learning and forgetting to predict retention of each possible schedule. With these predictions of the relative strength of each schedule in hand, we then empirically test some of the most promising schedules with several task types.

A theoretical study of hybrid practice

Hybrid practice is a mixed training schedule that blends distributed and massed practice. In this section, we provide a theoretical approach based on the memory equations of the ACT-R cognitive architecture to examine a wide range of learning schedules. We will use the predictions to choose schedules to test empirically.

The baseline ACT-R model

ACT-R is a functional cognitive architecture that contains theories about the mechanisms of human cognition (Anderson 2007; Anderson et al. 2004). The ACT-R architecture includes a declarative memory module and procedural memory module that store and retrieve information that corresponds to declarative knowledge and procedural knowledge. Declarative knowledge in ACT-R is represented by a set of chunks (Anderson and Schooler 1991; Miller 1956; Simon

1974). Every chunk has a numeric value, activation, and this activation reflects the strength of the chunk; higher activation leads to faster retrieval with less errors. The activation value increases with more presentations and retrieval requests. The activation A_i of a chunk i has two components, base-level activation and noise; Equations (1) and (2) show how these concepts are related.

$$A_i = B_i + \varepsilon. \tag{1}$$

The base-level activation for chunk i is:

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right), \tag{2}$$

where n: the number of presentations for chunk i; t_j: the time since the jth presentation; d: the decay parameter that is set using the base-level learning parameter.

Another important equation that decides the possibility of retrieval of a chunk is the recall probability function (Equation (3)). This is a function of a chunk's activation value and a retrieval threshold. It reflects the probability of recall at a particular moment that a model receives a retrieval request for a particular chunk.

$$P(A_i) = \frac{1}{1 + e^{(\tau - A_i)/s}}, \tag{3}$$

Evaluating a learning schedule based on an activation-based model

Pavlik (2007) and Pavlik and Anderson (2005) investigated the effects of practice and spacing on retention of a Japanese-English vocabulary task. They found more frequent practice and a longer retention interval produce the relative benefit of spacing from their empirical data, and they proposed an activation-based model that is an extension of the base-level activation equation of ACT-R 6.0. They argued that the activations of chunks increase with retrieval requests; however, these increments decay as a power function of time, so the decay rate for each presentation is not a constant, but a function that depends on the activation at the time of the presentation. Equation (4) shows the decay rate, d_i, is computed for the ith presentation of an item as a function of the activation m_{i-1}.

$$d_i(m_{i-1}) = ce^{m_{i-1}} + a, \tag{4}$$

where c: decay scale parameter; a: intercept of the decay function.

This equation provides a steady decrease for the long-term retention of a chunk when the distances between presentations are close. In contrast, as exposure spacing gets wider, activation decreases between presentations, and decay is therefore lower for new presentations, and long-term retention effects do not decrease as much (Pavlik and Anderson 2005). Using this decay function, they revised the base-level activation equation (Equation (5)), which shows the decay (d_i) is the function from Equation (4) rather than a constant variable, and the activation strength is based on the time since the previous presentation of an item and the decay value from the decay function.

$$m_n(t_{1\ldots n}) = \ln\left(\sum_{i=1}^{n} t_i^{-d_i}\right) \tag{5}$$

Evaluating the space of schedules

We use these revised equations to examine all the possible training schedules given eight learning sessions and one-retention session. Although these equations are mainly related to declarative memory learning, we believe that acquired declarative knowledge might be proceduralised with certain amount of practice, so these equations may predict the procedural learning too. The learning sessions we examined correspond to potential empirical work that we report in the next section. We assume the sessions could occur Monday through Thursday over two weeks with a 21 days retention session after the last (eighth) learning session. This is just one of many possible training situations. The training schedule could be varied in numerous ways, and we note this choice as a limitation in the conclusions.

AQ3]

We generated the recall probability (Equation (3)), which indicates the chance of recall for a learned item at each learning and retention session, for all 6435 possible schedules through the revised equation (Equation (5)) because the original equation (Equation (2)) could not predict the spacing effect (see the results based on the original equation at acs.ist. psu.edu/paik/Recall-Prob-Baseline-Model.xls, as well as the full table).

Table 1 shows the results for several schedules based on the revised equations, and Figure 1 illustrates the four exemplar schedules we tested in our empirical study. The *Schedule* column in Table 1 shows how many 30-minute training sessions are in each day. The columns that have a number from 1 to 8 are the probability of recall for each day, and the *Ret21* stands for the retention test after 21 days.

The results in Table 1 predict that a hybrid-distributed training schedule (HD: 1-0-1-1-1-0-1-3) will provide the best recall probability at the retention session. The fully distributed training schedule (D: 1-1-1-1-1-1-1-1) is not the best – its rank is 754 among 6435 (12%) – and the fully massed training schedule (M: 0-0-0-0-0-0-0-8) ranks 6103 among 6435 (94%). These results show that the fully distributed training schedule does not lead to the greatest memory activation, and that relatively, the distributed and massed schedules, when directly compared, look like what previous studies have found: a paradox of learning and retention.

Other parameters have been used in these equations, and we briefly report that the use of these other parameters does not make a substantial difference. Pavlik and Anderson also examined the results of the previous canonical studies for learning and retention through their model as Raaijmakers (2003) did. They explored the studies of Rumelhart (1967), Young (1971), Glenberg (1976), Bahrick (1979) and Bahrick and Phelphs (1987). Their analyses lead them to adjust the decay intercept (*a*) and decay scale (*c*) parameters of the memory equations. These adjusted parameters are presented in Table 2, and we predicted activation strength according to Equation (4) with these adjusted parameters.

We also used these parameters and generated the recall probabilities for possible schedules, and we found that the parameters that were used for Bahrick (1979), Bahrick and Phelphs (1987) and Glenberg (1976) resulted in the same prediction as our previous ones (the hybrid training schedule, 1-0-1-1-1-0-1-3, was the best). The parameters that were used for Rumelhart (1967) and Young (1971) predicted that 1-0-0-2-1-0-1-3 was the best training schedule. However, the 1-0-1-1-1-0-1-3 ranks at the 12th and 3rd in each result, respectively and both schedules rank higher than the fully distributed training schedule.

Table 1. Recall probability (from Equation (3)) at learning and retention sessions based on Pavlik and Anderson's memory strength Equations (4) and (5). D is the distributed, M the massed, HD and HM the hybrid-distributed and hybrid-massed schedules chosen to be tested empirically.

