The Pennsylvania State University

The Graduate School

College of Engineering

A NOVEL TRAINING PARADIGM FOR KNOWLEDGE AND SKILLS ACQUISITION: HYBRID SCHEDULES LEAD TO BETTER LEARNING FOR SOME BUT NOT ALL TASKS

A Dissertation in

Industrial Engineering

by

Jaehyon Paik

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2011

The dissertation of Jaehyon Paik was reviewed and approved* by the following:

Frank E. Ritter
Professor of Information Sciences and Technology
Dissertation Co-Advisor
Co-Chair of Committee

David A. Nembhard
Associate Professor of Industrial and Manufacturing Engineering
Dissertation Co-Advisor
Co-Chair of Committee

Andris Freivalds
Professor of Industrial and Manufacturing Engineering

Ling Rothrock
Associate Professor of Industrial and Manufacturing Engineering

Darrell Velegol
Professor of Chemical Engineering

Paul Griffin
Professor of Industrial and Manufacturing Engineering
Head of the Department of Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School

ABSTRACT

Studies on effective practice schedules have been generally investigated by comparing the performance of two relative extreme practice schedules, distributed and massed, at a retention test. The results of most of these studies have consistently shown that distributed practice schedules result in higher retention rates than massed practice schedules because of the spacing effect associated with human memory. These studies, however, failed to show either the optimal interval between learning sessions, or to consider the knowledge types used to perform tasks. Furthermore, these studies did not provide any theoretical supports except the spacing effect for predicting performance at the specific time of the schedules. To address these problems, I explored both theoretically and empirically. First, I investigated ACT-R's learning and forgetting theories to help identify an optimal practice schedule. Second, psychological experiments were conducted to validate these theories. In the experiments, four kinds of tasks were tested using four practice schedules (distributed, massed, Hybrid1, and Hybrid2). Finally, models were developed using ACT-R, and then compared with empirical data.

ACT-R's learning theories suggest that hybrid practice schedules (schedules consisting of distributed and massed practice) could produce better performance than an exclusively distributed practice schedule. The results of empirical data indicated a more complex picture. Like previous studies, the results of experiments showed a higher correlation between retention and distributed practice schedules than retention and massed practice schedules. There were, however, no significant differences between distributed and hybrid practice schedules when testing a declarative memory task. Nevertheless, the results also suggested that some hybrid schedules might produce better performance than distributed practice schedules for perceptual-motor skills. Specifically, my results show that the Hybrid1 practice schedule produced greater skill retention

than the distributed practice schedule for the perceptual-motor task, indicating that the spacing of distributed and massed practiced with respect to each other also influences performance.

When comparing the ACT-R models' performance with that of the participants', the models could predict the learning and forgetting trends of the participants in each group for declarative memory tasks; however, there were differences in the correct responses between the models' prediction and the human data. These results indicated that ACT-R could be used to predict the learning and forgetting trends of practice schedules, however, revisions might be necessary to fully map the models' predictions to the participants' specific responses.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I'd like to thank my advisor, Dr Frank Ritter. His continuous supports, guidance and encouragement enable me to finish my research. His enthusiasm in research had challenged and motivated me. Furthermore, he was always willing to help me with my research and my personal problems. The other committee members, Dr. David Nembhard, Dr. Andris Freivalds, Dr. Ling Rothrock, and Dr. Darrell Velegol, gave constructive criticisms and advices in this research. Dr. Jong Kim provided basic idea for this research, and my lab colleagues, Jonathan Morgan, Changkun Zhao, Jeremiah Hiam, and Chris Dancy, gave useful comments to fulfill this study. Finally, I thank to my family for their supports and love.

Chapter 1

Introduction

Training is important for all industries. The military and transportation industries invest enormous time and money to create well-qualified operators. Hospitals and medical schools continually train residents and nurses to not only improve surgery skills but also reduce medical mistakes. The cost of education and training is a large portion of the overall expenditure in many industries. According to the *2009 Industry Report*, U.S. companies that have more than 100 employees spent $52.2 billion on employees' learning and development cost in 2009. Figure 1-1 shows that U.S. companies have invested more than $50 billion in employees' training each year for the years 2003 to 2009.

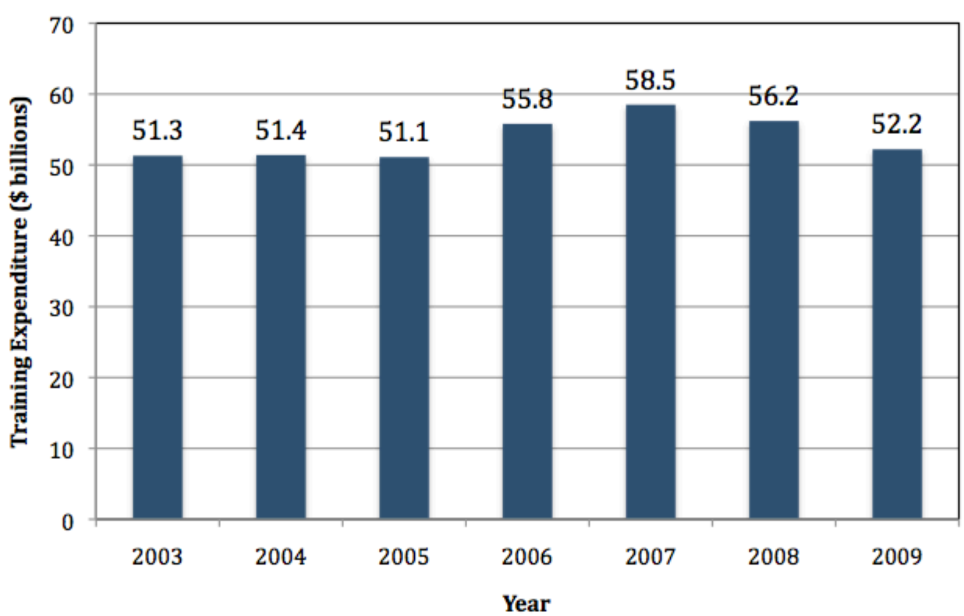


Figure **1-1**: Training expenditure from 2003 to 2009 (Training Magazine, 2009)

Due to the importance of training (illustrated in figure 1.1's training expenditures), scientists, especially in the fields of education and psychology, have studied training strategies. Training strategies for learning declarative and to some extent procedural skills have been researched, e.g., second-language vocabulary (Atkinson, 1972; Bloom & Shuell, 1981), computerized spelling drills test (Fishman, Keller, & Atkinson, 1968), pair-word memory (Pavlik & Anderson, 2005), and mathematical permutation (Rohrer & Taylor, 2006). Generally, this literature has compared the performance of two relatively extreme practice schedules, such as distributed practice schedule and massed practice schedule, at the retention test, and the results of most of these studies consistently indicated that the distributed practice schedule outperforms the massed practice schedule.

Scientists in medicine also have studied training strategies for medical school students. To get better retention on testing for medical school students and surgeons, massed and distributed studies on trainings of surgical skills in laboratories (Moulton et al., 2006), as well as in virtual reality environments (Gallagher et al., 2005; Mackay, Morgan, Datta, Chang, & Darzi, 2002) have been conducted. These complex tasks require not only declarative memory, but also perceptual-motor skill. The results also showed that a distributed practice schedule has better performance than a massed practice schedule on a retention test.

According to the results of the previous studies, the training schedule should be designed to be a distributed schedule regardless of knowledge types. However, I think there may exist schedules that can produce better performance than a distributed practice schedule in retention. In the next section, I describe the problems of current learning, retention, and training studies, and present contributions of my dissertation.

1.1 Problems with Learning and Retention, and How to Train

The results of these previous studies on the practice schedule may have lead many scientists who investigate learning and retention to compare two relatively extreme practice schedules without examining other options. Recent work (Kim, Ritter, & Koubek, in press), however, suggests an alternative view, one that examines the relationship between task types and practice schedules. For perceptual-motor skill, Kim et al. suggest that a massed or somewhat massed way might be more effective. They cited a study that investigated the effects of massed practice on stroke patients (Vearrier, Langan, Shumway-Cook, & Woollacott, 2005). However, the study of Vearrier et al. did not compared a massed practice schedule with a distributed practice schedule, so they do not provide direct evidence for the relative benefits of massed practice schedules, although the results are consistent with a view that massed or semi-massed may be better than distributed.

Kim et al. (in press) also argued that practice should be considered with learning and forgetting frameworks. Both frameworks divided learning and forgetting into three stages, declarative stage, transitional stage, and procedural stage, and if knowledge is procedurlized in one's mind, he/she does not need to retrieve his/her declarative knowledge, and forgetting rarely happens. So, the studies on effective practice schedules should not compare two relatively extreme practice schedules, but be focused on transferring knowledge from the initial stage to the procedural stage.

We can more easily imagine perceptual-motor skill learning if we consider the steps involved in learning how to ski. I think to learn how to ski, a massed practice schedule, (e.g. five hours in a row in one day), might be a better schedule than a distributed practice schedule (one hour per day over five days), and if this is true, we need to consider the possibility of other practice schedules that might produce better performance in retention. It might even the best to spend an hour for three days and two hours on the last day. Thus, we should not conclude that a distributed practice schedule is the best schedule in all kinds of tasks.

Most of the previous studies comparing practice schedules have not examined the gap between training sessions, or the interval between the last training session and retention session. Those studies mainly depended on psychological experiments. Participants were divided into two groups, massed and distributed groups, and followed each schedule, so there was only one kind of gap between training sessions in each schedule, and there was also only one kind of interval between the last session of training and retention. However, recent work (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Pavlik & Anderson, 2005) showed that the performance on a retention test varied according to the degree of spacing between training sessions, and furthermore the optimal gap between the learning sessions depended on the interval between the last learning session and the retention session.

We also need to consider theoretical aspects that most of the previous studies did not support. In the previous studies, participants were divided into two groups without any theoretical supports except spacing effect, and comparing two groups in a retention test. However, we can generate more than two training schedules using several training days. For example, eight hours training could be divided into one hour in a day over eight days, or two hours in a day over four days or any in between. Thus, we need to figure out the performances of the possible schedules by using learning and forgetting theories, and the result may help us to predict the most effective practice schedule.

1.2 Preview of Contributions

The scope of this dissertation encompassed both practical and theoretical contributions including measuring training efficiency, exploring the optimal training strategy based on the knowledge type, and increasing our understanding of the theories of learning and decay process of a promising cognitive architecture. The results of this dissertation could produce a paradigm shift of how to design training programs in academy, industry, and the military with respect to knowledge types of tasks. The detailed contributions are presented below.

The first contribution of this dissertation is to provide a new paradigm of training different from the widely used method in most training. Most of the previous studies have focused on massed practice and distributed practice schedules, but this dissertation shows a better practice schedule could exist, and it may produce better performance in knowledge and skills acquisition and retention. Furthermore, the later researchers who explore training schedules for better retention could consider testing a hybrid practice schedule in their research.

The second contribution of this dissertation is to provide theoretical support for hybrid practice. The results of the base-level learning equation of ACT-R, and its extension, could predict performance of all possible practice schedules, and the results showed a hybrid practice schedule shows 2.5% better performance on retaining knowledge in a retention test than a distributed practice schedule.

The third contribution of this dissertation is to explore the most efficient training schedule for three kinds of knowledge types with four tasks. A declarative memory task, two procedural memory tasks, and a perceptual-motor task were tested with four different practice regimens, and from the results of these experiments, most efficient practice regimen with respect to the specific task and knowledge type are provided.

The fourth contribution of this dissertation is that ACT-R could predict the long-term learning and forgetting process of human memory. The most of the previous ACT-R models have focused on microscopic psychological tasks, such as modeling simple summation processes etc. However, model for learning, forgetting, and retaining aspects of human mind with long-term duration has not been verified in this area, and cognitive modeling could provide meaningful implication for these kinds of practical problems (Gray & Altmann, 2001). Comparing the performance of ACT-R models with empirical data, I can validate whether the ACT-R cognitive architecture could represent learning, forgetting, and retaining knowledge according to the various training schedules with long-term duration or not.

The fifth contribution of this dissertation is to start to examine a new theory of skill retention (Kim et al., in press). According to this theory of skill retention, the forms of forgetting are different in each learning stage, so learning should occurred with different manners and degrees with respect to the learning stages. The hybrid practice schedules that I used in the experiment of this dissertation might be the schedules that could be explored to test the skill retention theory.

1.2 Introduction to the Chapters and Summary

My dissertation consists of seven chapters. The first chapter includes a brief introduction for research topic, problem statements, and expected research contributions. The second chapter is a literature review for learning and forgetting theories, and possible tasks to use in my study. It starts with introducing the learning and forgetting framework, and describes relevant previous research on distributed and massed training schedules. Finally, it ends with reviewing possible tasks that have been used in previous research on learning. In chapter 3, I explore hybrid practice schedules that are supported by the base-level learning equation and revised equations of the ACT-R architecture. Chapter 4 describes the study methodology and its results to explore the topic of my dissertation. In chapter 5, I describe detailed information of the cognitive architecture, ACT-R 6.0. ACT-R models that were developed to represent human behavior for the tasks in the study are also introduced in this chapter. In chapter 6, I compare the results of human data with ACT-R models. Finally, chapter 7 contains conclusions, summaries of the dissertation, and details on its contributions.

Chapter 2

Review of Theories and Tasks

In this chapter, I review the theoretical foundations for knowledge and skill acquisition, and provide a forgetting theory that is recently developed. Reviews for the retention effects according to the training schedules and knowledge types are provided, and finally, the possible tasks that I could use in my thesis for exploring the retention effects according to the different kinds of schedules are presented.

2.1 A Knowledge and Skills Acquisition Framework

Citing Proctor and Dutta's (1995) review of skill acquisition, Kim et al. (2008; in press) provide a good review of the acquisition and degradation of knowledge and skills with respect to declarative and procedural knowledge. Consequently, I will begin with Kim et al.'s review to discuss the learning framework featured in this study before segueing to a discussion of extensions to the framework, namely its application to other knowledge types.

Studies on a framework of knowledge and skills have been conducted several scientists Fitts (1954, 1964) decomposed the knowledge and skills acquisition as three stages: cognitive, associative, and autonomous stages. In the cognitive stage, completion time for a task is slow and errors occur frequently, so task completion time is relatively long in this stage, however, in the autonomous stage, tasks are always proceduralized in this stage, so task completion time is short and errors occur rarely. Anderson (1982) also developed a theory of cognitive skill acquisition with three stages of declarative, transitional, and procedural, corresponding to Fitts' three stage of learning. Rasmussen (1986) also proposed a framework as knowledge-based, rule-based, and skill-based for pertaining to skilled performance.

Recently Ohlsson (2011) showed a framework as getting-started, mastery, and optimization based on the formulation of Fitts. He argued that there are no sharp boundaries between the stages. Regardless of terms that they used, each stage of the framework corresponds to the stage of other framework, and figure 2-1 shows the framework of knowledge and skills acquisition.

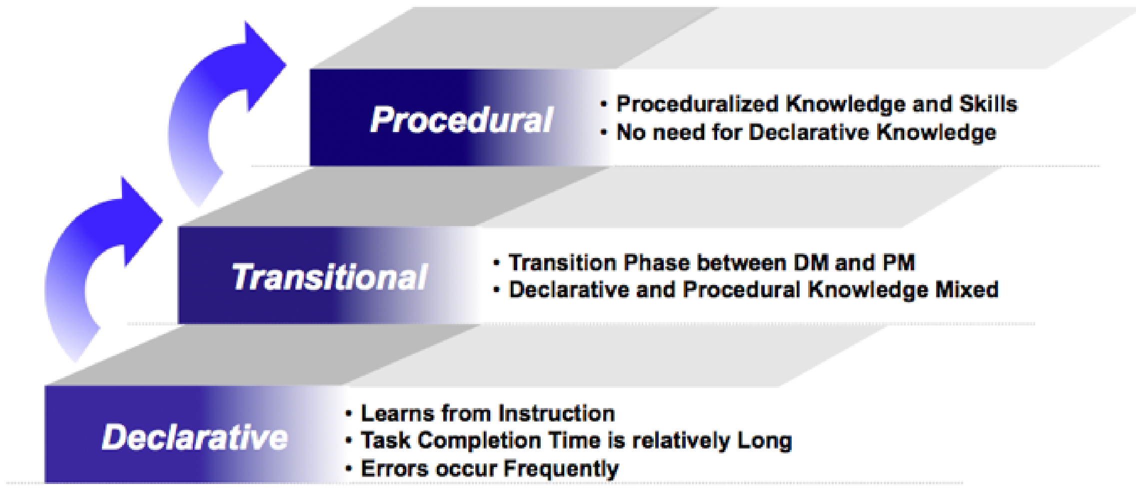


Figure **2-1**: Framework of knowledge and skills acquisition (taken from Anderson, 1982).

For the first stage, the declarative stage, humans learn knowledge and skills from instruction and they are stored as declarative type in human memory. Therefore, humans need to retrieve this declarative type memory, so it takes relatively long time to complete a task and errors occur frequently. In the procedural stage, the knowledge and skills are mostly procedural, so humans do not need to retrieve the information of the task from the declarative memory. The completion time is relatively short and errors occur rarely in this stage. In the transitional stage, humans have knowledge and skills that are partially declarative and partially procedural, and with continuous practice, the knowledge and skills are gradually procedural. This stage is a transitional phase from declarative to procedural phase. Most of the cognitive architectures, such as a Soar and an ACT-R, have this transitional mechanism called knowledge compilation to represent knowledge and skills acquisition, and proceduralization.

Newell and Rosenbloom (1981) proposed an impasse-driven learning mechanism called chunking, which is the foundation learning mechanism in the Soar cognitive architecture. Within Soar, learning takes place with impasses and subgoaling. If there is a lack of sufficient knowledge in a current problem space, Soar generates a subgoal to resolve the impasse. When this impasse is resolved, new procedural knowledge is created as a production. This process is called chunking, and if Soar encounters a similar condition in the future, it can apply this newly learned production to resolve it without any impasse.