		Recall probability at the end of session								
Rank	Schedule	1	2	3	4	5	6	7	8	Ret21
1	**HD 10111013**	**0**	**0.0452**	**0.2773**	**0.2991**	**0.4748**	**0.6452**	**0.7834**	**0.8418**	**0.4625**
2	11011013	0	0.0719	0.2123	0.2879	0.4688	0.6420	0.7819	0.8409	0.4623
3	10111103	0	0.0452	0.2773	0.2992	0.5413	0.6026	0.7683	0.8350	0.4622
4	11011103	0	0.0719	0.2123	0.2879	0.5348	0.5991	0.7667	0.8342	0.4621
5	10111022	0	0.0452	0.2773	0.2992	0.4249	0.7076	0.7446	0.8304	0.4616
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
360	**D 11111111**	**0**	**0.0712**	**0.2922**	**0.4851**	**0.4226**	**0.6191**	**0.7160**	**0.7726**	**0.4508**
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
5847	**HM 00002321**	**0**	**0.1234**	**0.2978**	**0.5868**	**0.7224**	**0.7166**	**0.8093**	**0.8020**	**0.3845**
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6282	**M 00000044**	**0**	**0.1234**	**0.4252**	**0.6234**	**0.6301**	**0.7626**	**0.8222**	**0.8565**	**0.3654**
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6412	00000008	0	0.1234	0.4252	0.6234	0.7321	0.7952	0.8349	0.8618	0.3478
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6435	70000001	0	0.1234	0.4252	0.6234	0.7321	0.7952	0.8349	0.4538	0.3338

Note: The numbers in the schedules column stand for the number of training sessions in each training day (e.g. 10111013 schedule stands for one session at the first, third, fourth, fifth and seventh days, no session at the second and sixth days and three sessions at the eighth day.), the columns from 1 to 8 are the day number and the Ret21 stands for the retention test after 21 days.

Figure 1. The four training schedules that were chosen to be tested. The four top boxes represent Monday to Thursday of the first week, and the four bottom boxes stand for the same days of the second week. The numbers represent the number of 30-minute blocks the tasks are practised, so, for example, 3 represents 90 min. of practice on the tasks. (Fully massed, eight sessions on a single day, are difficult to schedule, and the 4-4 schedule, which we will use later, has a similar retention prediction.)

These predictions are probably still approximations and would not fully reflect the effects of more sophisticated models with multiple retrieval strategies, previous knowledge, a more sophisticated theory of proceduralisation and individual differences. But, these predictions suggest that hybrid training schedules could lead to better performance than a distributed training schedule. In the next section, we present an empirical study that tests the hypothesis that the effects of training schedules will follow the order predicted by this theory, and that the hybrid training schedules will perform better than the massed and distributed schedules.

Empirical study of hybrid and traditional learning schedules

In this empirical study, we tested the distributed, massed and two hybrid schedules. The schedule labelled Hybrid-D is the best schedule for retention in Table 1. Hybrid-M is a highly ranked but somewhat arbitrary choice that provides a somewhat massed learning schedule. We used 4-4 for the massed schedule because it was difficult to have participants do a fully massed (eight block) session, and they are predicted to have similar learning.

We designed tasks with three knowledge types: (1) a declarative task, (2) a procedural task (perhaps with some declarative components), and (3) a perceptual-motor task, to see how learning schedule interacts with knowledge types. These are component skills in many complex, real-world tasks. These can be mapped onto taxons used, for example, in the IMPRINT tool for predicting task times (Booher and Minninger 2003) or to GOMS components (John and Kieras 1994).

Participants

Forty undergraduate students (during 2010–2011) at the Pennsylvania State University were recruited and randomly assigned to conditions. None of them had experience with any of the experimental tasks. Participants were provided extra course credit or $27. Each participant performed both learning and testing sessions alone in a separate room.

Table 2. Decay intercepts and decay scales that were explored by the study of Pavlik and Anderson (2005).

	Bahrick (1979); Bahrick and Phelps (1987)	Rumelhart (1967)	Young (1971)	Glenberg (1976)
Decay intercept (a)	0.217	0.149	0.300	0.058
Decay scale (c)	0.143	0.495	0.419	0.283

Materials

A Web-based Japanese-English vocabulary test based on Pavlik and Anderson's test (2005) was designed for declarative knowledge learning and retention, as this task is very similar to previous declarative knowledge tasks and is nearly a pure declarative knowledge task. Fifteen Japanese-English word pairs were used, and the number of correct answers was used for scoring. The English words have only four-letters (e.g. base, mail), and four-to-seven letter Japanese translations were used (e.g. dodai, yuubin). All the words were assigned randomly for each participant and each trial. One word was displayed in one page as a question and participants were requested to answer the question. The correct answer was provided with a 'Correct' or 'Wrong' label in the next page.

A Web-based Tower of Hanoi puzzle was used for procedural knowledge learning and retention. This task has been used before to study procedural learning and models of it (e.g. VanLehn 1989) have consisted of nearly completely procedural knowledge. Participants were asked to move six disks from the leftmost rod to the rightmost rod with three rules: (1) only one disk can be moved at a time, (2) only the top disk can be moved to another rod, and (3) a larger disk cannot be placed on a smaller disk, and those rules were provided to the participants in every trial so that they did not need to retrieve the rules. Participants were not told the optimal procedure but asked to move the disks as efficiently as possible. The optimal number of moves, 63 moves, was displayed for motivational purpose along with their number of moves so far. The number of disk moves in moving six disks from the leftmost rod to the rightmost rod was used as the participant's score.

For the perceptual-motor skill acquisition and retention, an inverted pendulum task was used. Models of it would be made up nearly completely of perceptual-motor knowledge. This task was performed using an application (BalanceMe) running on an iPod touch that is controlled by an accelerometer in the iPod. The goal of inverted pendulum tasks is to keep the stick (pendulum) vertical by moving the pivot point (the bottom of the stick or pendulum); our task showed a stick, which is on a ball in the screen, and participants were asked to keep balancing the stick by tilting the device. The task is illustrated in Figure 2, which shows a side view and the participants' view of the task. The time duration of balancing was provided by the application itself, and was used as the participant's score.

Design and procedure

This experiment was a mixed design with a between-subjects independent variable with four levels (distributed, massed, Hybrid-D and Hybrid-M schedules, from Table 1), and task types as a within-subject factor. Participants were randomly assigned to one of the four different groups, and were run individually. Table 3 shows the training schedules, which consisted of eight (30 min. each) learning sessions and one 30 min. retention session. Each session had three tests including the Japanese-English vocabulary test, the Tower of Hanoi puzzle and the inverted pendulum task. Thus, all of the participants performed nine sessions with each task according to their training schedule.

The distributed practice group had eight learning sessions in eight days over two weeks (one session per day and four sessions per week). The massed practice group had eight learning sessions in two days over one week (four sessions at the first day and another four sessions at the second day). The Hybrid-D practice group performed eight learning sessions in six days over two weeks (one session at the first, third, fourth, fifth and seventh day, and three sessions at the eighth day). The
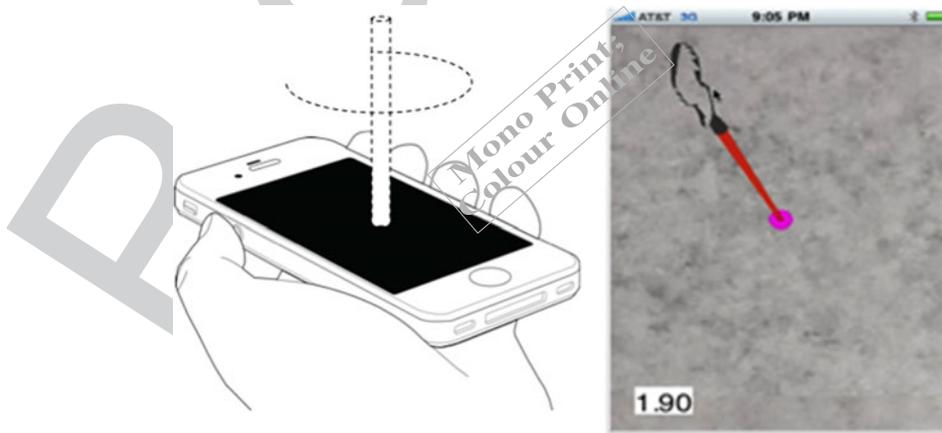
Figure 2.   A depiction of the task (left) for the inverted pendulum task (BalanceMe iPhone app) and the user interface (right). The task is to move the base, holding an iPod in your hand, so that the simulated stick stays balanced vertically.

Table 3. The four training schedules for the learning and retention experiment.

		Mon.	Tue.	Wed.	Thu.
1st week	Distributed	D1	D2	D3	D4
	Hybrid-M				
	Hybrid-D	H_D1		H_D2	H_D3
	Massed				
2nd week	Distributed	D5	D6	D7	D8
	Hybrid-M	$H_M1,2$	$H_M3,4,5$	$H_M6,7$	H_M8
	Hybrid-D	H_D4		H_D5	$H_D6,7,8$
	Massed			M1,2,3,4	M5,6,7,8
5th week	Distributed				D–Ret. test
	Hybrid-M				H_M–Ret. test
	Hybrid-D				H_D–Ret. test
	Massed				M–Ret. test

Hybrid-M practice group performed eight learning sessions in four days over one week with somewhat massed sessions (two sessions at the first day, three sessions at the second day, two sessions at the third day and one session at the fourth day). The retention tests occurred 21 days later (three weeks) after the last learning session of each group. Participants were asked not to do mental rehearsal or practise the tasks between sessions.

Results

We conducted a one-way analysis of variance (ANOVA) for the Japanese vocabulary task and the Tower of Hanoi task to test for differences among the training schedules. However, we conducted a survival analysis for the inverted pendulum task, because the data in this task increases rather than decreases with practice and may reach a limit. Survival analysis (Mills 2011) is a branch of statistics that mainly focuses on time-to-event data, such as death of organisms or failure in systems. It is widely used in engineering, economics and sociology to analyse how long a particular system or phenomenon is sustained with particular conditions so that we can predict the outcome of such conditions. As we noted in the previous section, we measured how long participants could keep balancing the stick on the ball in the device and used this as the dependent measure in the survival analysis.

The Japanese vocabulary task

Accuracy and latency were measured to test the effectiveness of the training schedules in the Japanese vocabulary task. We excluded one of the participants in the massed practice group, because his/her accuracy was an outlier (more than 2 standard deviations above the mean).

Table 4 shows the results of the descriptive statistics for the task, and Figure 3(a) shows the mean accuracy of the four practice groups at the first learning session, the last learning session and the retention session. Table 4 shows that the distributed practice group showed the lowest accuracy, and the Hybrid-D practice group showed the highest accuracy at the last learning session. However, the distributed practice group showed the highest accuracy, and the massed practice group showed the lowest accuracy at the retention session.

Table 4. Descriptive statistics for the training schedules on the Japanese vocabulary task.

		Last learning session		Retention session	
		Mean	SEM	Mean	SEM
Accuracy	D	56%	10%	51%	8%
	M	64%	7%	28%	5%
	HD	65%	6%	43%	9%
	HM	58%	9%	40%	7%
Latency	D	3.09 s	0.28 s	4.01 s	0.47 s
	M	3.00 s	0.27 s	4.79 s	0.79 s
	HD	3.03 s	0.27 s	3.50 s	0.35 s
	HM	3.18 s	0.31 s	3.41 s	0.42 s
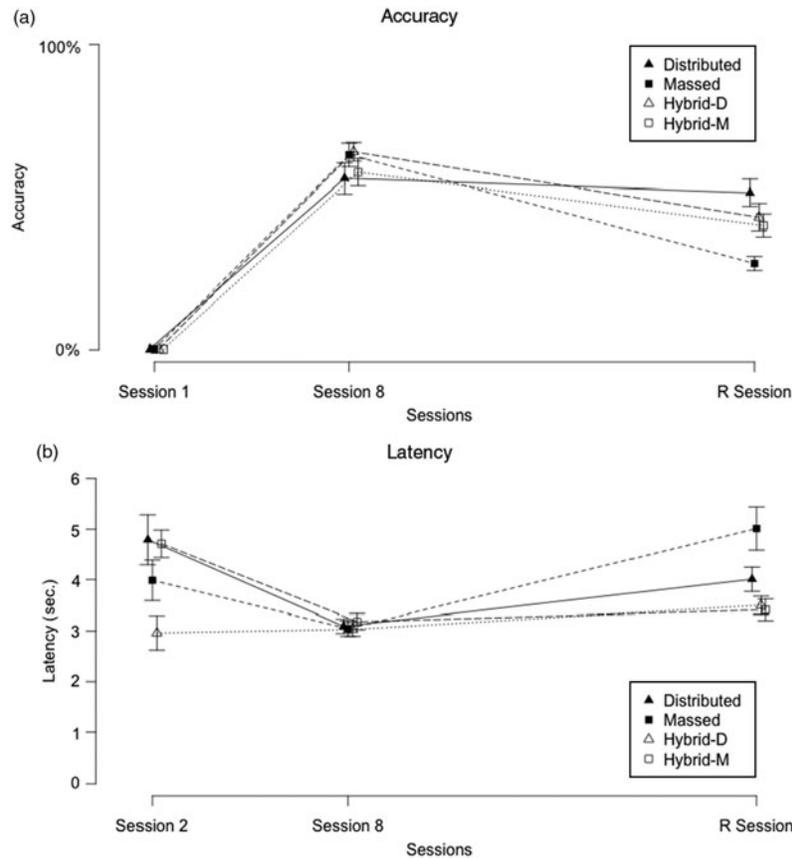
Figure 3.   The means of accuracy and latency for each session with respect to four groups ($N = 9$ for the massed group and $N = 10$ for the other groups) on the Japanese vocabulary task (error bars are standard errors of the mean, SEM). (a) Accuracy; (b) Latency.