Anderson (1982) explained the knowledge compilation with proceduralization and composition in ACT* (Anderson, 1983), a predecessor of ACT-R. The proceduralization is the process of replacing domain-specific declarative knowledge with general knowledge. The composition is sequences of productions are collapsed into a single production. The proceduralization and composition processes convert declaratively encoded knowledge into production form. However, these two mechanisms are not enough to explain the speedup of performance by practice, so generalization and specialization are used to refine procedural knowledge. However, there is a lack of empirical evidence to support these process, and some problems that are caused by newly learned procedural knowledge brings the system to an endless loop, so Taatgen and Anderson (2002) proposed the production compilation that is based on knowledge compilation. It was developed to model complex skill acquisition within the ACT-R architecture. It combines both proceduralization and composition mechanism into a single mechanism. Two rules are combined into a single rule by eliminating the retrieval request in the first rule and the retrieval condition in the second rule. It enables a model to perform a specific task more rapidly, because it reduces a two-steps retrieval into one step, so it can represent the speedup of performance by practice.

2.2 Knowledge and Skills Degradation

The knowledge and skills degradation, forgetting, has been studied, but not much compare to the learning, even if it had been studied, most of the studies focused on to the declarative memory decay or the procedural memory decay based on the extension of the declarative memory decay (Chong, 2004). Thus, there was no whole framework for knowledge and skills degradation.

Recently, researchers (Kim et al., in press) proposed a theory of skill retention with three stages of learning and forgetting based on Anderson's learning theory. They added the three forgetting curves in each learning stage, and they showed the forms of forgetting are different according to the stages. They also argued that knowledge in declarative memory degrades with lack of use, and it leads to the inability to perform the task in the first stage (declarative stage). In the second stage, transitional, they argued with lack of use, the declarative knowledge can be forgotten, leading to miss steps, however, procedural memory is basically immune to decay. Finally, they suggested that task knowledge is available in both declarative and procedural forms, but procedural knowledge predominantly drives performance in the third stage, procedural. With lack of use, declarative knowledge may be degraded, however, learners can still perform the task.

They also argued that different strategies should be applied to a training regime with regard to both the skill type and learning stages. That is, some tasks that require perceptual-motor skill should be trained with a massed way rather than a distributed way, and the training strategies should be different depending on the stage learners are in (i.e. distributed practice in the declarative stage, and distributed with massed practice in the transitional stage). Figure 2-2 shows a graph that describes this theory.

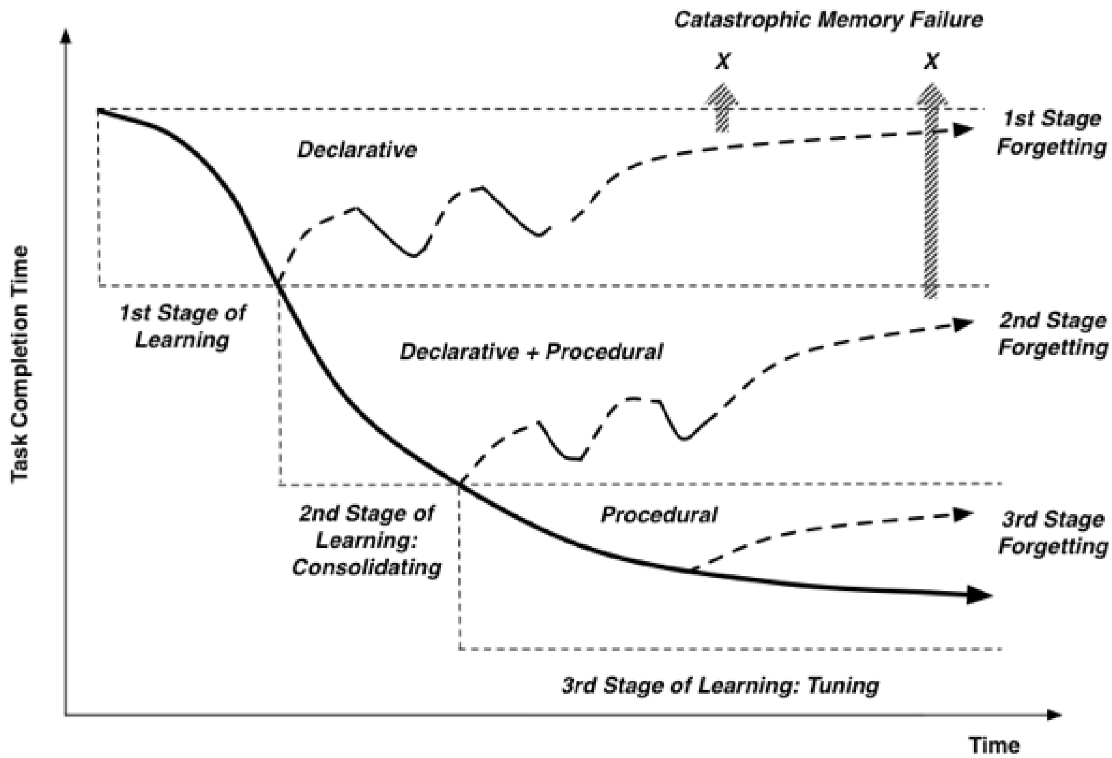


Figure **2-2**: A theory of skill retention with three stages of learning and forgetting (taken from Kim et al., in press).

2.3 Training Strategies for Retaining Knowledge and Skills

In previous sections, I have discussed the stages of knowledge and skills acquisition and degradation. To reach the final stage called the procedural stage, humans need to practice continually. However, humans forget their learned knowledge and skills as times go by because of decay process of human memory. Therefore, how to retain learned knowledge and skills has been an issue in filed of education and training.

Generally, the research on knowledge and skills acquisition has used two kinds of practice schedules; a) massed practice schedule, and b) distributed practice schedule. In a massed practice approach, the interval between training sessions is very short, however, in a distributed practice approach, the space between training sessions is wide. These intervals can be represented by a frequency of practice, therefore, if the frequency of practice is high at the beginning of learning session, this can be considered as massed practice, and if the frequency is low and uniformly distributed, it is thought of as distributed practice (Kim, 2008).

Due to the spacing effects of human mind, it has noticed that the distributed practice approach provides longer retention of learned knowledge and skills regardless of task types. In next three consecutive sections, I review previous studies that were reviewed in Kim at al. (in press) paper, for spacing effects according to different kinds of knowledge types.

2.3.1 Training Strategies for Declarative Tasks

Bloom and Shuell (1981) studied effects of massed and distributed practice on the learning and retention of French vocabulary for high school students. In this study, the distributed practice consists of 10-minute units on each of three successive days; the massed practice consists of all three units being completed during a 30-minute period on a single day. The performance of two groups was almost identical on a test given immediately after completion of study, however, in the second test given 4 days later, the distributed practice group showed better (35%) performance than the massed practice group.

Fishman et al. (1968) showed the same results in their Computerized Spelling Drills test. In this studied, conditions of massed and distributed practice were investigated using a within-subjects design in a situation involving computerized spelling drills. Two sets of three words each were presented once every other day over a period of 6 days in the distributed condition. The six other sets of words were presented one by one during 6 days. The probability of a correct answer for words in the massed condition was higher than that for the distributed condition during the learning sessions, but on 10 days and 20 days later retention tests, the distributed condition showed better performance that the massed one.

Bahrick et al. (1979; 1993) investigated the retention and spacing effect in learning foreign language vocabulary for a long-term period, 9-year longitudinal study. Four subjects learned and relearned 300 English-foreign language word pairs, and they participated in either 13 or 26 relearning sessions with intervals of 14, 28, or 56 days. Then, retention was tested 1, 2, 3, or 5 years after training terminated. The highest retaining performance was the closest spacing interval (14-day) at the last training session. However, after a year since the last training session, the highest retaining performance was the widest spacing of training (56-day), and the lowest retaining performance was found in the closest spacing interval. In the 2, 3, and 5 years retention test, the same pattern of results was shown, and this indicated that the more widely spaced training produced the greater retention performance on this vocabulary-learning task.

Pavlik and Anderson (2003, 2004, 2005) investigated the effect of practice and spacing on retention of Japanese-English vocabulary paired associates. In their studies, they found that the relative benefit of spacing increased with increased practice and with longer retention intervals. From the experiment, they proposed an activation-based model that proposes that a declarative memory unit receives an increment of strength whenever it is activated, but that these increments decay as a power function of time. They found that the decay rate for each presentation depended on the activation at the time of the presentation.

To sum up, the massed practice approach showed better performance than the distributed one at the training day, however, the distributed practice approach shows better performance than the massed approach in retention test for these tasks. According to Bahlick, the more widely spaced training showed the greater performance on this type of task. Pavlik and Anderson proposed an activation-based model that predicts the spacing effect with the activation mechanism of ACT-R cognitive architecture.

2.3.2 Training Strategies for Procedural Tasks

Rohrer and Taylor (2006) investigated the effects of overlearning and distributed practice on the retention of mathematics knowledge. In this study, 216 college students learned to solve one kind of mathematics problem before completing one of various practice schedules. In the first experiment, students were divided into four groups: distributed practice with a 1-week retention interval, distributed practice with a 4-week retention interval, massed practice with a 1-week retention interval, and massed practice with a 4-week retention interval. The difference between massed and distributed practice with 1-week retention interval was almost, 75% and 70%, respectively, however, in the 4-week retention interval, the distributed practice showed better performance than massed practice (64% vs. 32%), suggesting long-term retention of mathematics knowledge is better achieved with largely spaced distributed practice.

In their second experiment, they investigated the effects of overlearning on retention by varying the number of practice problems with a single session (massed practice). The result showed that there are no significant benefits of overlearning on retention of mathematics knowledge for 1-week and 4-week retention intervals. According to these results, Rohrer and Taylor insisted long-term retention was boosted by distributed practice and unaffected by overlearning, so the mathematics textbook should be modified, because most of the mathematics textbooks rely on overlearning and massed practice.

Kim (2008) investigated effects of distributed practice on the retention of procedural motor skills using a novel spreadsheet task, called the Dismal spreadsheet task because it uses the dismal spreadsheet (Ritter & Wood, 2005). The Dismal spreadsheet task is a sequential task consisting of 14 subtasks, and performing all of the subtasks takes about 40 minutes to do the first time, but with practice it takes about 25 minutes. In this study, participants were divided into six groups with two different interface modalities (mouse and keyboard), and three different retention intervals (6-day, 12-day, and 18-day). The results showed that there were no significant differences on learning performance by the two-modality groups. However, with regard to relearning, there is a significant difference of the mean task completion time on 6-day and 12-day retention, and 12-day and 18-day retention the in mouse users group, but there is no significant difference of the mean task completion time on those retentions in the keyboard users group. These indicate that there is statistical evidence that relearning effects can be affected by the modality and by the retention interval.

To sum up, the distributed practice schedule appears to have better performance on knowledge and skills retention tests in procedural memory tasks. According to Rohrer and Taylor, the over-learning or over-training for this type of task did not lead to better performance. The procedural task using motor skills also indicated that relearning effects could be affected by the tool of user and retention interval.

2.3.3 Training Strategies for Perceptual-Motor Tasks

The studies on spacing effects of perceptual (or procedural) motor skills have been mainly conducted in medicine area for finding an efficacy of practice regimen.

Mackay et al. (2002) investigated effects of practice distribution in procedural skills training. In this investigation, 41 novice subjects were recruited, and randomized to three groups (one massed practice group and two distributed practice groups with different rest periods) to receive training on the MIST VR surgical trainer (Wilson, Middlebrook, Sutton, Stone, & McCloy, 1997), a virtual reality trainer for laparoscopic surgery that also assesses performance.

The result demonstrated a benefit for distributed practice over massed practice in learning laparoscopic surgical skills on the MIST VR surgical trainer.

Moulton et al. (2006) investigated approaches for teaching surgical skills. In their study, thirty-eight junior surgical residents, randomly assigned to either massed (1 day) or distributed (weekly) practice regimens, were trained in a new skill called microvascular anastomosis. Each group had the same amount of time (330 min.) in practice, and performance was gathered pre-training, immediately post-training, and 1 month post-training. The anastomotic skill test for a live, anesthetized rat was conducted for final test, and to validate this test, computer-based and expert-based measures were used. The result of this study indicated the distributed group performed significantly better on the retention test in most outcome measures, stating current model of training skills using short courses should be modified.

To sum up, the spacing effect is proved in learning and retention of perceptual (procedural) motor skills task. The result of training on the MIST VR surgical trainer indicates that motor skills learning can be acquired effectively by distributed practice schedule. Furthermore Moulton et al. (2006) insisted that the current curriculum of surgical skills that follows massed form should be modified into distributed form.

In this section, I describe studies on training. Those studies compare two relatively extreme schedules, distributed and massed schedules, and results consistently showed that the distributed schedule is better than the massed schedule regardless of knowledge types.

2.4 Possible Tasks

In this section, I present the candidate tasks that can be used to explore the learning and retention effects of declarative memory, procedural memory, and perceptual-motor skill.

2.4.1 Tasks to test Declarative Memory Learning and Retention

Declarative memory represents the aspect of human memory that stores factual information. For instance, when we are using a web browser to find a specific web site, such as http://www.google.com, we need to know about command "Command + L" or "Ctrl + L" to jump to the URL window for typing the address. This kind of information is stored in declarative memory, and the knowledge of "Command + L" or "Ctrl + L" is considered as declarative knowledge.

To investigate the effects of practice on learning and retention, various kinds of foreign vocabulary memory have been used. Bloom and Shuell (1981) used French, Bahrick et al. (1993) used French and German, and Pavlik and Anderson (2005) used Japanese vocabulary pairs. The foreign language vocabulary have been used by many researchers to find the effects of practice, so this task could be a candidate for exploring the effects of practice on declarative memory learning and retention.

The nonsense syllable learning task that was developed by Ebbinghaus (1964) is a word-like string of letters that is not intended to have any established meaning. This task has been extensively used in experimental psychology for measuring learning and retention. It is similar to the foreign vocabulary memory, because it demands that participants memorize a paired-associate (or stimulus-response) learning. For instance, the subject should response "BIX" when he/she saw a "@" mark (@-BIX pair). The lists of these pairs are practiced during the experiment, and subjects repeated the lists until they were memorized. The experiment shows the learning and retention of this paired-associate test, so the nonsense syllable experiment could be another candidate for exploring the effects of practices on declarative memory learning and retention.

2.4.2 Tasks to test Procedural Memory Learning and Retention

Procedural memory indicates the process of specific task. This information is not easily verbalized, but can be used without consciously thinking about it. For instance, when we try to open web browser, such as IE, Firefox, and Safari, first we need to check the location of the icon, and move mouse to that location. Finally, by double-clicking the icon, we can open the browser.

The Dismal spreadsheet task (Kim, Koubek, & Ritter, 2007) that was designed to study learning and forgetting consists of 14 subtasks, and was used to explore procedural knowledge degradation (Kim, 2008). This task has sequential process to do whole subtasks, and the task completion time could be decreased with practice. Therefore, the Dismal spreadsheet task can be a candidate task for exploring the effects of practice on learning and retention of procedural memory.

The Tower of Hanoi game that is a mathematical game or puzzle has been used in psychological research on problem solving (Anzai & Simon, 1979). It consists of three rods, and a number of disks of different sizes that can be slide on to any rod. The disks are sorted by ascending order in one rod at the first time, and the objective of the game is moving all disks from the leftmost rod to the rightmost rod with following rules: (a) only one disk can be moved at one time, (b) the top disk among the whole stacks can be moved to another rod, and (c) a large disk cannot be placed on a small disk. Participant uses his/her procedural memory to solve this problem, and the completion time and the number of movements can be measured, and might be varied according to the schedules, so this task can be the one of the candidate tasks for learning and retention of procedural memory.

The mathematic problems that were used in Rohrer and Taylor's studies (2006) could be a candidate for a procedural memory task, because it requires procedural memory of human being to solve this particular kind of permutation problems. The problems, such as *aaabbb, aabbbb,* *aabbccc*, etc., were shown to the participants, and participants solved these problems according to the way to solve them. For example, for the sequence of *aabbbb* (with six letter, two occurrence of *a*, and four occurrence of *b*), the solution is like below.

$$aabbbb = \frac{6!}{2! \times 4!} = \frac{6 \times 5 \times 4 \times 3 \times 2 \times 1}{(2 \times 1) \times (4 \times 3 \times 2 \times 1)} = 15$$

Judging alphabetic arithmetic problem (Zbrodoff, 1995) could be used as a procedural memory task. In this study, participants were presented with an equation like A + 2 = C and had to respond yes or no whether the equation was correct based on counting in the alphabet. So, A + 2 = C is correct, but B + 3 = F is not. She designed 2 sets of each of 6 kinds of problem, and each problem has different addend, such as 2, 3, and 4. From the experiment, she found that participants could answer very fast in case of A + 2 = C, however, it took more time when the participants solved the large addend with later alphabet, F + 4 = K.

More interestingly, the participants could memorize the answers of some problems, so they could provide correct answers without using counting. This means that participants used their procedural memory at first, then, with practice, they could use their declarative memory to solve problems.

From the possible tasks that I presented above, the task completion time or the number of correct answers could be gathered, and these data could be used to compare the effectiveness of three kinds of learning regimens.

2.4.3 Tasks for Perceptual-Motor Skill Learning and Retention

Perceptual-motor skill indicates the process of transition from perceiving to movement of motor. When human being perceives something from the sensory input, the mental process selects and controls the movement, and then muscle effectors carry out the movement. For instance, when we meet a confirm massage (ex. Yes or No) from a website, and we are willing to select "Yes", our intention is transferred, so we can move our hand and finger, then click "Yes".

The complex tracking task, also called the Wicken's task (Martin-Emerson & Wickens, 1992) has been used in several psychological experiments to explore human multi-tasking. In this test, participants viewed upper and lower windows on a display screen. Participants performed a tracking task using their right hand joystick in upper window, and pressed the index or middle finger of left hand depending on the direction that an arrow appears in the lower window (right arrow – index finger, and left arrow – middle finger). This task is the one of the candidate tasks for perceptual-motor skill learning and retention.

2.5 Summary

In this chapter, I present the framework of knowledge and skill acquisition that has been insisted by various scientists. I also present a new theory of skill retention with three stages of learning and forgetting based on Anderson's learning framework. All of the theories consistently showed that when humans procedurlize their learned knowledge through continuous practice, forgetting rarely happens, and even if it happens, it could not influence to complete tasks. Through examining these learning, forgetting, and retention theories, I found that the importance of training is not the training schedule itself, but the considering how to transfer learners into the procedural stage of the learning framework. I also show the previous research that compared two kinds of training schedules, a distributed practice schedule and a massed practice schedule, regardless of knowledge types, and concluded that training should occur in a distributed way because spacing effects exist in human memory. Finally, I show some possible tasks that were used in some previous studies, and can be used in my research.