A series of ANOVA were conducted to examine whether there were any statistical differences in learning (last learning session) and retaining (retention session) knowledge among the four training schedules. The results revealed that there were no significant differences in accuracy across schedules at the last learning session, F (3,35) $= 0.256$, $p = 0.857$, $\eta^2 = 0.021$, and at the retention session, F (3,35) $= 1.45$, $p = 0.246$, $\eta^2 = 0.11$.

Figure 3b shows the mean latency of the four groups at the second learning session (participants could not answer in Session 1 because they did not yet know the associations), the last learning session and the retention session. The mean latency at the last learning was almost identical; however, the performances of all groups at the retention session showed little differences. We conducted a series of ANOVA among groups to examine whether there were any statistical differences in latency at the last learning session and retention session, and no statistical differences were observed (all $F's$ (3,35) < 3, $p's > 0.05$).

We also conducted a series of independent t-tests to examine the effects of training schedules in this task and found that the massed practice group has lower accuracy than the distributed practice group at the retention session (t (17) $= 2.24$, $p = 0.038$, $d = 1.08$), which is consistent with several previous studies. However, using Bonferroni adjusted alpha level 0.0083 (0.05/6), we could not find any reliable differences between the two schedules. The corrected alpha is too small in our study, which might lead to Type 2 errors (misses), so we will not provide any further analyses of this task.

But visually, the results are suggestive that hybrid learning schedules may be useful, and the results at least suggest that further study is warranted of non-standard training schedules that are often found in the real world, what with cancelled classes, holidays, weekends and other interruptions to a learning schedule.

The Tower of Hanoi

The number of disk moves and task completion time for the Tower of Hanoi were examined for the effect of training schedules on learning and retention. We excluded five trials where the number of disk moves was greater than 500 or where

Table 5. Descriptive statistics for the four training schedules on the Tower of Hanoi task.

		First learning session		Last learning session		Retention session	
		Mean	SEM	Mean	SEM	Mean	SEM
Moves	D	148.77	18.44	84.60	10.65	75.77	5.20
	M	130.55	14.35	73.88	6.59	85.00	11.25
	HD	176.25	35.03	80.25	11.48	98.00	15.33
	HM	220.00	44.21	92.88	12.84	100.44	7.89
Latency(in s)	D	383.57	36.78	149.38	22.35	126.35	10.56
	M	312.88	39.93	138.63	22.78	171.82	34.20
	HD	349.46	72.55	103.60	14.38	126.62	19.06
	HM	575.86	91.72	155.57	26.49	167.07	14.52

task completion time was greater than 1200 sec. Table 5 shows the descriptive statistics for the task, and Figure 4 shows the average number of disk moves and task completion time at the first learning session, the last learning session and the retention session with respect to the four practice groups.

We performed a series of ANOVA for the number of disk moves and task completion time among the groups at the first learning session. We found that there was no statistical difference in the number of disk moves among groups; however, there were statistical differences in task completion time, $F(3,31) = 3.41$, $p = 0.02$, $\eta^2 = 0.24$. *Post hoc* comparison using the Tukey HSD test indicated that the average first learning session task completion time for the massed practice group ($M = 312.9$ s, SD = 119.8) was significantly different than the Hybrid-M group ($M = 575.8$, SD = 275.2) with the massed group taking less time. However, there were no significant differences in comparing the other groups.
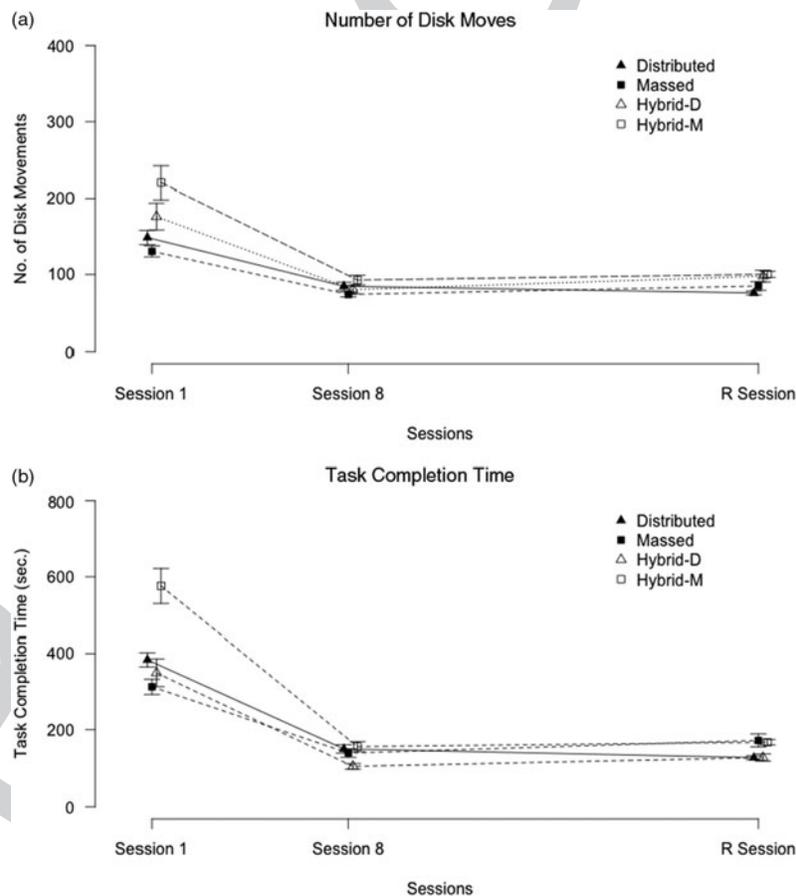
Figure 4. (a) The means of the number of disk moves and (b) task completion time for the first learning session ($N = 9$ for the distributed, massed, and Hybrid-M groups and $N = 8$ for the Hybrid-D group), the last learning session and retention session with respect to four practice groups (error bars are SEM) on the Tower of Hanoi task. (a) Number of disk moves (b) task completion time.

496   The number of moves and task completion time gradually decreased during the learning sessions, and the performance
497 of all groups were almost identical at the last learning session (see Figure 4). To examine any differences in learning among
498 the groups, we conducted a series of ANOVA for the performance of the last learning session. The results of analyses
499 indicate there were no significant differences among the groups in both measurements (because of differences between
500 massed and Hybrid-M groups in the task completion time at the first learning session, we analysed the change in
501 performance for the task completion time).

502   We also examined whether or not there were any performance differences according to the training schedules in
503 retaining knowledge for the Tower of Hanoi task. Figure 4 provides an overview, showing the performances of all practice
504 groups at the retention session were almost identical to the last learning session, which indicates participants retained their
505 learned knowledge regardless of training schedule, and the acquired procedural knowledge was not decayed by time.
506 A more detailed plot (Paik 2011, Figures 4-9 and 4-10) shows nearly identical curves for all schedules by block. While the
507 time between sessions varied for the data points, their values did not. The results of the ANOVA in both measurements also
508 showed that there were no reliable differences among the groups.