In next chapter, I provide the new practice schedules that I found through examining and exploring in detail the learning and forgetting theories of ACT-R and its extension.

Chapter 3

Theoretical Study of Hybrid Practice

Hybrid practice is a mixed training schedule that includes distributed and massed practice, however, it has not been investigated and examined in previous studies. In this chapter, I present several hybrid schedules that may be able to generate better performance than distributed and massed practice in retention test, and these schedules were found through the equations of ACT-R cognitive architecture and its extension.

3.1 Hybrid Practice Schedules Based on the Baseline ACT-R model

ACT-R uses chunks to present declarative knowledge in human mind. Every chunk has a numeric value that is called activation. The activation reflects the strength of chunk, and this value could increase with more presentation (retrieval request). The activation A_i of a chunk i is computed from two components, base-level and noise component, and the equations are described in equation 3.1 and 3.2.

$$A_i = B_i + \varepsilon \qquad \text{Equation 3.1}$$

The base-level activation for a chunk i is:

$$B_i = \ln\left(\sum_{j=1}^{n} t_j^{-d}\right) \qquad \text{Equation 3.2}$$

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 :bll (base-level learning) parameter

As we know from equation 3.2, the base-level activation values can be increased with high frequency and short interval between the presentations of the object. In contrast, this value could be decreased with low frequency and long interval.

Another factor that decide the possibility of retrieval of chunk is a retrieval threshold. This value (τ) is a constant and can be adjusted by modeler, and the probability of recall (equation 3.3) is a function of chunk's activation value and retrieval threshold. When a model makes a retrieval request and a matching chunk exists, that chunk will be retrieved only if its activation value exceeds the retrieval threshold. The equation 3.3 presented the probability of retrieval of each chunk, *i*.

$$P(A_i) = \frac{1}{1+e^{\frac{\tau - Ai}{s}}} \qquad \text{Equation 3.3}$$

Through the equations 3.1, 3.2, and 3.3, I am able to figure out the recall probabilities at the specific learning and retention sessions. To simplify the activation strength of each chuck, the noise component, ε, in equation 3.1 is ignored, and the retrieval threshold, τ, and the noise parameter, *s*, in equation 3.3 are set as constants (Pavlik & Anderson, 2005).

In my research, the total number of learning session is eight that could take place Monday through Thursday over two weeks, and the retention session takes place 21 days later from the end of the last learning session. Under these conditions (The schedules end on the last day of training, so the schedules never have 0 at the end), I can have 6,435 candidate schedules. Table 3-1 presents the recall probabilities at each training session and retention session, and is sorted by the rank of value at 21st day retention based on these equations.

Table **3-1**: Recall probabilities based on the base-level equation at training and retention days.

Rank	Schedules	1	2	3	4	5	6	7	8	Ret21
1	00000008	0	.9000	.9633	.9791	.9858	.9894	.9917	.9932	.2219
2	00000017	0	.0857	.9161	.9671	.9808	.9868	.9900	.9921	.2210
3	00000107	0	.0447	.9118	.9660	.9803	.9865	.9899	.9919	.2201
4	00000026	0	.9000	.2717	.9286	.9704	.9823	.9876	.9906	.2200
5	00001007	0	.0303	.9098	.9656	.9801	.9864	.9898	.9919	.2192
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
53	00000044	0	.9000	.9632	.9791	.5936	.9465	.9756	.9848	.2180
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
558	00002321	0	.9000	.2717	.9286	.9704	.6437	.9495	.7212	.2088
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
2210	10111013	0	.0447	.1891	.1414	.2383	.3925	.9355	.9722	.1986
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
4397	11111111	0	.0857	.2146	.3286	.2198	.3856	.4867	.5575	.1898
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6435	70000001	0	.9000	.9632	.9791	.9858	.9894	.9916	.3141	.1694

As we can see the table 3-1, the best massed practice schedule (0-0-0-0-0-0-0-8) could generate the best recall probability at the retention day, and the distributed practice schedule (1-1-1-1-1-1-1-1) is by far not the best, its rank is 4,397 among 6,435 (68%). The differences between the massed practice and distributed practice are more than 3.2% in total recall probability at the retention day, and almost 17% in relatively. From this result, we can figure out that the baseline ACT-R model could not predict the spacing effect of human mind, so the needs for extending current learning and forgetting equation have risen.

In next section, I present an extension of current ACT-R equation that was developed by Pavlik and Anderson (2005) for predicting the spacing effect of human mind.

3.2 Hybrid Practice Schedules Based on an Activation-Based model

Pavlik and Anderson (Pavlik, 2007; 2003, 2004, 2005) investigated the effect of practice and spacing on retention of Japanese-English vocabulary paired associates. In their studies, they found that the relative benefit of spacing increased with increased practice and with longer retention intervals. From the experiment, they proposed an activation-based model based on the base-level learning equation of ACT-R 6.0. In this model, they argue a declarative memory unit receives an increment of strength whenever it is activated, but these increments decay as a power function of time. They found that the decay rate for each presentation depended on the activation at the time of the presentation. Equation 3.4 is a decay function that was originally a constant in ACT-R 6.0, and shows how the decay rate, d_i, is calculated for the *i*th presentation of an item as a function of the activation m_{i-1}.

$$d_i(m_{i-1}) = ce^{m_{i-1}} + a$$ Equation 3.4

c: decay scale parameter
a: intercept of the decay function

This equation results in a steady decrease in the long-run retention for more presentations in a sequence where presentations are closely spaced. As spacing gets wider in such as sequence, activation has time to decrease between presentations, decay is therefore lower for new presentations and long-run retention effects do not decrease as much (Pavlik & Anderson, 2008). Using the decay function from equation 3.4, the revised base-level learning equation is presented in equation 3.5.

$$m_n(t_{1\ldots n}) = \ln\left(\sum_{i=1}^{n} t_i^{-d_i}\right)$$ Equation 3.5

From the equation 3.5, an activation-based model, I can figure out the recall probabilities at the training day and retention day. The conditions are the same as the previous section, that is 6,435 training schedules, and the decay scale parameter, *c*, is 0.217 and the intercept of the decay function, *a*, is 0.177. Those values have used by Pavlik and Anderson (2005, 2008). The noise component, ε, in equation 3.1 is also ignored, and the retrieval threshold, τ, and the noise parameter, *s*, in equation 3.3 are set as constants. Table 3-2 shows the recall probabilities at each training session and retention session, and is sorted by the rank of value at 21st day retention.

Table **3-2**: Recall probabilities based on the activation-based model at training and retention days.

Rank	Schedules	1	2	3	4	5	6	7	8	Ret21
1	10111013	0	.0452	.0744	.1807	.4203	.6531	.8093	.8688	.5073
2	10021013	0	.0345	.0913	.1867	.4191	.6510	.8084	.8684	.5071
3	10111103	0	.0452	.0744	.1807	.4920	.6158	.7873	.8598	.5055
4	10021103	0	.0345	.0913	.1867	.4903	.6137	.7864	.8594	.5053
5	11011013	0	.0712	.0630	.1653	.4097	.6490	.8080	.8682	.5049
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
754	11111111	0	.0712	.0919	.3364	.3831	.6207	.7316	.7943	.4819
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
5532	00002321	0	.1099	.4414	.7070	.7323	.7323	.8194	.8190	.4068
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
5738	00000044	0	.1234	.1764	.4944	.6049	.7626	.8368	.8721	.3982
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6103	00000008	0	.1234	.1764	.4944	.7150	.8070	.8503	.8757	.3785
•••	•••	•••	•••	•••	•••	•••	•••	•••	•••	•••
6435	70000001	0	.1234	.1764	.4944	.7150	.8070	.8503	.4842	.3083

As we can see table 3-2, the hybrid practice schedule (1-0-1-1-1-0-1-3) could generate the best recall probability at the retention day, and the distributed practice schedule (1-1-1-1-1-1-1-1) is not the best, its rank is 754 among 6,435 (11.7%). The differences between the best hybrid practice and distributed practice are more than 2.5% in total recall probability at the retention day, and 5.2% in relatively. However, the distributed schedule is better than the massed practice schedule (0-0-0-0-0-0-0-8) by almost 11%. Thus, the distributed practice schedule could produce better performance than the massed one at retention that have been shown in many previous studies, however, the distributed with cramming is the best.

Pavlik and Anderson also predicted the results of the previous canonical studies through their model like as Raaijmakers (2003) did. Raaijmakers extended the search of associateive memory (SAM) model (Raaijmakers & Shiffrin, 1981) to account for the spacing effect and successfully fit the model to the data sets of Rumelhart (1967), young (1971), and Glenberg (1976). In addition to these data sets, Pavlik and Anderson also predicted the data sets of Bahrick (1979) and Bahrick and Phelps (1987). To fit these data sets, Pavlik and Anderson used different values for the decay intercept (a) and decay scale (c) parameters, and table 3-3 shows these different values with respect to the previous studies.

Table **3-3**: Decay intercepts and Decay scales with respect to the previous studies.

	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

I also investigated practice schedules with these different decay parameters. For the results of the decay parameters of Bahrick, Bahrick and Phelps, and Glenberg showed that the best practice schedule is 1-0-1-1-1-0-1-3 like as previous table 3-2 showed, however the orders of rank are little bit different among the results. For the results of the decay parameters of Rumelhart and Young showed that 1-0-0-2-1-0-1-3 is the best training schedule, and 1-0-1-1-1-0-1-3 ranks at the 12th and 3rd in each result respectively. Both rank higher than the fully distributed practice schedule. From these results, we can assume that ACT-R theories support some hybrid practice schedules could produce better performance than a distributed practice schedule. However, this equation could not explain the relationship between retention interval (RI) and inter-study interval (ISI) that were found in recent experiment (Cepeda et al., 2009; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Cepeda et al., 2008).

3.3 A Study on Optimal Study-Interval and Retention

Several studies on predicting optimal spacing and retention for knowledge acquisition (Cepeda et al., 2009) have been explored to examine the relationship between retention interval and inter-study interval. To figure out this relationship, they use 32 obscure but true trivia facts (e.g., “What European nation consumes the most spicy Mexican food?” Answer: “Norway) with two study sessions and one retention session. They used seven kinds of inter-study interval (interval between the first and second study sessions) and four kinds of retention interval (interval between the second study session and the final test). Table 3-3 shows four kinds of retention interval, seven kinds of the inter-study interval, and the number of subjects in each experimental condition.

Table **3-3**: Number of subjects in each experimental condition.

Retention Interval (days)	Inter-Study Interval (days)	Number of subjects
7	0	60
7	1	66
7	2	79
7	7	77
7	21	70
7	105	45
35	0	72
35	1	69
35	4	75
35	7	66
35	11	41
35	21	61
35	105	23
70	0	55
70	1	67
70	7	59
70	14	51
70	21	49
70	105	37
350	0	45
350	1	34
350	7	43
350	21	25
350	35	41
350	70	26
350	105	28

Under these conditions, subjects participated two study sessions and one test session by following their schedules. The final test session consists of two tests, recall (subjective test) and recognition (objective test with five potential answers) tests. The results that are presented in figure 3-1 shows that the final performance increased with increasing inter-study interval initially and then decreased as inter-study interval was increased further. This shows that the spacing effect exists in humans, however, the more widely spacing could not generate better performance at retention test. It also showed that there might be a threshold with respect to each retention interval; so, we need to consider these thresholds in making training schedules.

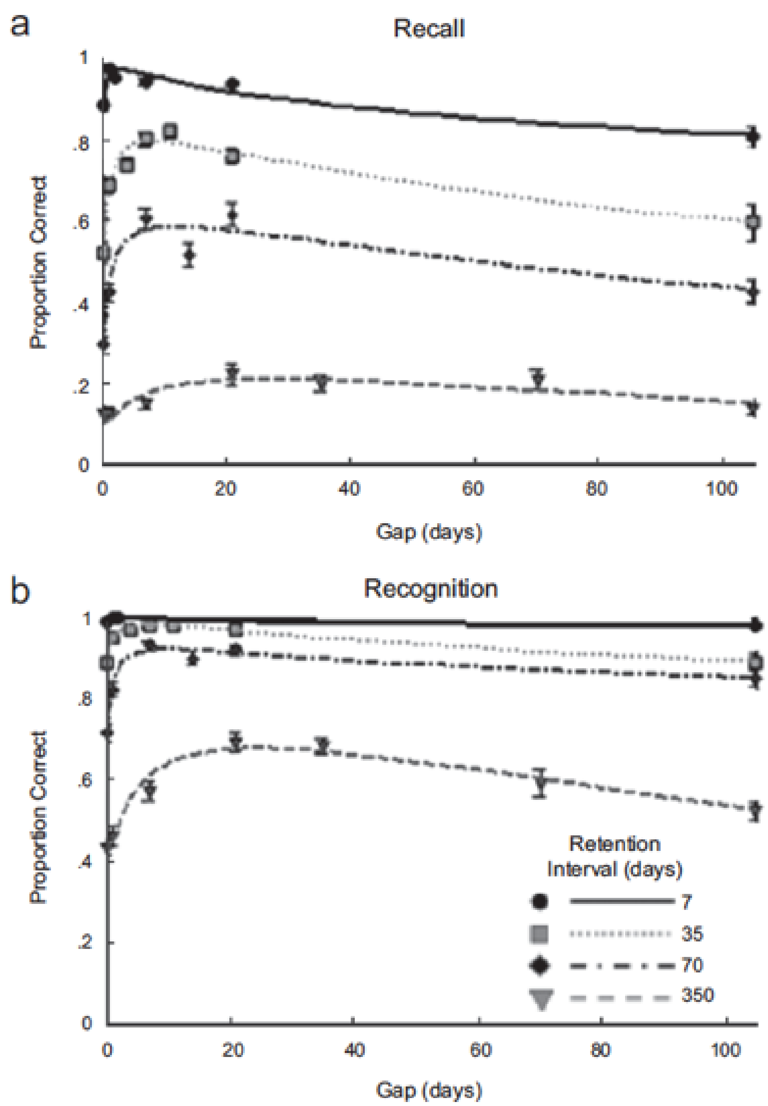


Figure **3-1**: Performance on the final (a) recall and (b) recognition tests as a function of gap, for each of the four retention intervals (taken from Cepeda et al., 2008).

3.4 Other Theories for Performance Prediction

Lebiere and Best (2009) recently argued that many cognitive architectures have difficulties to predict the ability to imperfectly but robustly enumerate a set of alternatives that could be found in many human activities. They argued that the conflict between mechanisms of long-term reinforcement and the need for short-term inhibition of recent items is a primary source of the difficulties, so they revised the original base level learning equation to solve this problem. They added two new parameters: a short-term decay rate d_s, and a time scaling parameter t_s, and the equation is presented in equation 3.6.

$$B_i = \log \sum_{j=1}^{n} t_j^{-d} - \log\left(1 + \frac{t_n^{-ds}}{t_s}\right) \quad \text{Equation 3.6}$$

Using the equation 3.6, they tested different ranges of the short-term decay rate and time scaling parameter. They found that this new mechanism not only prevents degenerate behavior in memory retrieval, but also emerges as a source of the power law distribution. However, it still needs to be validated with empirical data.

The Predict Performance Equation (PPE, Jastrzembski, Addis, Krusmark, Gluck, & Rodgers, 2010; Jastrzembski & Gluck, 2009; Jastrzembski, Gluck, & Gunzelmann, 2006) is a mathematical model of learning and forgetting developed to capture performance across training histories, and to generate precise, quantitative point predictions of performance. The PPE is implemented in the Predictive Performance Optimizer (PPO, Jastrzembski, Gluck, Rodgers, & Krusmark, 2009) —a cognitive tool designed to help learners and instructors predict optimal spacing between the training sessions. The PPE equation can be presented in equation 3.7 and 3.8.

$$Performance = S \bullet St \bullet N^{C} \bullet T^{-d} \quad \text{Equation 3.7}$$

S: a scalar to accommodate any variable of interest
C: learning rate

d: decay rate

$$St = \left[\frac{\sum lag}{P} \bullet \frac{P_i}{T_i} \bullet \frac{\sum_i^j (lag_{\max_{i,j}} - lag_{\min_{i,j}})}{N_i} \right] \qquad \text{Equation 3.8}$$

lag: amount of true time passed between training events
P: true amount of time amassed in practice

Jastrzembski et al. could validate these equations and their extension with a team coordination Unmanned Air Systems (UAS) reconnaissance task and F-16 simulator air-to-air combat data, and they found that the equations could predict the performance of their empirical data closely. However, I remain to explore these equations as future work.

3.5 Summary

In this chapter, I show the theoretical study of hybrid practice. I use the base-level learning equation, and the revised base-level learning equation that are the theories of ACT-R for the declarative knowledge learning and forgetting, and I provide the best practice schedule that is supported by the ACT-R cognitive architecture. The results of the base-level learning equation predict that the massed practice schedule (0-0-0-0-0-0-0-8) is the best, however, the results of the revised base-level learning equation predict that the hybrid practice schedule (1-0-1-1-1-0-1-3) produce better performance than the other schedules. However, Cepeda et al. (2008) validated the revised base-level learning equation with their empirical data and found that the equation could not predict the performance of various learning interval and long term retention, so this equation might be needed to modified in some way.

Lebiere and Best (2009) revised the base-level learning equation, and found that their new equation not only could prevent degenerate behavior in memory retrieval, but also emerge as a source of the power law distribution. However, it still needs to be validated with empirical data.

Jastrzembski et al. (2010) showed the Predictive Performance Equation, and by using this equation they provided how to predict the performance of learner at any specific time.

In next chapter, I present the procedure and results of the empirical study to test these theories.

Chapter 4

Empirical Study: Exploring the Best Practice Schedule

This chapter describes the detailed research design and method to explore a new paradigm of training strategy for knowledge and skills acquisition and retention. Participants conducted four sets of tasks with four kinds of training regimens. Dependent measures and results of the experiment are presented here.

4.1 Method

4.1.1 Participants

Forty-six undergraduate students at the Pennsylvania State University were recruited in this experiment. Four participants did not show up at the training day, and two participants' data were excluded because they studied during their rest periods. For the reimbursement, participants were provided extra course credit or $27. The participants did not have knowledge of Japanese vocabulary and Tower of Hanoi game.

The required sample size, 40, were calculated and approximated from the power analysis for ANOVA design (Cohen, 1988). In this task, I used 0.8 as an effect size delta value, because its range is 0 to 3 typically, and 0.75 is medium effect size, so 0.8 is greater than the medium effect size. The sample size, delta values and power parameters specified in the analysis are presented in Table 4-1.