The inverted pendulum task

512 The learning curves of the inverted pendulum task look different than the other tasks; they use an inverted measure (time the
513 pendulum can be held), where the task time goes up with practice but also the curves are quite distinct. Table 6 shows the
514 descriptive statistics for this task. The results of the ANOVA for the first learning session in the task showed there were no
515 significant differences among the groups, $F(3,36) = 0.81, p = 0.49, \eta^2 = 0.06$.

516   Figure 5 shows that each group starts out with approximately the same balance time. Figure 5 also shows that each
517 training schedule had a different shaped learning curve and final performance. The balance time of the participants in the
518 distributed group had little variance, and performance was very low during the learning sessions and did not increase at the
519 retention session. The participants in the Hybrid-M group performed the best at the last learning session and showed little or
520 no forgetting at the retention session.

521   We conducted a survival analysis to examine which training schedule gives better performance during the learning
522 session and the retention session. Figure 6(a) shows the results of survival analysis of the last learning session for all groups.
523 The graph shows that the participants who learned this task in a distributed way could not survive (keep balancing) longer
524 than the other participants who trained on different schedules. Participants in the Hybrid-M group could survive longer than
525 the other groups.

526   To examine the learning effect of different training schedules, we conducted a Cox proportional hazard regression and
527 found that there was a significant difference ($p = 0.01$) between the distributed group and Hybrid-M group with the Hybrid-
528 M group balancing longer.

529   We also conducted survival analysis of the retention session, and the results are presented in Figure 6b. As the graph
530 shows, the results are almost identical to those of the last learning session, except the massed practice group survived longer
531 than the Hybrid-D group (opposite to their relative order in the last learning session). The participants in Hybrid-M group still
532 survived longer than the other groups in this session. The results of the Cox proportional hazard regression shows that there
533 was a significant difference ($p = 0.02$) between the distributed practice group and the Hybrid-M practice group. In this task,
534 the schedules lead to reliably different learning curves and the differences appear to be of both theoretical and practical
535 importance.

Discussion and conclusions

539 We first summarise the theoretical analysis and study results. Then, we consider the practical implications and future work
540 that arises from these results.

Table 6.   Descriptive statistics for the inverted pendulum task.

		First learning session		Last learning session		Retention session	
		Mean	SEM	Mean	SEM	Mean	SEM
Duration time	D	17.54	5.80	29.87	7.39	25.38	4.77
	M	13.65	4.21	74.12	31.51	109.00	39.10
	HD	9.32	2.22	98.32	44.03	76.66	38.41
	HM	11.68	1.85	148.26	40.35	146.35	38.91
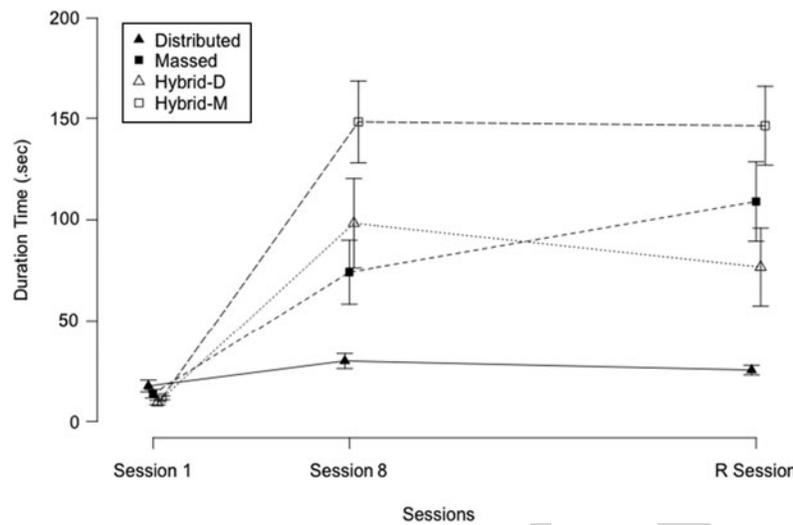
Figure 5. The means of balance time for the first learning session ($N = 10$ for each point), the last learning session and the retention session with respect to four practice groups (error bars are SEM) for the inverted pendulum task.
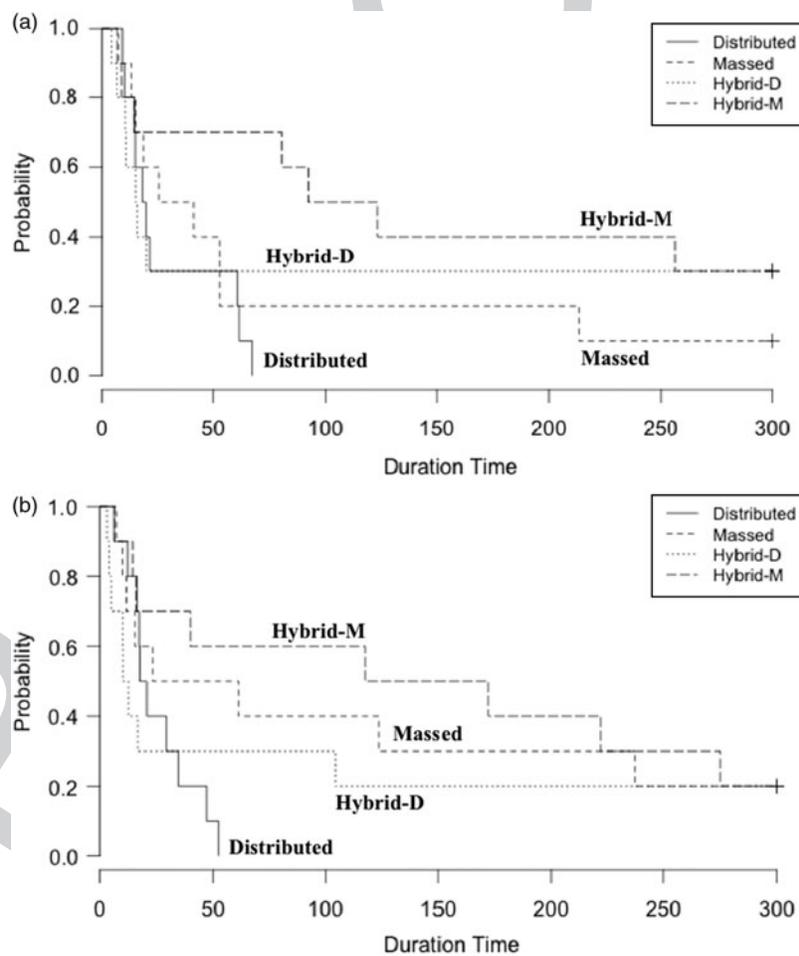
Figure 6. Survival probability curves with respect to all practice groups in the (a) last learning session and (b) retention session. The Hybrid-M and distributed curves are reliably different in both cases with $p = 0.01$ and $p = 0.02$, respectively.