Table **4-1**: Power analysis for the experiment design.

N	Delta
2	0.067
3	0.089
4	0.111
5	0.135
6	0.159
7	0.185
8	0.211
9	0.237
10	0.264
12	0.318
14	0.371
16	0.424
18	0.475
20	0.524
25	0.634
30	0.726
35	0.799
40	***0.855***
50	0.929

Significant Level (Alpha) is 0.05

4.1.2 Materials

To explore the objectives of this study, four kinds of experiment environments were developed. Each of the experiment environments satisfies the requirements of possible tasks that I mentioned in chapter 2.

The web-based Japanese-English vocabulary test is based on Pavlik and Anderson's (2005) Japanese-English vocabulary test. It was developed for declarative knowledge learning and retention. Pavlik and Anderson used 104 Japanese-English words that were from the Medical Research Council (MRC) Psycholinguistic database (Coltheart, 1981). Among these 104 Japanese-English word pairs, fifteen Japanese-English word pairs were used. These words have familiarity ratings between in English 406 and 621, with a mean of 565, and have imagability ratings between 343 and 566, with a mean of 480. These ratings are composed according to a procedure described in the MRC Psycholinguistic Database manual (Coltheart, 1981). The overall MRC database means for familiarity and imagability are 488 (SD 120) and 438 (SD 99) respectively, so the words had higher familiarity and imagability ratings than the database averages. Only four-letter English words were used (e.g. base, mail, date, etc.), and four-to-seven letter Japanese translations were used (e.g. dodai, yuubin, nendai, etc). Japanese words were presented using English characters. Word assignment to conditions was randomized for each participant and each trial. The problems were displayed one problem per one page, and the answer was displayed with "Correct" or "Wrong" sign. The number of correct answers and the task completion time of correct answer were recorded. Figure 4-1 is the screen shot of the web-based Japanese-English vocabulary task that I used.

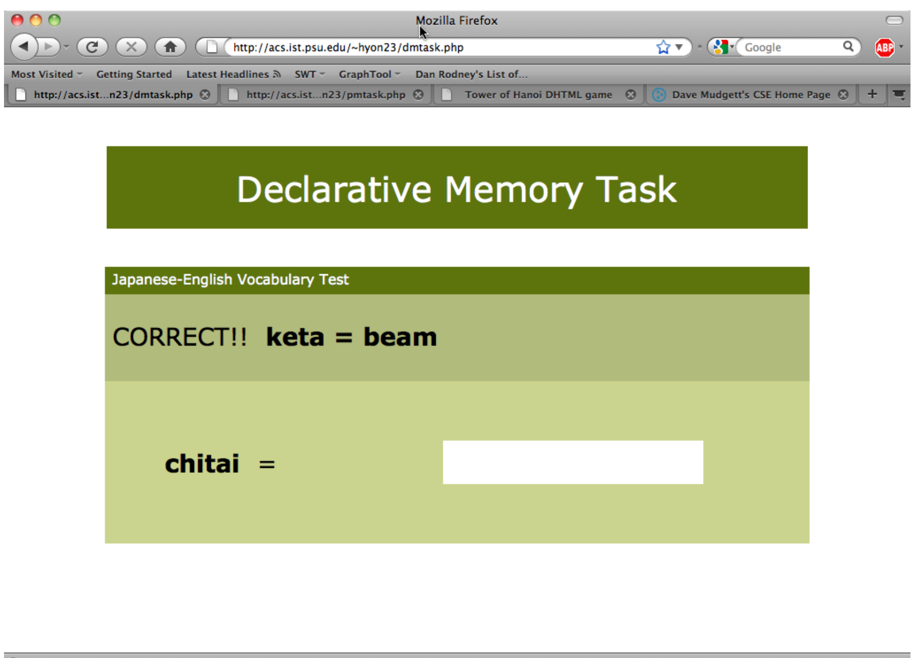


Figure **4-1**: The web-based Japanese-English vocabulary task.

For procedural knowledge learning and retention, two tasks were used, one is a web-based Tower of Hanoi puzzle, and the other is a web-based mathematical permutation problem-solving task. A Tower of Hanoi puzzle that was modified from its original style has three rods with six disks. Participants were asked to move six disks from the leftmost rod to the rightmost rod with three rules: (a) only one disk can be moved at one time, (b) the top disk among the whole stacks can be moved to another rod, and (c) a large disk can not be placed on a small disk. The lowest number of movements, 63 moves, was displayed for motivational purposes, and a user's moves were counted and displayed. Figure 4-2 is the screenshot of the web-based Tower of Hanoi puzzle.

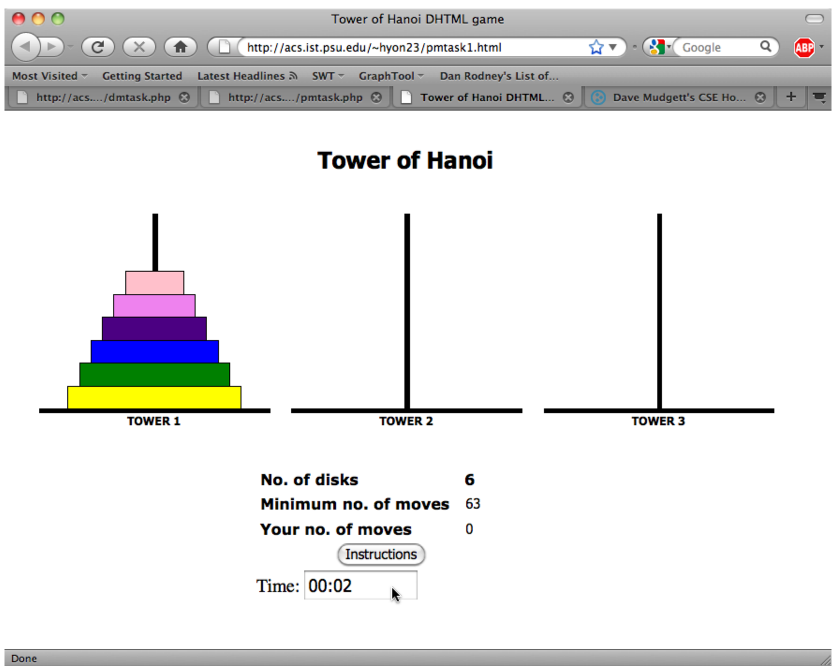


Figure **4-2**: The web-based Tower of Hanoi game.

A web-based mathematical permutation problem-solving task that was used by Rohrer and Taylor (2006) is solving a series of particular kind of problem. Participants were asked to provide the number of possible unique ordering of different sequences. A sequence is a set of n letters including k different letters and at least one repeated letter (e.g. *aabbbb*, n =6, k = 2). The number of possible unique ordering is equal to $n!/n_1!n_1!...n_k!$, where n_i is the number of repetitions of letter i. This task consists of 12 problems that have particular form, such as *aaabbb*, *aabbccc*, etc, and these problems were displayed one-by-one per page, and the answer also was displayed with a “Correct” or “Wrong” sign. Participants were asked to solve all problems in each session, and the number of correct answers and the task completion time of each correct answer were presented when the task was completed. To calculate the problems, blank paper and pencil were provided. Figure 4-3 is the screenshot of the web-based mathematical permutation problem-solving task.

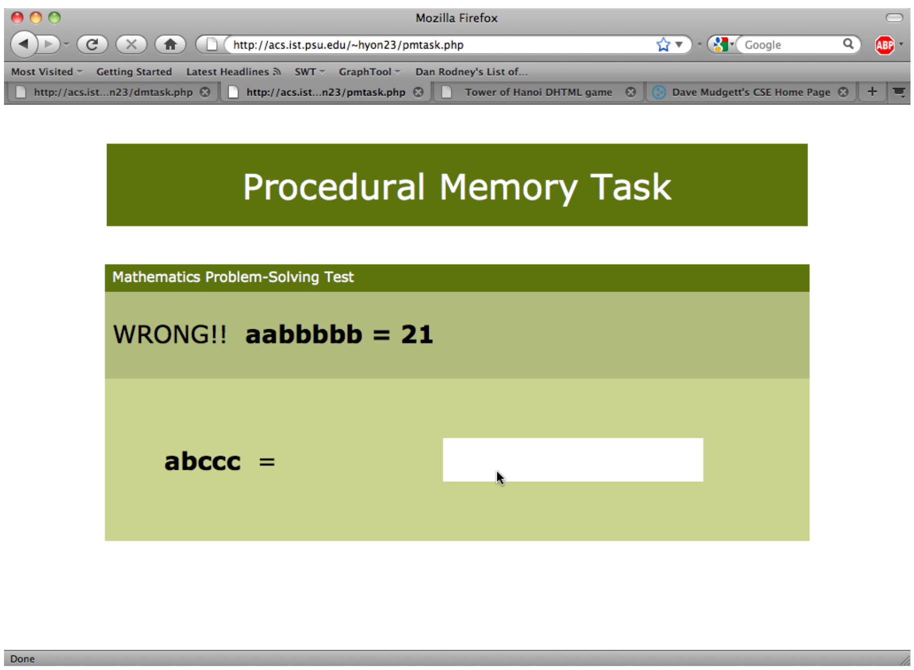


Figure **4-3**: The web-based mathematical permutation problem-solving task.

For the perceptual-motor learning and retention task, an Inverted Pendulum task was used. This is an iPhone or iPodtouch application that is controlled by accelerometer. Participants used an iPodtouch, and they needed to keep balancing a stick that is on the ball in the game through tilting the device. The time duration of balancing is provided by the application itself. Figure 4-4 shows the screen shot of this application.

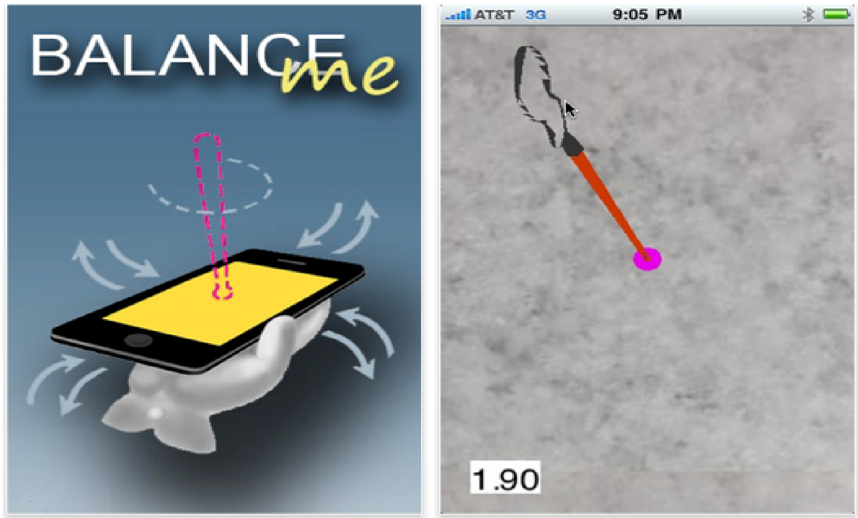


Figure **4-4**: Screenshots of the iPodTouch based Inverted Pendulum task.

These experiments were held at the ACS (Applied Cognitive Science) Usability lab. A Mac desktop with wide screen monitor, keyboard and mouse were provided to conduct experiments.

From the Japanese-English vocabulary test, the number of correct answers (accuracy) and the time taken for the correct answers (latency) were recorded in the system. The number of disk movements and the task completion time were recorded from the Tower of Hanoi puzzle. Like the Japanese-English vocabulary test, the number of correct answers (accuracy) and time taken for the correct answers (latency) were recorded in the system from the mathematical permutation problem-solving task. Finally, duration time was obtained for the inverted pendulum game. Table 4-2 shows the dependent measures for each knowledge type in this study.

Table **4-2**: The dependent measures according to the task type.

Task Type	Task	Dependent Measure
Declarative Memory	Japanese-English Vocabulary	Number of Correct Answers
		Task Completion Time
Procedural Memory	Tower of Hanoi	Number of Disk Moves
		Task Completion Time
	Mathematical Permutation Problem	Number of Correct Answers
		Task Completion Time
Perceptual Motor	Inverted Pendulum Task	Time duration of Balancing

4.1.3 Design

The experiment of this study is a between subjects experiment with one independent variable of four levels (massed, distributed, Hybrid1, and Hybrid2 schedules). The participants were randomly assigned to four groups with regard to training regimens. All of the regimens consist of eight learning sessions and one retention test session, and each session has four tests including the Japanese-English vocabulary test, the Tower of Hanoi puzzle, the Permutation problem-solving test, and the inverted pendulum. Therefore, all of the participants performed 36 tests on a schedule according to their training regimen.

The schedule of the distributed practice group was eight learning sessions in eight days over two weeks (four sessions per week). The schedule of the massed practice group was the same eight learning sessions in two days over one week (four sessions at the 1st day and another four session at the 2nd day). The Hybrid1 practice group performed eight learning sessions in four days over one week with unevenly distributed (two sessions at the 1st day, three sessions at the 2nd day, two sessions at the 3rd day, and one session at the 4th day). The Hybrid2 practice group performed eight learning sessions in six days over two weeks (one session at the 1st, 3rd, 4th, 5th, and 7th day, and three session at the 8th day). The retention test of the all regiments took place three weeks later (21 days retention) after the last session of each regimen. To obtain more accurate results, participants were asked not to do mental rehearsal or practice of the tasks during the rest period. Table 4-3 shows the four different training regimens for this study.

Table **4-3**: Four different training schedules for the learning and retention experiments.

		Mon.	Tue.	Wed.	Thu.	Fri.
1st Week	Distributed	D1	D2	D3	D4	
	Hybrid1					
	Hybrid2	$H_2$1		$H_2$2	$H_2$3	
	Massed					
2nd Week	Distributed	D5	D6	D7	D8	
	Hybrid1	$H_1$1$H_1$2	$H_1$3$H_1$4$H_1$5	$H_1$6 $H_1$7	$H_1$8	
	Hybrid2	$H_2$4		$H_2$5	$H_2$6$H_2$7$H_2$8	
	Massed			M1M2M3M4	M5M6M7M8	
3rd Week	Distributed					
	Hybrid1					
	Hybrid2					
	Massed					
4th Week	Distributed					
	Hybrid1					
	Hybrid2					
	Massed		M – Ret. Test			
5th Week	Distributed					
	Hybrid1				H_1 – Ret. Test	
	Hybrid2				H_2 – Ret. Test	
	Massed				D – Ret. Test	

4.1.4 Procedure

Before starting the experiments, participants were asked to read an implied informed consent form, and then they had explanation of the purpose of the study and performed the four tasks. The sequence of the four tasks was randomly assigned, and whenever participants complete one task, dependent measures, such as completion task time, number of movements, and so on, were presented automatically to the participants.

4.2 Results and Discussion

I present the experiment outcomes from the 40 participants who completed the tasks according to the practice schedules in this section. The performance test for the first learning session (before test) is shown in 4.2.1. In section 4.2.2, the general performance of each group with respect to the four tasks is presented. The comparing of performance improvement between the first learning session and the retention session is presented in 4.2.3. Finally, the learning trends for each group with respect to all tasks are presented in 4.2.4.

4.2.1 Testing the Random Assignment of Groups

In this section, I present the performance test on the first learning session of all practice schedules with respect to the whole tasks to validate the participants were randomly assigned to each group. A series of one-way analysis of variance (ANOVA) were conducted.

4.2.1.1 The Japanese Vocabulary Test

The accuracy and the latency of the Japanese vocabulary test are all zero at the first learning session. These results show that the participants in each group did not have any knowledge about Japanese words, and it reflects that there are no significant differences for recruited participants in each group.

4.2.1.2 The Permutation Problem-solving Task

Two kinds of measurement, accuracy and latency, were analyzed to compare the practice schedules for the Permutation problem-solving task. There were little differences for accuracy and latency at the first learning session, so I performed a one-way analysis of variance (ANOVA). I assumed that all of the variables are normally distributed and have equal variance. Table 4-4 shows the results of ANOVA for the accuracy and latency for this task.

Table **4-4**: ANOVA table for the accuracy and the latency of the Permutation problem-solving task.

		df	*Sum of Squares*	*Mean Square*	*F*	*P*
Accuracy	Between Groups	3	0.06	0.02	1.07	.38
	Within Groups	36	0.70	0.02		
	Total	39	0.72			
Latency	Between Groups	3	1,372.72	457.57	2.21	.10
	Within Groups	36	7,440.51	206.68		
	Total	39	8,813.23			

There were no significant differences between groups for the accuracy of the Permutation problem-solving task at the first learning session, $F(3,36) = 1.07$, $p > .05$, and the latency of this task at the first learning session, $F(3,36) = 2.21$, $p > .05$. These results indicate that the participants were randomly assigned into each group, so the participants of each group did not lead to significant differences for performing the task during the learning and retention sessions.

4.2.1.3 The Tower of Hanoi Puzzle

For the Tower of Hanoi puzzle, two kinds of measurement, task completion time and the number of disk movements, were analyzed to compare the practice schedules. A one-way analysis of variance (ANOVA) was conducted, and the results are presented in table 4-5.

Table **4-5**: ANOVA table for the task completion time and the number of disk movements of the Tower of Hanoi puzzle.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Task Completion Time	Between Groups	3	302,438.64	100,812.88	0.92	.44
	Within Groups	36	3,928,955.65	109,137.66		
	Total	39	4,231,394.29			
Number of Disk Movements	Between Groups	3	48,924.20	16,308.07	0.81	.49
	Within Groups	36	720,937.40	20,026.04		
	Total	39	769,861.60			

The task completion time of the Tower of Hanoi puzzle at the first learning session is not significant between groups, $F(3,36) = 0.92$, $p > .05$, and the number of disk movements at the first learning session is not significant between groups either, $F(3,36) = 0.81$, $p > .05$. From the results of ANOVA, the participants were randomly divided into each group, so the participants of each group did not lead to significant difference for performing the task during the learning and retention sessions.

4.2.1.4 The Inverted Pendulum Task

For the Inverted Pendulum task, the duration time was analyzed to compare the practice schedules. A one-way analysis of variance (ANOVA) was conducted, and table 4-6 shows the results.