Summary of analysis and results

In this study, we explored a range of training schedules that might be superior to the equally spaced distributed and massed training schedules previously used with respect to three types of knowledge. We theoretically examined the 6435 possible training schedules that arise from eight learning sessions with a 21-day retention session. We predicted the performance of all training schedules through the base-level learning equations of ACT-R.

By using a revised base-level learning equation, we could predict the performance of learning and retention sessions for all training schedules. From those schedules, we selected four: (1) the one that predicts the highest performance at the retention test (Hybrid-D, 1-0-1-1-1-0-1-3), (2) a massed (M: 0-0-0-0-0-0-4-4), (3) a distributed (D: 1-1-1-1-1-1-1-1) schedule that has been widely used in previous studies, and (4) a hybrid-massed schedule (Hybrid-M, 0-0-0-0-2-3-2-1). We then examined these schedules with three different tasks that include a declarative knowledge task, procedural knowledge task and perceptual-motor task.

The results of the declarative knowledge task showed that there were no significant differences among the groups, which is a different result from numerous previous studies. We believe that this result might be from the structure of our massed training schedule. Our massed training schedule was not a completely pure massed training schedule (4-4 sessions, not 8 sessions in a day) compared to previous studies. The space between the 2-4 sessions might help participants retain knowledge. It could also be the case that the sample was small and/or unusual. It is also possible that for this type of task, the most efficient training schedule could vary, but the hybrid-distributed schedule might be good. A hybrid schedule might also be seen in practice more often than a pure distributed schedule.

The results of the procedural knowledge task (Tower of Hanoi) showed no significant differences among the four different practice groups – the learning curves are almost identical. Similarly, knowledge was retained during the retention interval for all groups; the performance between the last learning session and retention session was almost identical; the procedural knowledge did not decay over this time period. More detailed analysis based on session finds a similar pattern at that level (Paik 2011). This result suggests that the most important factor for acquiring and retaining knowledge and skills in a task that requires procedural knowledge is not the training schedule, but simply the amount of practice.

The lack of change in performance over time from the last practice to the retention test also suggests that decay for procedural knowledge is substantially lower than for declarative knowledge. This is an assumption that many learning theories make (Anderson 1982; Fitts 1954, 1964; Kim, Ritter, and Koubek 2013; Ohlsson 2011; Rasmussen 1986) and a few studies appear to find (e.g. Cooke, Durso, and Schvaneveldt 1994; Stefanidis et al. 2006). However, our results are different from the study of Rohrer and Taylor (2006), which examined the effect of training schedules in a mathematical computation task, and showed that the distributed group was superior to the massed group. We believe that the main reason for the different result might be due to different tasks. Both tasks require procedural knowledge; however, the mathematical task that they used also requires declarative knowledge, such as declarative math facts and memorisation of how to solve the problem (using a novel formula), and most of the participants in the massed practice group might forget this formula at the four-week retention test, because the declarative knowledge is more sensitive to decay over time.

For the perceptual-motor skill task, which we argue is the most important contribution in this study, we found Hybrid-M led to better learning and retention, and that there were significant differences between the Hybrid-M and the distributed practice groups. This result examines and supports a theory of skill retention (Kim, Ritter, and Koubek 2013) and our hypotheses that learning should occur with different degrees and ways in each stage of learning, and that the tasks that require perceptual-motor skills should be trained with a massed or somewhat massed schedule to reach the procedural stage of learning. The result also indicates that the learning in the task that required perceptual-motor skill was more influenced by the training schedules than the other tasks, and massed or somewhat massed training schedules lead to better performance than the purely distributed schedule.

We also found that our results were different from the previous studies, which showed that distributed practice was better than massed practice in medical skill acquisition and retention (Moulton et al. 2006). We believe the reason is the same as the differences in the procedural task, which is that the previous medical skill task depended on a large amount of declarative knowledge; however, our inverted pendulum task was more purely a perceptual-motor skill task and does not appear to require declarative knowledge. It may also be the case that this task itself is more sensitive to the effect of different training schedules. We can note these problems as an area for future work.

From the results of our study, we could say that a fully distributed schedule is not superior to the other schedules in retaining acquired knowledge and skills except for declarative knowledge. Learning procedural tasks appears to mainly depend on the amount of time to practise rather than the different schedules. Furthermore, our results suggest that tasks that require perceptual-motor skill should be learned with some intensive practice that can be obtained from massed or hybrid-

massed schedules. In real life, however, there are tasks that require both declarative and perceptual-motor skills, such as learning with a virtual surgery system. Traditionally, the way to train with that system is more likely to use a distributed schedule, but it is probably better to train with a mixed schedule; a massed or hybrid way to train how to control the system, and a distributed way to train where and why to move the controller.

The difficulty for the classification of tasks based on a particular knowledge type might be a problem of this study. The Japanese vocabulary and the inverted pendulum tasks mainly required declarative knowledge and perceptual-motor skill respectively, so we can categorise those two tasks fairly clearly as declarative and perceptual-motor tasks. The Tower of Hanoi task, however, might require mixed knowledge because there are rules that participants should know to solve the problem. We provide the rules before participants performed the task to make participants mainly use their procedural knowledge instead of mixed knowledge, but it may have both types of knowledge. A study that provides more clear information about task categorisation might be needed. In addition, further tasks that can be categorised particular knowledge types or that use mixed knowledge types should be studied to provide general guideline for learning and retention.

Although there are unresolved issues and room for future work, our study suggests a new training paradigm by examining schedules based on learning equations and testing non-equally distributed learning schedules. We showed that a hybrid-distributed/massed schedule, which is different from the widely used approach in the previous studies, did lead to greater learning in a perceptual-motor task. Most of the previous studies simply examined distributed practice and massed practice, and compared performance at retention. However, our study, especially in the theoretical analysis, examined all the possible training schedules and predicted the performance for all learning and retention sessions.

Theoretically, the schedules are different, and the distributed schedule is not the best. In addition, the declarative schedule did not lead to greater learning empirically in these tasks – this could be due to low power, but there were differences pair-wise for the declarative vs. massed, and for all schedules on the perceptual-motor task, which makes this hypothesis less likely. Our theoretical analyses in this study can be used by other researchers who want to explore the range of training schedules by gathering data on them.

Practical suggestions for training schedules

For practitioners and industrial trainers, this work makes suggestions about how to train complex tasks. The results suggest that mostly declarative tasks can continue to be trained with distributed training schedules, but that they can be safely and perhaps more profitably trained with a less distributed schedule and may be more robustly learned across training schedule types than we may have initially thought. These tasks would include, for example, learning industrial signs (Chan and Ng 2010).

The results suggest that procedural tasks, similar to many plant operations (e.g. Sauer et al. 2008) and vehicle operations (e.g. Bellet and Tattegrain-Veste 1999; Petersen and Barrett 2009) could be trained with hybrid schedules because time-on-task seems to be the most important aspect for learning – across training schedules learning appears to occur at the same rate.