Table **4-6**: ANOVA table for the duration time of the Inverted Pendulum task.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Duration Time	Between Groups	3	363.53	121.17	0.81	.49
	Within Groups	36	5,392.96	149.80		
	Total	39	5,756.49			

The task duration time of the Inverted Pendulum task at the first learning session is not significant different between groups, $F(3,36) = 0.81, p > .05$. These results indicate that the participants in each group were randomly divided into each group, so participants of each group did not influence to the performance during the whole learning and retention sessions.

4.2.2 General Performance for Each Group

All of the participants conducted eight learning sessions and one-retention session. One session consists of four tasks, and the order of the tasks was decided randomly in each session. The statistical results are presented in the next section.

Figure 4-5 shows the experiment results of the accuracy for the Japanese-English vocabulary task in each session with respect to the four practice schedules. As we can see, the proportion correct at the first learning session is zero (nobody knew this Japanese vocabulary), and it increases with practice. However, the increase and decrease rates are different by each session and each practice schedule. The massed practice schedule shows that the accuracy increases during the first four learning sessions, and there is almost no change between the fourth and fifth sessions. The reason is there was one-day rest period between those two sessions, so a little knowledge degradation might happen during the rest period. The distributed schedule shows that the accuracy increases during the first four learning sessions, and it decreases at the fifth session. There were four days rest period between the fourth and fifth sessions, so some amount of knowledge degradation happened during the rest period. The Hybrid1 practice schedule shows that the accuracy increases during the first two learning sessions, however, the large amount of knowledge degradation happened at the third learning session. It indicates that two concatenation learning sessions are not enough to retain learned vocabularies. The Hybrid2 practice schedule shows that the accuracy increases during whole learning sessions, even though there are four days rest period between the third and fourth learning sessions.

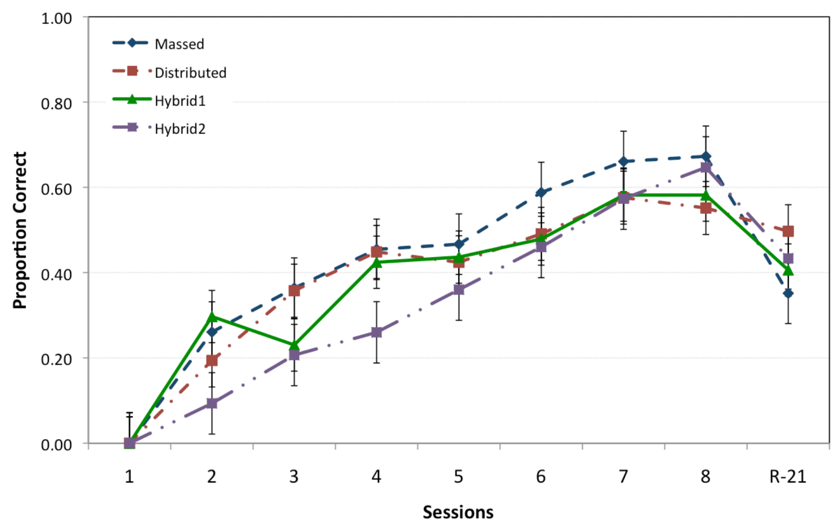


Figure **4-5**: The accuracy for the Japanese-English vocabulary test with respect to the four practice schedules.

The average proportion correct for the massed practice schedule at the last learning session is .67, which is the highest, and for the distributed schedule shows the lowest performance, .56. However, the massed practice schedule has the lowest performance, .35, and the distributed practice schedule has the highest performance, .51, at the retention session (R-21). The other hybrid schedules, Hybrid1, and Hybrid2, have .41, and .43, respectively. The massed practice schedule shows the highest decrease (.67 to .35) among the schedules, and the distributed practice schedule shows the lowest decrease (.55 to .50). The Hybrid1 practice schedule decreases from .58 to .41, and the Hybrid2 practice schedule decreases from .65 to .43. These results are presented in table 4-7.

Table **4-7**: The accuracy for the Japanese-English vocabulary test at the last learning session and the retention session with respect to the four practice schedules.

	Last Learning Session	Retention Session (R-21)
Massed	.67	.35
Distributed	.56	.51
Hybrid1	.58	.41
Hybrid2	.65	.43

Figure 4-6 shows the experiment results of the latency for correct answers of the Japanese-English vocabulary task in each session with respect to the four schedules. The latency at the first learning session is zero, because the number of correct answers is zero at that session. The latencies among the schedules are almost the same at the last learning session, however, the massed practice schedule shows the highest, 4.78 sec., and two of the hybrid practice schedules show similar latencies, 3.41 sec. and 3.51 sec., at the retention session. These results are presented in table 4-8.

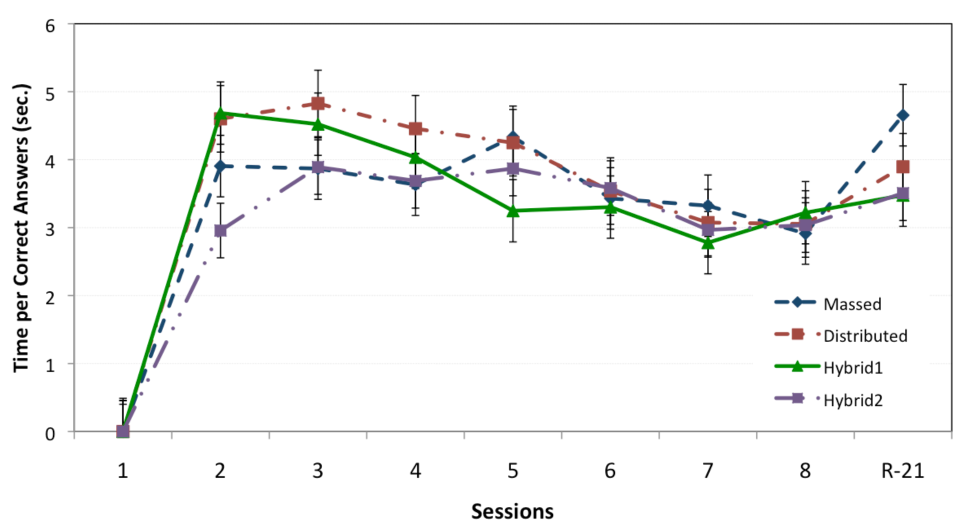


Figure **4-6**: The latency for the Japanese-English vocabulary test with respect to the four practice schedules.

Table **4-8**: The latency (sec.) for the Japanese-English vocabulary test at the last learning session and the retention session with respect to the four practice schedules.

	Last Learning Session	Retention Session (R-21)
Massed	2.99	4.78
Distributed	3.09	4.00
Hybrid1	3.17	3.41
Hybrid2	3.04	3.51

The accuracy and latency were also used in the Permutation problem-solving task. Figure 4-7 and 4-8 show the experiment results with respect to the four practice schedules. From the both figures, we can figure out that different kinds of practice schedules do not affect the accuracy and latency in this kind of task. The latency, of course, decreases with practice, but there is no difference among the practice schedules. These results are presented in table 4-9 and 4-10.

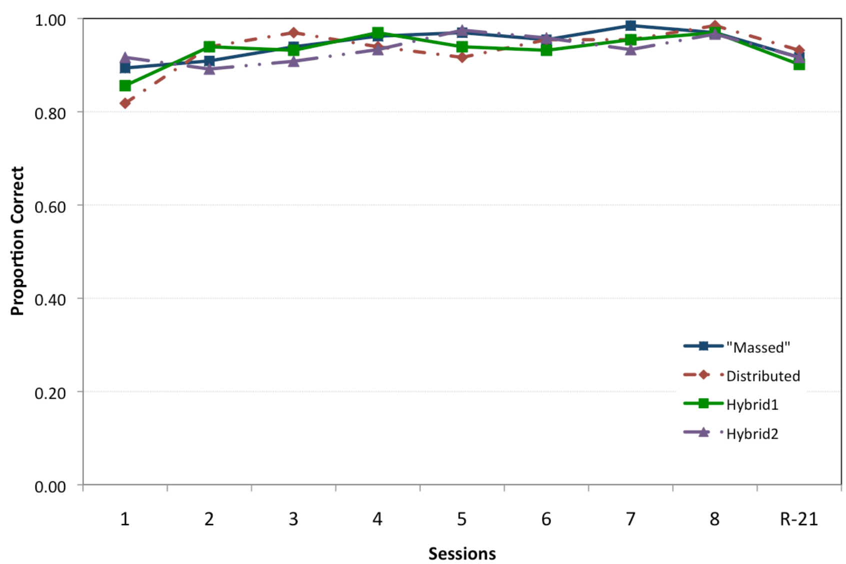


Figure **4-7**: The accuracy for the Permutation problem-solving task with respect to the four practice schedules.

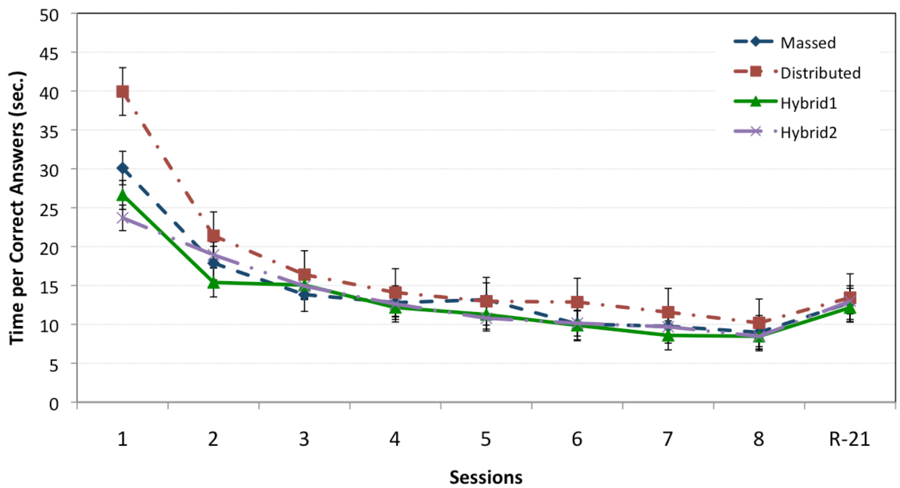


Figure **4-8**: The latency for the Permutation problem-solving task with respect to the four practice schedules.

Table **4-9**: The accuracy for the Permutation problem-solving task at the last learning session and the retention session with respect to the four practice schedules.

	Last Learning Session	Retention Session (R-21)
Massed	.97	.91
Distributed	.98	.93
Hybrid1	.97	.89
Hybrid2	.97	.92

Table **4-10**: The latency (sec.) for the Permutation problem-solving task at the last learning session and the retention session with respect to the four practice schedules.

	Last Learning Session	Retention Session (R-21)
Massed	8.93	12.92
Distributed	10.44	13.91
Hybrid1	8.57	12.13
Hybrid2	8.46	12.99

For the Tower of Hanoi puzzle, the task completion time (figure 4-9) and the number of disk movements (figure 4-10) were recorded and analyzed. The task completion times at the first learning session are somewhat different among the schedules. The participants in the massed practice group completed the task in 401.60 sec., ones in the distributed group completed the task in 465.22 sec., ones in the Hybrid1 group completed the task in 638.28 sec., and ones in the Hybrid2 group completed the task in 519.57 sec. at the first learning session. Those values decrease gradually with practice, and participants in the massed group completed the task in 138.57 sec., ones in the distributed group in 153.38 sec., ones in the Hybrid1 group in 203.16 sec., ones in the Hybrid2 group in 102.57 sec. at the final learning session. Finally, the task completion time of the massed group is 172.33 sec., the distributed group is 167.72 sec., the Hybrid1 group is 185.47 sec., and the Hybrid2 group is 124.44 sec. at the retention session. These results are presented in table 4-11.

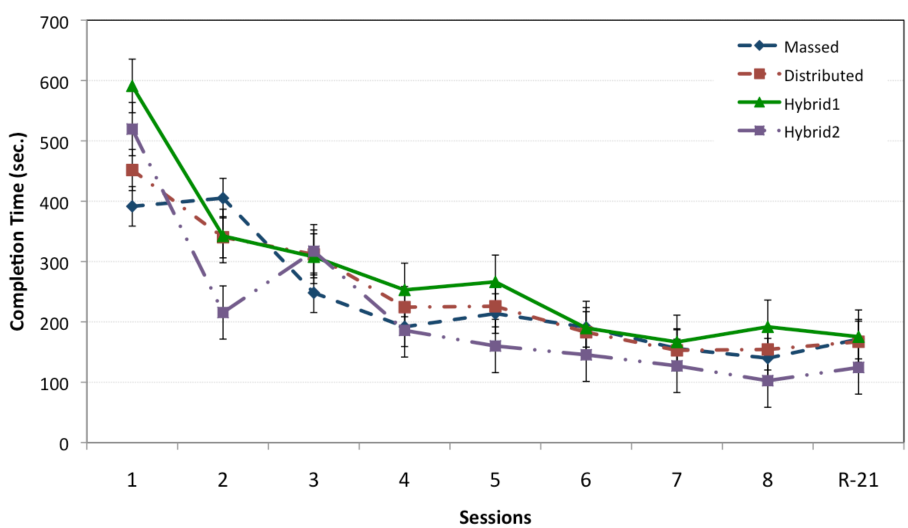


Figure **4-9**: The task completion time for the Tower of Hanoi puzzle with respect to the four schedules.

Table **4-11**: The task completion time (sec.) for the Tower of Hanoi puzzle at the last learning session and retention session with respect to the four schedules.

	Last Learning Session	Retention Session (R-21)
Massed	138.57	172.33
Distributed	153.38	167.72
Hybrid1	203.16	185.47
Hybrid2	102.57	124.44

As we can see the figure 4-10, the shapes of graph for the number of disk movements are almost identical to the ones for the task completion time. The participants in the massed group moved disks 167.5 times, ones in the distributed group moved 183.90 times, ones in the Hybrid1 group moved 248 times, and ones in the Hybrid2 group moved 241 times. Those values are gradually decrease with practice, and the massed group moved disks 73.40 times, the distributed group moved 84.10 times, the Hybrid1 group moved 115.70 times, and the Hybrid2 group moved 78.2 times at the final training session. At the retention session, the number of disk movements for the massed group is 84.20 times, for the distributed group is 81.40 times, for the Hybrid1 group is 108.50 times, and for the Hybrid2 group is 94 times. Table 4-12 shows these results.

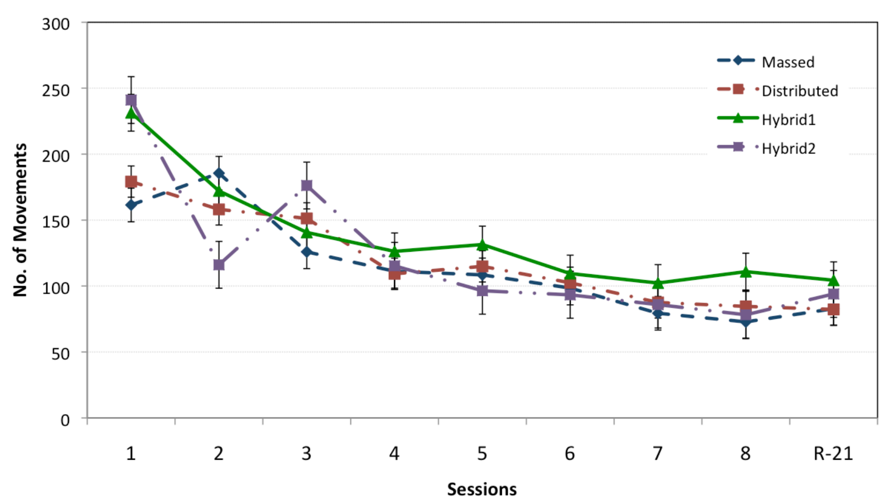


Figure **4-10**: The number of disk movements for the Tower of Hanoi puzzle with respect to the four schedules.

Table **4-12**: The number of disk movements for the Tower of Hanoi puzzle at the last learning session and retention session with respect to the four schedules.

	Last Learning Session	Retention Session (R-21)
Massed	73.4	84.2
Distributed	84.1	81.4
Hybrid1	115.7	108.5
Hybrid2	78.2	94.0

There are very little differences between the last learning session and the retention session for the task completion time and the number of disk movement in the Tower of Hanoi puzzle regardless of the different practice schedules. It shows that the knowledge for solving the Tower of Hanoi puzzle was proceduralized in the declarative and procedural memories of the participants during the learning sessions. It is also consistent with that when the proceduralized happens in human mind, the degradation rarely happens.

For the Inverted Pendulum task, the duration time was recorded and analyzed. Figure 4-11 shows the experiment results of the duration time for the Inverted Pendulum task in each session with respect to the four practice schedules. The average duration time for the participants in the massed group is 13.68 sec., for ones in the distributed group is 17.54 sec., for ones in the Hybrid1 group is 11.68, and for ones in the Hybrid2 group is 9.32 sec. at the first learning session. These duration times increase with practice, and participants in the massed group reached 74.12 sec., ones in the distributed group reached 29.87 sec., ones in the Hybrid1 group reached 148.26 sec., and ones in the Hybrid2 group reached 98.32 sec. at the final learning session. At the retention session, the massed group kept balancing during 109.00 sec., the distributed group kept balancing during 25.39 sec., the Hybrid1 group kept balancing during 146.36 sec., and the Hybrid2 group kept balancing during 76.67 sec. These results are presented in table 4-13.

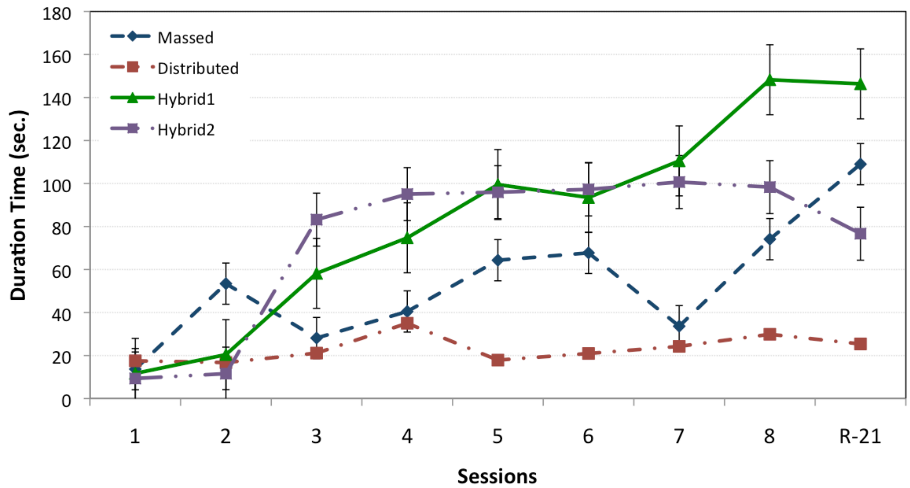


Figure **4-11**: The duration time of the Inverted Pendulum task with respect to the four schedules.

Table **4-13**: The duration time (sec.) of the Inverted Pendulum task at the last learning session and retention session with respect to the four schedules.

	Last Learning Session	Retention Session (R-21)
Massed	74.12	109.00
Distributed	29.87	25.39
Hybrid1	148.26	146.36
Hybrid2	98.32	76.67

The participants in the massed group performed better balancing at the retention session than the last learning session, and ones in the Hybird1 group shows almost the same performance between the retention session and the last learning session. It shows that the perceptual-motor skills of participants in those groups were procedurlized in their minds during the learning sessions. In contrast, the participants in the distributed group could not proceduralized their perceptual-motor skills through the distributed practice schedule.

4.2.3 Comparison of Performance Improvement

To examine the efficiency of each practice schedule, I analyzed the performance improvement that is described the equation 4.1, by comparing the performance at the first learning session to the performance at the retention session. The performance improvement of each group is compared using analysis of variance and a t-test.

$$\textit{Performance Improvement} = \frac{\textit{Performance Difference between the first learning session and the retention session}}{\textit{Performance at the first training session or retention session}} \qquad \text{Equation 4.1}$$

4.2.3.1 The Japanese Vocabulary Test

The performance improvement for the accuracy and latency of the Japanese vocabulary test are the same as the accuracy and the latency at the retention session, because, the accuracy and latency of this task at the first training session were zero. Figure 4-12 shows the boxplots for the accuracy and latency of the Japanese vocabulary test with respect to the practice schedules.

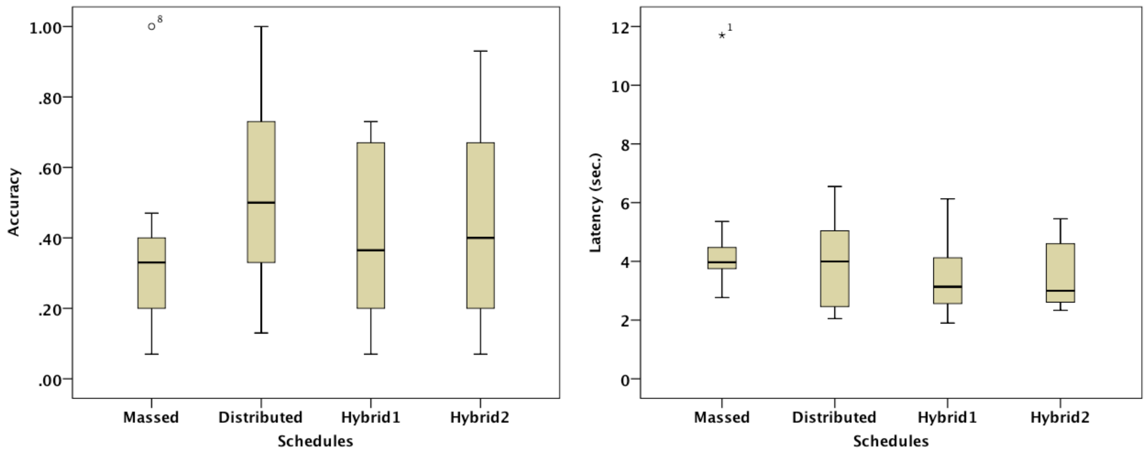


Figure **4-12**: Boxplots (*N* = 10 per groups) showing the accuracy and latency for the correct answers of the Japanese vocabulary test at the retention session with respect to all groups (There are outliers in the massed practice group for the results of accuracy and latency).

Two one-way analyses of variances were also conducted to check the performance improvement according to the practice schedules, and these results are presented in Table 4-14. The accuracy and latency for the correct answers of the Japanese vocabulary test for the performance improvement are not significant different among the groups, $F(3,36) = .61, p > .05$, and $F(3,36) = 1.33, p > .05$ respectively.

Table **4-14**: ANOVA table for the accuracy and the latency for the correct answers of the Japanese vocabulary test for the performance improvement.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Accuracy	Between Groups	3	0.13	0.04	0.61	.61
	Within Groups	36	2.57	0.07		
	Total	39	2.69			
Latency	Between Groups	3	11.72	3.91	1.33	.28
	Within Groups	36	105.87	2.94		
	Total	39	117.59			

These results are different from the results of the previous works that showed a distributed practice schedule provides better performance at the retention session than a massed practice schedule. I think the outliers in the massed group and two hybrid schedules that were compared together cause little difference among the schedules. So, I exclude the outlier that shows perfect accuracy (1.0) at the retention session in the massed group, and did the analysis of variance again, and the results are presented in table 4-15. There were still no significant different among groups, $F(3,35) = 1.44$, $p > .05$ for accuracy, and $F(3,35) = 1.71$, $p > .05$ for latency of correct answers.

Table **4-15**: ANOVA table for the accuracy and the latency for correct answers of the Japanese vocabulary test for the performance improvement excludes outlier in the massed group.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Accuracy	Between Groups	3	0.26	0.08	1.44	.25
	Within Groups	35	2.10	0.06		
	Total	38	2.36			
Latency	Between Groups	3	14.83	4.94	1.71	.18
	Within Groups	35	101.39	2.89		
	Total	38	116.22			

To examine the difference between each pair of groups, I performed a series of independent-sample t-tests. Each group compared with the other groups, so six in total independent t-tests were conducted. There are no significant differences in the accuracy and latency on the retention session between all of the groups; between the massed and the distributed groups, $t(18) = -1.31$, $p > .05$ for accuracy and $t(18) = 0.83$, $p > .05$ for latency, between the massed and the Hybrid1 groups, $t(18) = -0.48$, $p > .05$ for accuracy and $t(18) = 1.52$, $p > .05$ for latency, between the massed and the Hybrid2 groups, $t(18) = -0.66$, $p > .05$ for accuracy and $t(18) = 1.46$, $p > .05$ for latency, between the distributed and the Hybrid1 groups, $t(18) = 0.9$, $p > .05$ for accuracy and $t(18) = 0.91$, $p > .05$ for latency, between the distributed and the Hybrid2 groups, $t(18) = 0.62$, $p > .05$ for accuracy and $t(18) = 0.82$, $p > .05$ for latency, and between the Hybrid1 and the Hybrid2 groups, $t(18) = -0.22$, $p > .05$ for accuracy and $t(18) = -0.17$, $p > .05$ for latency. These results are presented in table 4-16.

Table **4-16**: Independent-sample t-tests for the accuracy and the latency for correct answers of the Japanese vocabulary test.

	Schedules		Mean Diff.	SE Diff	*df*	*t*	*p*	95% CI of the Difference LB	95% CI of the Difference UB
Accuracy	Massed	Distributed	-0.16	0.12	18	-1.31	.21	0.41	0.10
	Massed	Hybrid1	-0.05	0.11	18	-0.48	.63	-0.29	0.18
	Massed	Hybrid2	-0.08	0.12	18	-0.65	.52	-0.37	0.18
	Distributed	Hybrid1	0.11	0.12	18	0.90	.38	-0.14	0.35
	Distributed	Hybrid2	0.08	0.13	18	0.62	.54	-0.19	0.35
	Hybrid1	Hybrid2	0.03	0.11	18	-0.22	.83	-0.27	0.22
Latency	Massed	Distributed	0.78	0.93	18	0.83	.42	-1.18	2.74
	Massed	Hybrid1	1.37	0.90	18	1.52	.15	-0.53	3.26
	Massed	Hybrid2	1.27	0.87	18	1.46	.16	-0.56	3.10
	Distributed	Hybrid1	0.59	0.64	18	0.91	.37	-0.76	1.94
	Distributed	Hybrid2	0.49	0.60	18	0.82	.42	-0.77	1.76
	Hybrid1	Hybrid2	-0.09	0.55	18	-0.17	.86	-1.26	1.06

As I did the second ANOVA that excludes outlier in the massed group, I performed the three independent-sample t-tests (massed vs. distributed, massed vs. Hybrid1, and massed vs. Hybrid2) without the outlier in the massed group. As we see from table 4-17, there is significant difference between the massed and the distributed groups in accuracy, $t(18) = -2.23$, $p < .05$, two-tailed, and there is no significant difference in latency, $t(18) = 1.04$, $p > .05$. There are no significant differences between the massed and the Hybrid1 group, and between the massed and the Hybrid2 group with respect to the accuracy and the latency. These results show that the distributed schedule provides better performance than the massed schedule, and those results are consistent with the previous studies.

Table **4-17**: Three independent-sample t-tests for the accuracy and the latency for correct answers of the Japanese vocabulary test without the outlier in the massed group.

	Schedules		Mean Diff.	SE Diff	*df*	*t*	*p*	95% CI of the Difference LB	UB
Accuracy	Massed	Distributed	-0.23	0.10	17	-2.23	.04*	-0.45	-0.01
	Massed	Hybrid1	-0.12	0.09	17	-1.40	.18	-0.32	0.06
	Massed	Hybrid2	-0.15	0.10	17	-1.45	.16	-0.37	0.07
Latency	Massed	Distributed	1.00	0.96	17	1.04	.31	-1.02	3.02
	Massed	Hybrid1	1.59	0.92	17	1.72	.10	-0.36	3.54
	Massed	Hybrid2	1.49	0.89	17	1.68	.11	-0.39	3.38

However, I found an interesting result between the distributed and the Hybrid2 schedules. As I explained in chapter 4, the distributed schedule has one session per one day over eight days, and between the fourth and fifth sessions there are four days rest period, so the training session ends at the 11th day from the first training session. The Hybrid2 schedule also ends its last training session at the 11th day from the first training session. So, we can compare the decrease rate of each schedule according to days between the last training session and the retention session. Figure 4-13 shows the accuracy of the distributed and the Hybrid2 schedule according to days, and I assumed that forgetting of the Japanese vocabulary follows a linear relation. We can find that the Hybrid2 schedule may provide better retention when the retention session takes place within 12 days from the end of the training day (before day 24). It also shows that the effective schedule varies according to the length of retention interval.

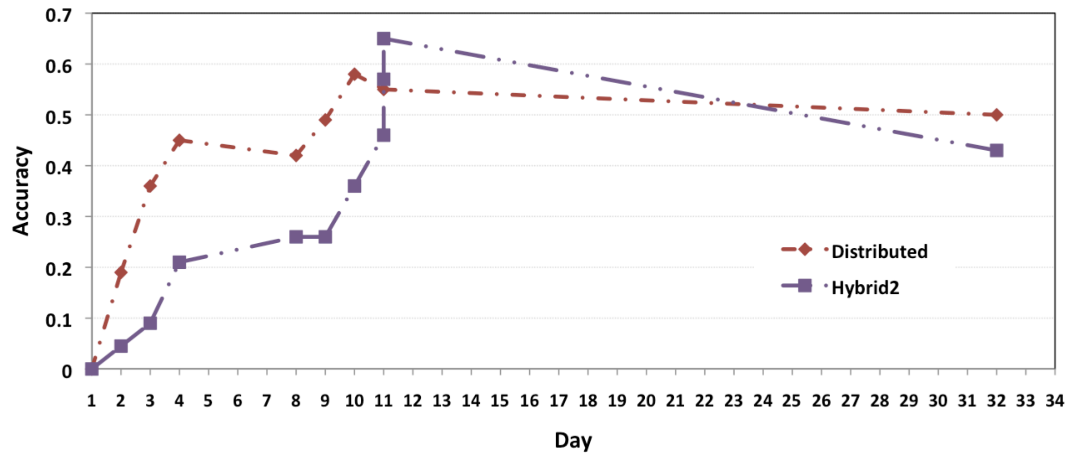


Figure **4-13**: The accuracy of the distributed and Hybrid2 schedules for the Japanese vocabulary test.

4.2.3.2 The Permutation Problem-Solving Task

The performance improvement for the accuracy and latency of the Permutation problem-solving task are calculated according to the equation 4.1. Figure 4-14 shows the results of the *performance improvement* (equation 4.1) of accuracy and latency for the correct answers of the Permutation problem-solving task with respect to the practice schedules. There are outliers in the distributed group of accuracy graph, and in the distributed, Hybrid1, and Hybrid2 groups of latency graph. I think those outliers are not from the different schedules, but from mistakes during the sessions or individual differences to solve the problems, because, as we see the figures 4-7 and 4-8 in the previous section, the shapes of the graphs for all groups are very similar.

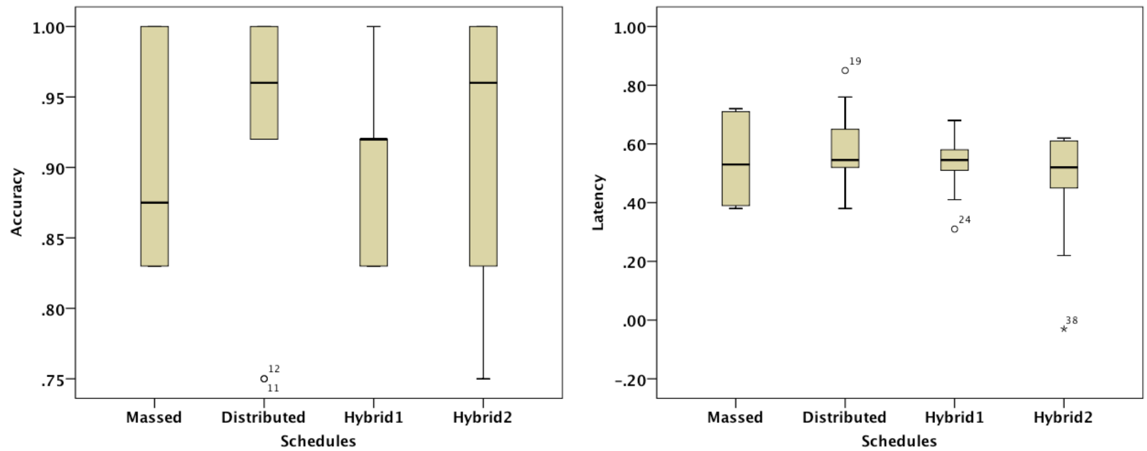


Figure **4-14**: Boxplots showing the performance improvement rate of accuracy and latency for correct answers of the Permutation problem-solving task with respect to all groups (There are outliers in the distributed, Hybrid1, and Hybrid2 practice groups.

To examine the differences in performance improvement among the groups, a series of one-way analyses of variance were conducted, and these results are presented in Table 4-18. The performance improvement of the accuracy and latency for the correct answers of the Permutation problem-solving task are not significantly different among the groups, $F(3,36) = 0.27$, $p > .05$, and $F(3,36) = 1.26$, $p > .05$ respectively.

Table **4-18**: ANOVA table for the improved performance rate of the accuracy and the latency for correct answers of the Permutation problem-solving task.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Accuracy	Between Groups	3	0.01	0.00	0.27	.85
	Within Groups	36	0.28	0.01		
	Total	39	0.29			
Latency	Between Groups	3	0.09	0.03	1.26	.30
	Within Groups	36	0.83	0.02		
	Total	39	0.92			

To examine the difference between each pair of groups, I performed a series of independent-sample t-tests. Each group compared with the other groups with respect to the accuracy and the latency, so totally 12 of the independent t-tests were conducted. Table 4-19 shows the results of the independent-sample t-tests.

Table **4-19**: Independent-sample t-tests for the performance improvement of the accuracy and the latency for the correct answers of the Permutation problem-solving task.

	Schedules		Mean Diff.	SE Diff	*df*	*t*	*p*	95% CI of the Difference	
								LB	UB
Accuracy	Massed	Distributed	-0.02	0.04	18	-0.46	.65	-0.11	0.07
	Massed	Hybrid1	-0.13	0.09	18	-1.40	.18	-0.32	0.06
	Massed	Hybrid2	-0.15	0.10	18	-1.45	.16	-0.37	0.07
	Distributed	Hybrid1	0.03	0.34	18	0.93	.37	-0.04	0.11
	Distributed	Hybrid2	0.01	0.05	18	0.20	.85	-0.09	0.10
	Hybrid1	Hybrid2	-0.03	0.04	18	-0.66	.52	-0.10	0.05
Latency	Massed	Distributed	-0.05	0.06	18	-0.82	.42	-0.18	0.08
	Massed	Hybrid1	1.59	0.92	18	1.72	.10	-0.36	3.54
	Massed	Hybrid2	1.49	0.89	18	1.68	.11	-0.39	3.38
	Distributed	Hybrid1	0.06	0.05	18	1.12	.28	-0.05	0.18
	Distributed	Hybrid2	0.13	0.08	18	1.67	.11	-0.03	0.30
	Hybrid1	Hybrid2	0.07	0.07	18	0.95	.35	-0.08	0.22

There are no significant differences between all groups in performance improvement of accuracy and latency for the correct answers in the Permutation problem-solving task as described in the table 4-19. These results are different with the previous study (Rohrer & Taylor, 2006) that showed the distributed practice schedule produces better performance than the massed practice schedule in similar permutation problem-solving task. The reason of these differences is maybe because the study of Rohrer and Talyor had just one learning session for the massed group (participants in this group solved 10 problems) and one retention test, and two learning session for the distributed group (participants in this group solved five problems in each session) and one retention test, so it was not enough to learn this kind of permutation problems. However, there were eight learning sessions in my experiment, so the knowledge to solve this kind of problem might be proceduralized in the memory of participants in all groups as we can see in figure 4-7 and 4-8.

4.2.3.3 The Tower of Hanoi Puzzle

The *performance improvements* for the task completion time and the number of disk movements of Tower of Hanoi puzzle are calculated according to the equation 4.1. Figure 4-15 shows the boxplots of the performance improvements of the task completion time and the number of disk movements for this task with respect to the practice schedules. There is an outlier in the Hybrid1 group of the task completion time graph. The reason for this outlier is this subject showed good performance at the first learning session, so the improved performance rate was low at the retention test. However, there is no effect of training schedule on learning the Tower of Hanoi puzzle that I present in section 4.2.1, so I leave it without any further analysis.