The results suggest that for perceptual-motor tasks, including many hospital (e.g. Warming et al. 2008) and manufacturing tasks (e.g. Koubek, Clarkston, and Calvez 1994), and even sport (Annett 1994), that massed and particularly hybrid-massed learning schedules will be better than other schedules and that distributed schedules in particular should be avoided.

Future work

The other problem that we found is the revised ACT-R base-level learning equation failed to fully predict our empirical results. We tested several parameters of decay intercept and decay scale that predicted the results of some canonical studies, and found that the Hybrid-D training schedules is superior to the purely distributed one in our experiment setting. However, the results of our empirical study (Japanese-English vocabulary task) showed no significant difference in accuracy between the distributed practice group and the Hybrid-D practice group. There might be noise or the theoretical performance difference between the two training schedules might be relatively too small (see Table 1) to find reliable differences of those schedules. The number of the participants might be seen as small ($N = 10$ per schedule), because when we use the power analysis for ANOVA design (Cohen 1988), the required sample size is 18 per each group with large effect size (with 0.05 for alpha, and 0.8 for power); however, we designed a multi-session (eight sessions) study with a test after a retention period, and the subject sessions in our study represent the resources of a more than four times more subjects than single-session studies. In addition, the sample size per schedule that we used is fairly common, that of 10–15 subjects per condition in a complex study (e.g. Pavlik and Anderson 2005; VanLehn 1989).

However, these results support a recent study (Cepeda et al. 2008) that found the revised base-level equation could not predict their results of an empirical study either. Therefore, the revised base-level learning equation for predicting complex learning and retention spacing over multiple sessions could probably still be improved.

Like all studies, this study is limited. It would be interesting to gather empirical data of further retention intervals, a larger number of training sessions, more total time training, additional tasks, less or more concurrent tasks, or of another retention interval, which might give more evidence for our results. In addition, we did not record many aspects of the participants' demographic information; this is not always or perhaps often done in some learning studies (Pavlik and Anderson 2005; Reder and Ritter 1992), but we would recommend doing so when to support secondary analyses. All of these areas represent limitations of the current study and areas for future work.

We also need theories that might be able to predict the performance of procedural and perceptual-motor skills tasks. The theory that we used in this study was based on declarative knowledge and our assumption was that acquired declarative knowledge might be proceduralised with certain amount of practice, so the declarative knowledge theory might predict the learning for the other knowledge types. However, the shape of learning curves might vary in different knowledge types (Kim, Ritter, and Koubek 2013), so particular theories or equations with respect to particular knowledge types to predict the learning performance are still needed, and further hybrid learning schedules might yet be examined.

References

[AQ4]

Anderson, J. R. 1982. "Acquisition of Cognitive Skill." *Psychological Review* 89 (4): 369–406. doi:10.1037/0033-295X.89.4.369.

Anderson, J. R. 2007. *How Can the Human Mind Exist in the Physical Universe?* New York, NY: Oxford University Press.

Anderson, J. R., D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. 2004. "An Integrated Theory of the Mind." *Psychological Review* 111 (4): 1036–1060. doi:10.1037/0033-295X.111.4.1036.

Anderson, J. R., and L. J. Schooler. 1991. "Reflections of the Environment in Memory." *Psychological Science* 2 (6): 396–408. doi:10.1111/j.1467-9280.1991.tb00174.x.

Annett, J. 1994. "The Learning of Motor Skills: Sports Science and Ergonomics Perspectives." *Ergonomics* 37 (1): 5–16. doi:10.1080/00140139408963617.

Bahrick, H. P. 1979. "Maintenance of Knowledge: Questions about Memory We Forgot to Ask." *Journal of Experimental Psychology: General* 108 (3): 296–308. doi:10.1037/0096-3445.108.3.296.

Bahrick, H. P., L. E. Bahrick, A. S. Bahrick, and P. E. Bahrick. 1993. "Maintenance of Foreign Language Vocabulary and the Spacing Effect." *Psychological Science* 4 (5): 316–321. doi:10.1111/j.1467-9280.1993.tb00571.x.

Bahrick, H. P., and E. Phelphs. 1987. "Retention of Spanish Vocabulary over 8 Years." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13 (2): 344–349. doi:10.1037/0278-7393.13.2.344.

Bellet, T., and H. Tattegrain-Veste. 1999. "A Framework for Representing Driving Knowledge." *International Journal of Cognitive Ergonomics* 3 (1): 37–49. doi:10.1207/s15327566ijce0301_3.

Bloom, K. C., and T. J. Shuell. 1981. "Effects of Massed and Distributed Practice on the Learning and Retention of Second-Language Vocabulary." *Journal of Educational Research* 74 (4): 245–248. doi:10.1080/00220671.1981.10885317.

Booher, H. R., and J. Minninger. 2003. "Human Systems Integration in Army Systems Acquisition." *Handbook of Human Systems Integration*: 663–698.

[AQ5]

Cepeda, N. J., N. Coburn, D. Rohrer, J. T. Wixted, M. C. Mozer, and H. Pashler. 2009. "Optimizing Distributed Practice." *Experimental Psychology* 56 (4): 236–246. doi:10.1027/1618-3169.56.4.236.

Cepeda, N. J., E. Vul, D. Rohrer, J. T. Wixted, and H. Pashler. 2008. "Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention." *Psychological Science* 19 (11): 1095–1102. doi:10.1111/j.1467-9280.2008.02209.x.

Chan, A. H., and A. W. Ng. 2010. "Effects of Sign Characteristics and Training Methods on Safety Sign Training Effectiveness." *Ergonomics* 53 (11): 1325–1346. doi:10.1080/00140139.2010.524251.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Cooke, N. J., F. T. Durso, and R. W. Schvaneveldt. 1994. "Retention of Skilled Search after Nine Years." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 36 (4): 597–605.

Cull, W. L., J. J. Shaughnessy, and E. B. Zechmeister. 1996. "Expanding Understanding of the Expanding-Pattern-of-Retrieval Mnemonic: Toward Confidence in Applicability." *Journal of Experimental Psychology: Applied* 2 (4): 365–378. doi:10.1037/1076-898X.2.4.365.

Fishman, E. J., L. Keller, and R. C. Atkinson. 1968. "Massed versus Distributed Practice in Computerized Spelling Drills." *Journal of Educational Psychology* 59 (4): 290–296. doi:10.1037/h0020055.

Fitts, P. M. 1954. "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement." *Journal of Experimental Psychology* 47 (6): 381–391. doi:10.1037/h0055392.

Fitts, P. M. 1964. "Perceptual-Motor Skill Learning." In Categories of Human Learning, edited by A. W. Melton, 381–391. New York, NY: Academic Press.

Gallagher, A. G., E. M. Ritter, H. H. Champion, G. G. Higgins, M. P. Fried, G. G. Moses, C. D. Smith, and R. M. Satava. 2005. "Virtual Reality Simulation for the Operating Room: Proficiency-Based Training as a Paradigm Shift in Surgical Skills Training." *Annals of Surgery* 241 (2): 364–372. doi:10.1097/01.sla.0000151982.85062.80.

Glenberg, A. M. 1976. "Monotonic and Nonmonotonic Lag Effects in Paired-Associate and Recognition Memory Paradigms." *Journal of Verbal Learning and Verbal Behavior* 15 (1): 1–16. doi:10.1016/S0022-5371(76)90002-5.

John, B. E., and D. E. Kieras. 1994. *The GOMS Family of Analysis Techniques: Tools for Design and Evaluation*. Pittsburgh, PA: School of Computer Science, Carnegie-Mellon University.

Kim, J. W., F. E. Ritter, and R. J. Koubek. 2013. "An Integrated Theory for Improved Skill Acquisition and Retention in the Three Stages of Learning." *Theoretical Issues in Ergonomics Science* 14 (1): 22–37. doi:10.1080/1464536X.2011.573008.

Koubek, R. J., T. P. Clarkston, and V. Calvez. 1994. "The Training of Knowledge Structures for Manufacturing Tasks: An Empirical Study." *Ergonomics* 37 (4): 765–780. doi:10.1080/00140139408963687.

Landauer, T., and R. A. Bjork. 1978. "Optimum Rehearsal Patterns and Name Learning." *Practical Aspects of Memory* 1: 625–632.

Mackay, S., P. Morgan, V. Datta, A. Chang, and A. Darzi. 2002. "Practice Distribution in Procedural Skills Training." *Surgical Endoscopy.* 16 (6): 957–961. doi:10.1007/s00464-001-9132-4.

Miller, G. A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (2): 81–97. doi:10.1037/h0043158.

Mills, M. 2011. *Introducing Survival and Event History Analysis*. Sage.

Moskaliuk, J., J. Bertram, and U. Cress. 2013. "Training in Virtual Environments: Putting Theory into Practice." *Ergonomics* 56 (2): 195–204. doi:10.1080/00140139.2012.745623.

Moulton, C. A. E., A. Dubrowski, H. MacRae, B. Graham, E. Grober, and R. Reznick. 2006. "Teaching Surgical Skills: What Kind of Practice Makes Perfect? A Randomized, Controlled Trial." *Annals of Surgery* 244 (3): 400–409.

Ohlsson, S. 2011. *Deep Learning: How the Mind Overrides Experience*. New York, NY: Cambridge University Press.

Paik, J. 2011. "A Novel Training Paradigm for Knowledge and Skills Acquisition: Hybrid Schedules Lead to Better Learning for Some but not All Tasks (PhD diss.)". Department of Industrial Engineering, The Pennsylvania State University, University Park.

Pavlik, P. I. 2007. "Timing Is in Order: Modeling Order Effects in the Learning of Information." In *In Order to Learn: How the Sequences of Topics Affect Learning*, edited by F. E. Ritter, J. Nerb, E. Lehtinen, and T. O'Shea, 137–150. New York, NY: Oxford University Press.

Pavlik, P. I., and J. R. Anderson. 2005. "Practice and Forgetting Effects on Vocabulary Memory: An Activation-based Model of the Spacing Effect." *Cognitive Science* 29 (4): 559–586. doi:10.1207/s15516709cog0000_14.

Petersen, A., and R. Barrett. 2009. "Postural Stability and Vehicle Kinematics during an Evasive Lane Change Manoeuvre: A Driver Training Study." *Ergonomics* 52 (5): 560–568. doi:10.1080/00140130802419495.

Pew, R. W., and A. S. Mavor, eds. 2007. *Human-System Integration in the System Development Process: A New Look*. Washington, DC: National Academy Press. books.nap.edu/catalog.php?record_id=11893.

Rasmussen, J. 1986. *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering*. New York, NY: Elsevier Science.

Reder, L. M., and F. E. Ritter. 1992. "What Determines Initial Feeling of Knowing? Familiarity with Question Terms, not with the Answer." *Journal of Experimental Psychology: Learning, Memory & Cognition* 18 (3): 435–451. doi:10.1037/0278-7393.18.3.435.

Rohrer, D., and K. Taylor. 2006. "The Effects of Overlearning and Distributed Practise on the Retention of Mathematics Knowledge." *Applied Cognitive Psychology* 20 (9): 1209–1224. doi:10.1002/acp.1266.

Rumelhart, D. E. 1967. *The Effects of Interpresentation Intervals on Performance in a Continuous Paired-Associate Task (Tech. Rep. No. 16)*. Stanford, CA: Stanford University, Institute for Mathematical Studies in the Social Sciences.

Sauer, J., D. Burkolter, A. Kluge, S. Ritzmann, and K. Schüler. 2008. "The Effects of Heuristic Rule Training on Operator Performance in a Simulated Process Control Environment." *Ergonomics* 51 (7): 953–967. doi:10.1080/00140130801915238.

Schmidt, R. A., and R. A. Bjork. 1992. "New Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training." *Psychological Science* 3 (4): 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x.

Shea, J. B., and R. L. Morgan. 1979. "Contextual Interference Effects on the Acquisition, Retention, and Transfer of a Motor Skill." *Journal of Experimental Psychology: Human Learning and Memory* 5 (2): 179–187. doi:10.1037/0278-7393.5.2.179.

Simon, H. A. 1974. "How Big Is a Chunk? By Combining Data from Several Experiments, a Basic Human Memory Unit Can Be Identified and Measured." *Science* 183 (4124): 482–488. doi:10.1126/science.183.4124.482.

Stefanidis, D., J. R. Korndorffer, S. Markley, R. Sierra, and D. J. Scott. 2006. "Proficiency Maintenance: Impact of Ongoing Simulator Training on Laparoscopic Skill Retention." *Journal of the American College of Surgeons* 202 (4): 599–603. doi:10.1016/j.jamcollsurg.2005.12.018.

VanLehn, K. 1989. *Learning Events in the Acquisition of Three Skills*, Paper presented at the Proceedings of the 11th Annual Conference of the Cognitive Science Society, Hillsdale, NJ.

Vearrier, L. A., J. Langan, A. Shumway-Cook, and M. Woollacott. 2005. "An Intensive Massed Practice Approach to Retraining Balance Post-Stroke." *Gait & Posture* 22 (2): 154–163. doi:10.1016/j.gaitpost.2004.09.001.

Warming, S. S., N. E. Ebbehøj, N. N. Wiese, L. H. Larsen, J. J. Duckert, and H. Tønnesen. 2008. "Little Effect of Transfer Technique Instruction and Physical Fitness Training in Reducing Low Back Pain Among Nurses: A Cluster Randomised Intervention Study." *Ergonomics* 51 (10): 1530–1548. doi:10.1080/00140130802238606.

Young, J. L. 1971. "Reinforcement-Test Intervals in Paired-Associate Learning." *Journal of Mathematical Psychology* 8 (1): 58–81. doi:10.1016/0022-2496(71)90023-X.