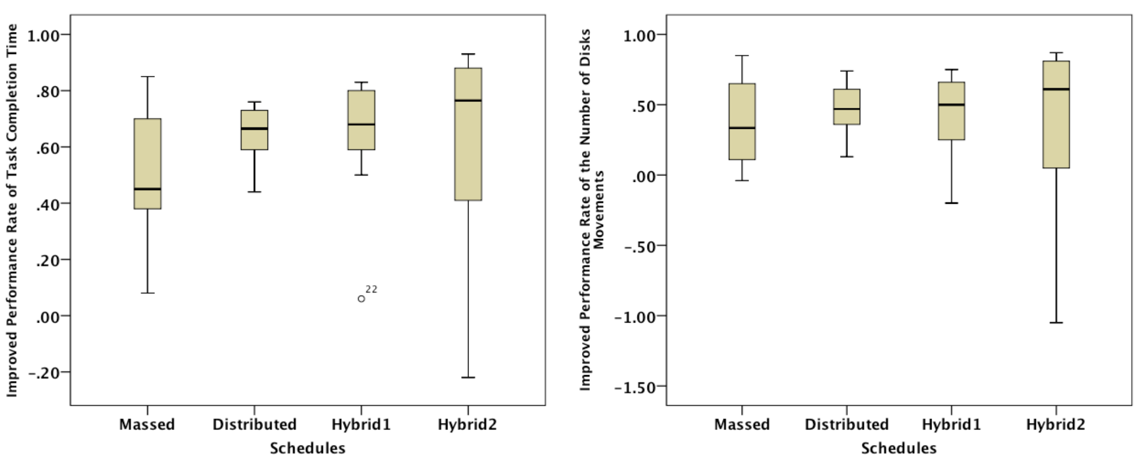


Figure **4-15**: Boxplots showing the performance improvement of the task completion time and the number of disk movements for the Tower of Hanoi puzzled with respect to all groups (There is an outlier in the Hybrid1 practice group).

To examine the differences in performance improvement among the groups, two times of one-way analyses of variance were conducted, and these results are presented in Table 4-20. The performance improvements of the task completion time and the number of disk movements for this task are not significantly different among the groups, $F(3,36) = 0.72$, $p > .05$ for the task completion time, and $F(3,36) = 0.26$, $p > .05$ for the number of disk movements.

Table **4-20**: ANOVA table for the performance improvements of the task completion time and the number of disk movements for the Tower of Hanoi puzzle.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Accuracy	Between Groups	3	0.15	0.05	0.72	.55
	Within Groups	36	2.42	0.07		
	Total	39	2.56			
Latency	Between Groups	3	0.11	0.04	0.26	.85
	Within Groups	36	5.12	0.14		
	Total	39	5.23			

To examine the difference between each pair of groups, I performed a series of independent-sample t-tests. Each group compared with the other groups with respect to the task completion time and the number of disk movements, so 12 in total independent t-tests were conducted. Table 4-21 shows the results of the independent-sample t-tests.

Table **4-21**: Independent-sample t-tests for the performance improvements of the task completion time and the number of disk movements for the Tower of Hanoi puzzle.

	Schedules		Mean Diff.	SE Diff	*df*	*t*	*p*	95% CI of the Difference LB	UB
Task Completion Time	Massed	Distributed	-0.15	0.08	18	-1.91	.08	-0.32	0.02
	Massed	Hybrid1	-0.14	0.10	18	-1.37	.19	-0.35	0.08
	Massed	Hybrid2	-0.08	0.14	18	-0.57	.58	-0.38	0.20
	Distributed	Hybrid1	0.01	0.08	18	0.15	.88	-0.15	0.18
	Distributed	Hybrid2	0.07	0.13	18	0.56	.59	-0.19	0.34
	Hybrid1	Hybrid2	0.06	0.14	18	0.41	.69	-0.24	0.36
Number of Disks Movements	Massed	Distributed	-0.12	0.11	18	-1.05	.31	-0.35	0.12
	Massed	Hybrid1	-0.08	0.132	18	-0.64	.53	-0.36	0.19
	Massed	Hybrid2	0.00	0.02	18	0.02	.99	-0.44	0.45
	Distributed	Hybrid1	0.03	0.11	18	0.31	.76	-0.19	0.26
	Distributed	Hybrid2	0.12	0.19	18	0.62	.55	-0.29	0.54
	Hybrid1	Hybrid2	0.09	0.21	18	0.42	.68	-0.35	0.53

There are no significant differences between all groups in performance improvements of the task completion time and the number of disk movements in the Tower of Hanoi puzzle as described in the table 4-21. These results may show that learning and forgetting for the task that requires procedural knowledge are not related with the different training schedules, but related with the amount of practice, or the effect of different practice schedules could not be found in my experiment.

4.2.3.4 The Inverted Pendulum task

The performance improvement for the duration time of the Inverted Pendulum task is calculated according to the equation 4.1. Figure 4-16 shows the boxplot of the performance improvement of the duration time for this task with respect to the practice schedules.

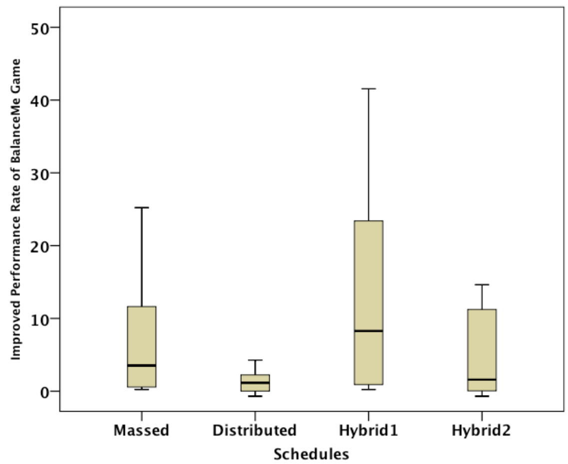


Figure **4-16**: A boxplot showing the improved performance rate of the duration time for the Inverted Pendulum task with respect to all groups (*N*=10 per schedule, 40 total, and there is significant difference between the distributed and Hybrid1 practice groups).

To examine the differences in performance improvement among the groups, a one-way analysis of variance was conducted, and the results are presented in Table 4-22. The performance improvement of the duration time for this task is significantly different among the groups, $F(3,36) = 3.35$, $p < .05$.

Table **4-22**: ANOVA table for the performance improvement of the duration time for the Inverted Pendulum task.

		df	*Sum of Squares*	*Mean Square*	*F*	*p*
Accuracy	Between Groups	3	843.16	281.05	3.35	.03
	Within Groups	36	3021.85	83.94		
	Total	39	3865.03			

To examine the difference between each pair of groups, I performed a series of independent-sample t-tests. Each group compared with the other groups with respect to the duration time, so totally six of the independent t-tests were conducted. Table 4-23 shows the results of the independent-sample t-tests.

Table **4-23**: An Independent-sample t-test for the performance improvement of the duration time for the Inverted Pendulum task.

Schedules		Mean Diff.	SE Diff	*df*	*t*	*p*	95% CI of the Difference LB	UB
Massed	Distributed	5.54	2.60	18	2.13	.06	0.07	10.99
Massed	Hybrid1	-6.94	5.47	18	-1.27	.23	-18.70	4.82
Massed	Hybrid2	2.41	3.16	18	0.76	.46	-4.26	9.09
Distributed	Hybrid1	-12.47	4.86	18	-2.57	.03*	-23.43	-1.52
Distributed	Hybrid2	-3.12	1.91	18	-1.64	.13	-7.37	1.13
Hybrid1	Hybrid2	9.35	5.18	18	1.81	.09	-1.98	20.67

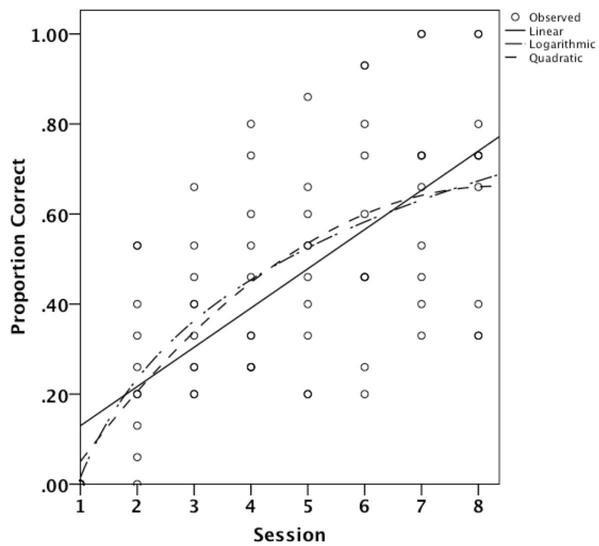
As we can see from table 4-23, there is significant difference between the distributed and Hybrid1 groups in duration time, $t(18) = -2.57$, $p < .05$, two-tailed, and marginally significant difference between the massed and distributed group, and there are no significant differences between all other groups. These results show that the Hybrid1 schedule provides better performance than the distributed schedule in this task that requires perceptual-motor skill. We can also assume that the massed or somewhat early massed practice schedules provide better learning effects in the task that requires perceptual-motor skills. These results prove the assumption that I mentioned in the chapter 1 that the task required the perceptual-motor skill should be trained in a massed way rather than a distributed way.

4.2.4 Learning Trends for Each Group

In this section, I present the estimated learning trends in each group with respect to the task measurements. I estimated to the curve with power, exponential, linear, log, and quadratic models. By doing this estimation, I can figure out the learning trends of each training schedule and each task, thus it enables us to estimate the performance of further learning sessions.

4.2.4.1 Massed Practice Group

Figure 4-17 shows the estimated learning trends for the proportion correct answers of the Japanese vocabulary test for the massed practice group. The linear, logarithmic, and quadratic models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-24.



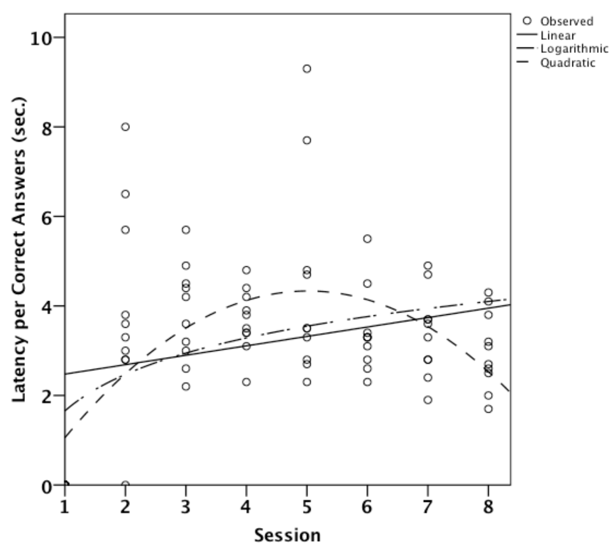
Linear: $y = 0.087x + 0.043,\ R^2 = 0.50$
Logarithmic: $y = 0.317\ln(x) + 0.015,\ R^2 = 0.54$
Quadratic: $y = -0.011x^2 + 0.189x - 0.128,\ R^2 = 0.53$

Figure **4-17**: Learning trends for the proportion correct answers of the Japanese vocabulary test for the massed practice group (N=10).

Table **4-24**: Statistical output of curve estimation for the proportion correct answers of the Japanese vocabulary test for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.50	78	77.41	.00
Logarithmic	.54	78	92.56	.00
Quadratic	.53	77	43.74	.00

Figure 4-18 shows the estimated learning trends for the latency per correct answer of the Japanese vocabulary test for the massed practice group. The linear, logarithmic, and quadratic models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-25.



Linear: $y = 0.21x + 2.267,\ R^2 = 0.07$

Logarithmic: $y = 1.175\ln(x) + 1.656,\ R^2 = 0.18$

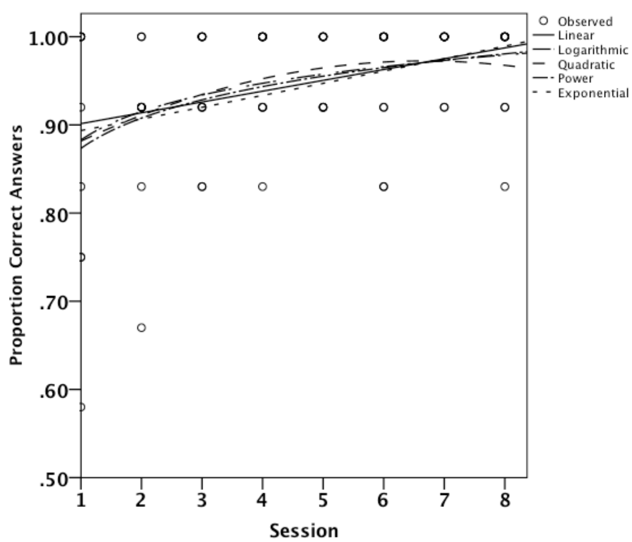
Quadratic: $y = -0.203x^2 + 2.037x - 0.777,\ R^2 = 0.33$

Figure **4-18**: Learning trends for the latency per correct answer of the Japanese vocabulary test for the massed practice group (*N*=10).

Table **4-25**: Statistical output of curve estimation for the latency per correct answer of the Japanese vocabulary test for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.07	78	5.93	.02
Logarithmic	.18	78	17.33	.00
Quadratic	.33	77	19.27	.00

Figure 4-19 shows the estimated learning trends for the proportion correct answer of the Permutation problem-solving task for the massed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend and shown in table 4-26.



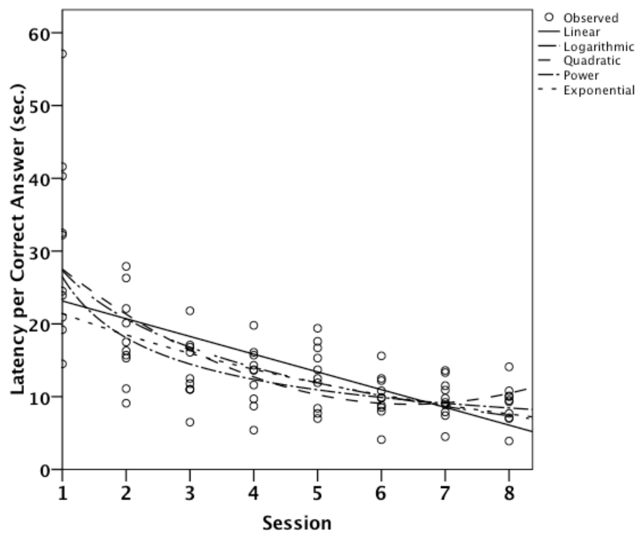
Linear: $y = 0.12x + 0.889,\ R^2 = 0.12$
Logarithmic: $y = 0.046\ln(x) + 0.883,\ R^2 = 0.14$
Quadratic: $y = -0.003x^2 + 0.038x + 0.847,\ R^2 = 0.14$
Power: $y = x^{0.056} + 0.873,\ R^2 = 0.14$
Exponential: $y = 0.881e^{0.015x},\ R^2 = 0.12$

Figure **4-19**: Learning trends for the proportion correct answers of the Permutation problem-solving task for the massed practice group (*N*=10).

Table **4-26**: Statistical output of curve estimation for the proportion correct answer of the Permutation problem-solving task for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.12	78	10.33	.02
Logarithmic	.14	78	12.44	.00
Quadratic	.14	77	6.36	.00
Power	.14	78	12.91	.00
Exponential	.12	78	10.37	.00

Figure 4-20 shows the estimated learning trends for the latency per correct answer of the Permutation problem-solving task for the massed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-27.



Linear: $y = -2.437x + 25.581,\ R^2 = 0.41$

Logarithmic: $y = -9.613\ln(x) + 27.358,\ R^2 = 0.14$

Quadratic: $y = 0.625x^2 - 8.064x + 34.96,\ R^2 = 0.52$

Power: $y = x^{-0.549} + 26.468,\ R^2 = 0.14$

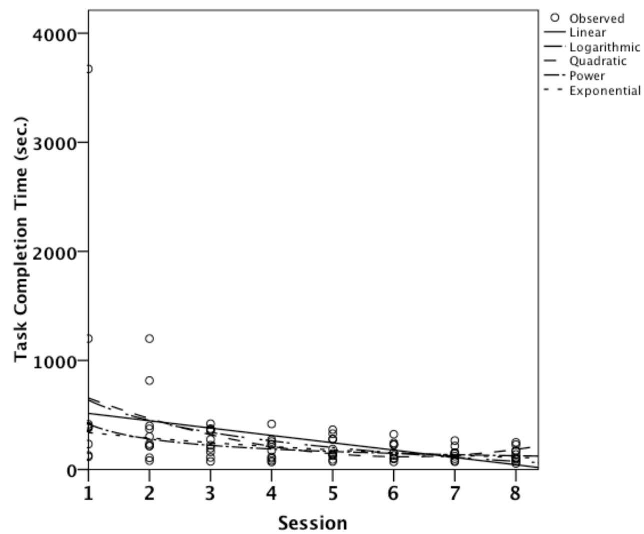
Exponential: $y = 24.8361e^{-0.148x},\ R^2 = 0.46$

Figure **4-20**: Learning trends for the latency per correct answer of the Permutation problem-solving task for the massed practice group (*N*=10).

Table **4-27**: Statistical output of curve estimation for the latency per correct answer of the Permutation problem-solving task for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.41	78	54.28	.02
Logarithmic	.53	78	86.75	.00
Quadratic	.52	77	41.44	.00
Power	.52	78	84.15	.00
Exponential	.46	78	65.06	.00

Figure 4-21 shows the estimated learning trends for the task completion time of the Tower of Hanoi puzzle for the massed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-28.



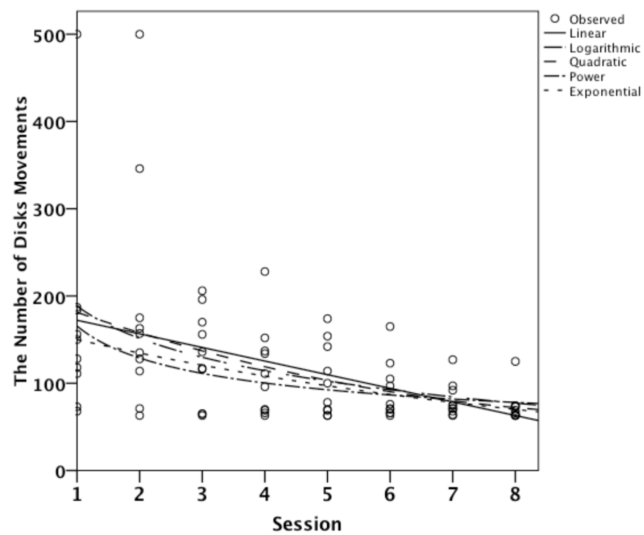
Linear: $y = -67.291x + 581.587,\ R^2 = 0.13$
Logarithmic: $y = -270.141\ln(x) + 636.87\ \ R^2 = 0.17$
Quadratic: $y = 20.155x^2 - 248.69x + 883.919,\ R^2 = 0.17$
Power: $y = x^{-0.577} + 417.524,\ R^2 = 0.28$
Exponential: $y = 396.651e^{-0.159x},\ R^2 = 0.25$

Figure **4-21**: Learning trends for the task completion time of the Tower of Hanoi Puzzle for the massed practice group (N=10).

Table **4-28**: Statistical output of curve estimation for the task completion time of the Tower of Hanoi puzzle for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.13	78	54.28	.00
Logarithmic	.17	78	86.75	.00
Quadratic	.17	77	41.44	.00
Power	.28	78	84.15	.00
Exponential	.25	78	65.06	.00

Figure 4-22 shows the estimated learning trends for the number of disk movements of the Tower of Hanoi puzzle for the massed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-29.



Linear: $y = -15.595x + 187.779,\ R^2 = 0.20$

Logarithmic: $y = -53.507\ln(x) + 188.528\ \ R^2 = 0.20$

Quadratic: $y = 1.295x^2 - 27.252x + 207.207,\ R^2 = 0.17$

Power: $y = x^{-0.362} + 165.626,\ R^2 = 0.25$

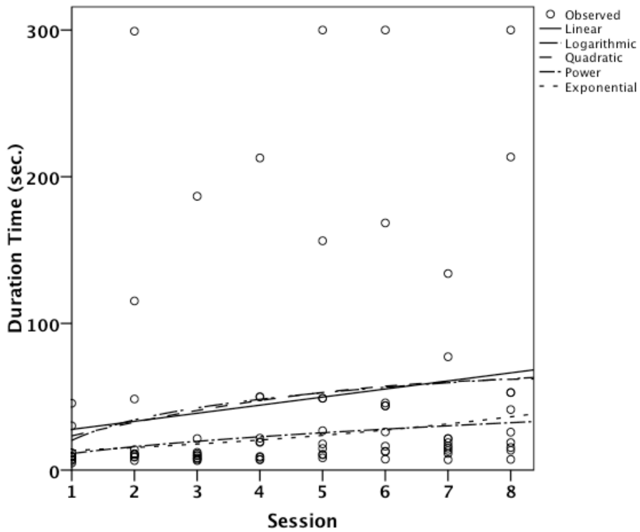
Exponential: $y = 167.292e^{-0.109x},\ R^2 = 0.27$

Figure **4-22**: Learning trends for the number of disk movements of the Tower of Hanoi Puzzle for the massed practice group (*N*=10).

Table **4-29**: Statistical output of curve estimation for the number of disk movements of the Tower of Hanoi puzzle for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.20	78	20.00	.00
Logarithmic	.20	78	19.27	.00
Quadratic	.17	77	10.22	.00
Power	.25	78	25.62	.00
Exponential	.27	78	29.01	.00

Figure 4-23 shows the estimated learning trends for the duration time of the BalanceMe Game for the massed practice group. The power and exponential models are significant ($p < .05$) to the estimated learning trends; however, the linear, logarithmic, and quadratic models are not significant ($p < .05$), shown in table 4-30.



Linear: $y = 5.508x + 22.232,\ R^2 = 0.03$

Logarithmic: $y = 20.198\ln(x) + 20.243\ \ R^2 = 0.03$

Quadratic: $y = -0.631x^2 + 11.182x + 12.775,\ R^2 = 0.03$

Power: $y = x^{0.509} + 11.224,\ R^2 = 0.09$

Exponential: $y = 11.573e^{0.143x},\ R^2 = 0.09$

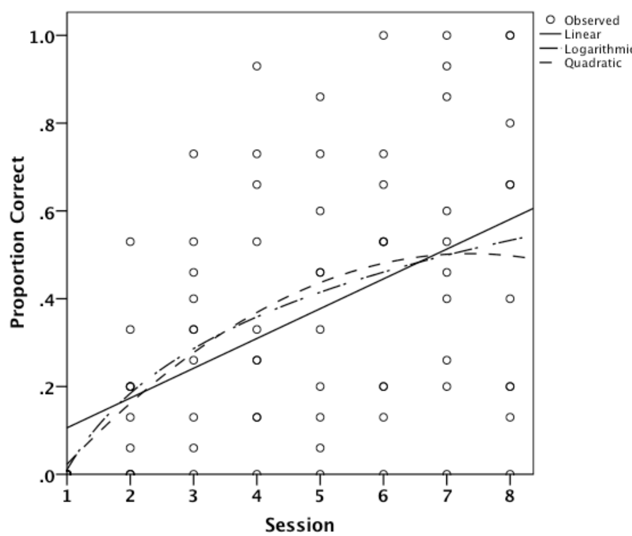
Figure **4-23**: Learning trends for the duration time of the Inverted Pendulum task for the massed practice group (N=10).

Table **4-30**: Statistical output of curve estimation for the duration time of the Inverted Pendulum task for the massed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.03	78	2.34	.13
Logarithmic	.03	78	2.61	.11
Quadratic	.03	77	1.22	.30
Power	.09	78	7.92	.00
Exponential	.09	78	7.56	.00

4.2.4.2 Distributed Group

Figure 4-24 shows the estimated learning trends for the proportion correct answers of the Japanese vocabulary test for the distributed practice group. The linear, logarithmic, and quadratic models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-31.



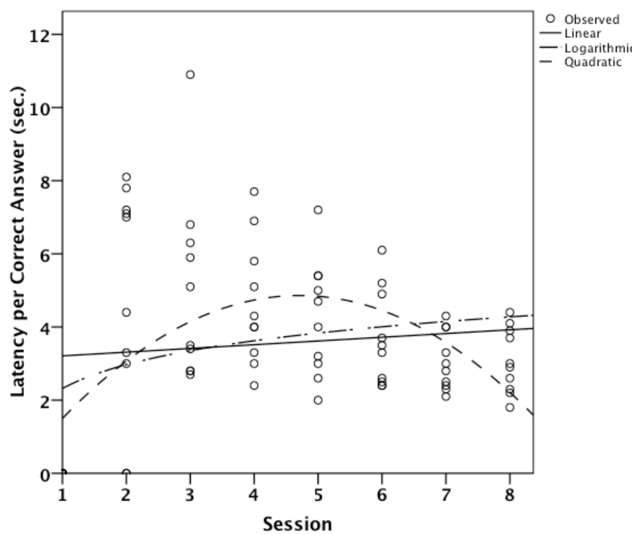
Linear: $y = 0.068x + 0.038,\ R^2 = 0.25$
Logarithmic: $y = 0.252\ln(x) + 0.01,\ R^2 = 0.29$
Quadratic: $y = -0.012x^2 + 0.175x - 0.14,\ R^2 = 0.29$

Figure **4-24**: Learning trends for the proportion correct answers of the Japanese vocabulary test for the distributed practice group (N=10).

Table **4-31**: Statistical output of curve estimation for the proportion correct answers of the Japanese vocabulary test for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.25	78	26.60	.00
Logarithmic	.29	78	31.55	.00
Quadratic	.29	77	15.38	.00

Figure 4-25 shows the estimated learning trends for the latency per correct answer of the Japanese vocabulary test for the distributed practice group. The quadratic model is significant ($p < .05$) to the estimated learning trend; however, the linear and logarithmic models are not significant ($p < .05$), and shown in table 4-32.



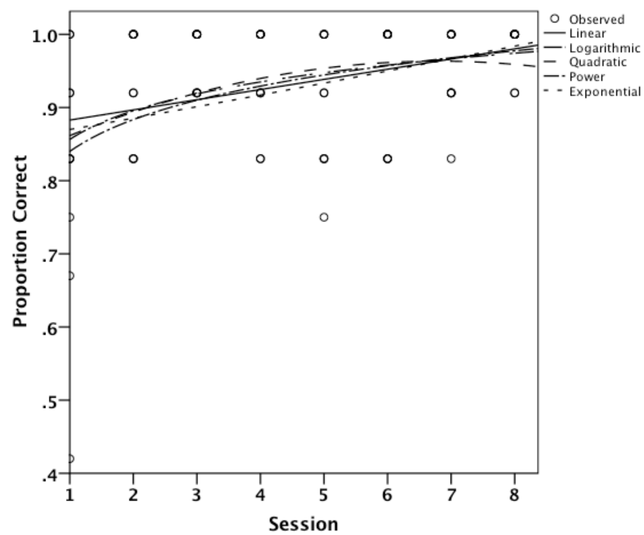
Linear: $y = 0.102x + 3.108,\ R^2 = 0.01$
Logarithmic: $y = 0.941\ln(x) + 2.318,\ R^2 = 0.08$
Quadratic: $y = -0.244x^2 + 2.301x - 0.557,\ R^2 = 0.26$

Figure **4-25**: Learning trends for the latency per correct answer of the Japanese vocabulary test for the distributed practice group (*N*=10).

Table **4-32**: Statistical output of curve estimation for the latency per correct answer of the Japanese vocabulary test for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.01	78	.84	.36
Logarithmic	.08	78	6.31	.14
Quadratic	.26	77	13.20	.00

Figure 4-26 shows the estimated learning trends for the proportion correct answer of the Permutation problem-solving task for the distributed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend and shown in table 4-33.



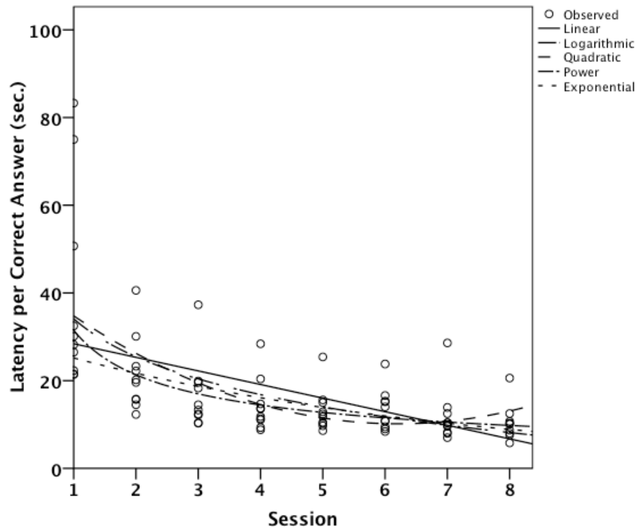
Linear: $y = 0.014x + 0.869,\ R^2 = 0.11$
Logarithmic: $y = 0.057\ln(x) + 0.856,\ R^2 = 0.15$
Quadratic: $y = -0.003x^2 + 0.041x + 0.823,\ R^2 = 0.13$
Power: $y = x^{0.073} + 0.84,\ R^2 = 0.15$
Exponential: $y = 0.855e^{0.018x},\ R^2 = 0.10$

Figure **4-26**: Learning trends for the proportion correct answers of the Permutation problem-solving task for the distributed practice group (N=10).

Table **4-33**: Statistical output of curve estimation for the proportion correct answer of the Permutation problem-solving task for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.11	78	9.47	.00
Logarithmic	.15	78	13.59	.00
Quadratic	.13	77	5.70	.00
Power	.15	78	13.50	.00
Exponential	.10	78	9.06	.00

Figure 4-27 shows the estimated learning trends for the latency per correct answer of the Permutation problem-solving task for the distributed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-34.



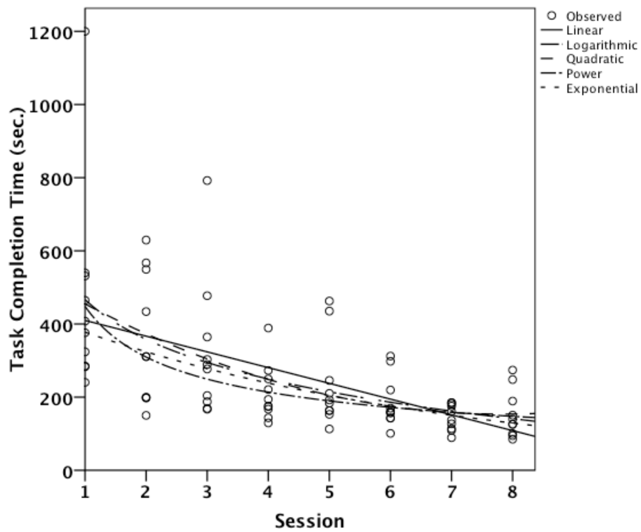
Linear: $y = -3.099x + 31.517,\ R^2 = 0.30$
Logarithmic: $y = -12.474\ln(x) + 34.105,\ R^2 = 0.41$
Quadratic: $y = 0.907x^2 - 11.264x + 45.124,\ R^2 = 0.41$
Power: $y = x^{-0.562} + 31.489,\ R^2 = 0.51$
Exponential: $y = 29.21e^{-0.149x},\ R^2 = 0.43$

Figure **4-27**: Learning trends for the latency per correct answer of the Permutation problem-solving task for the distributed practice group (*N*=10).

Table **4-34**: Statistical output of curve estimation for the latency per correct answer of the Permutation problem-solving task for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.30	78	34.12	.02
Logarithmic	.41	78	53.42	.00
Quadratic	.41	77	26.60	.00
Power	.51	78	80.49	.00
Exponential	.43	78	59.32	.00

Figure 4-28 shows the estimated learning trends for the task completion time of the Tower of Hanoi puzzle for the distributed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-35.



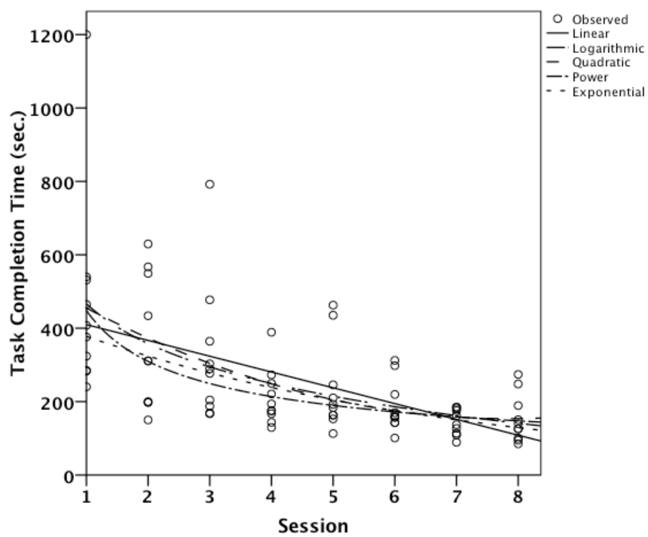
Linear: $y = -43.04x + 452.803,\ R^2 = 0.32$
Logarithmic: $y = -156.616\ln(x) + 466.731\ R^2 = 0.35$
Quadratic: $y = 6.452x^2 - 101.107x + 569.581,\ R^2 = 0.34$
Power: $y = x^{-0.534} + 448.304,\ R^2 = 0.43$
Exponential: $y = 440.208e^{-0.153x},\ R^2 = 0.43$

Figure **4-28**: Learning trends for the task completion time of the Tower of Hanoi Puzzle for the distributed practice group (N=10).

Table **4-35**: Statistical output of curve estimation for the task completion time of the Tower of Hanoi puzzle for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.32	78	36.00	.00
Logarithmic	.33	78	41.05	.00
Quadratic	.34	77	20.21	.00
Power	.43	78	59.44	.00
Exponential	.43	78	59.32	.00

Figure 4-29 shows the estimated learning trends for the number of disk movements of the Tower of Hanoi puzzle for the distributed practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-36.



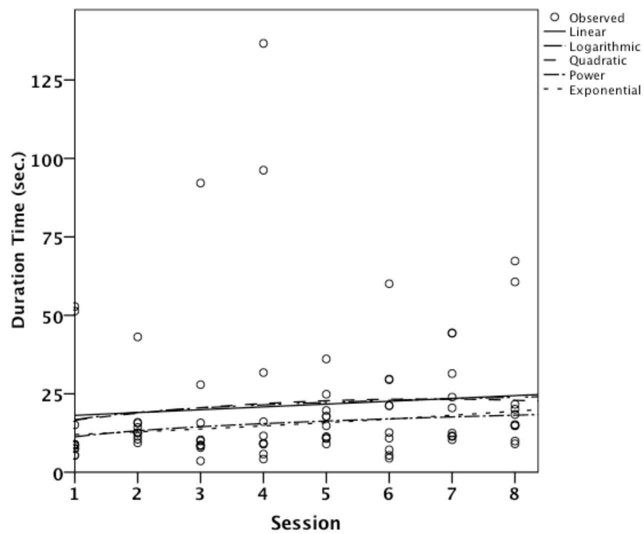
Linear: $y = -14.78x + 190.271,\ R^2 = 0.21$
Logarithmic: $y = -52.047\ln(x) + 192.754\ \ R^2 = 0.21$
Quadratic: $y = 1.561x^2 - 28.832x + 213.691,\ R^2 = 0.17$
Power: $y = x^{-0.346} + 174.169,\ R^2 = 0.26$
Exponential: $y = 174.256e^{-0.102x},\ R^2 = 0.28$

Figure **4-29**: Learning trends for the number of disk movements of the Tower of Hanoi Puzzle for the distributed practice group (N=10).

Table **4-36**: Statistical output of curve estimation for the number of disk movements of the Tower of Hanoi puzzle for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.21	78	20.24	.00
Logarithmic	.21	78	20.82	.00
Quadratic	.22	77	10.56	.00
Power	.26	78	27.50	.00
Exponential	.28	78	29.57	.00

Figure 4-30 shows the estimated learning trends for the duration time of the Inverted Pendulum task for the distributed practice group. The power and exponential models are marginally significant ($p < .05$) to the estimated learning trends; however, the linear, logarithmic, and quadratic models are not significant ($p < .05$), shown in table 4-37.



Linear: $y = 0.899x + 17.223,\ R^2 = 0.01$

Logarithmic: $y = 3.472\ln(x) + 16.666\ \ R^2 = 0.10$

Quadratic: $y = -0.213x^2 + 2.813x + 14.032,\ R^2 = 0.10$

Power: $y = x^{0.230} + 11.265,\ R^2 = 0.04$

Exponential: $y = 11.216e^{0.069x},\ R^2 = 0.04$

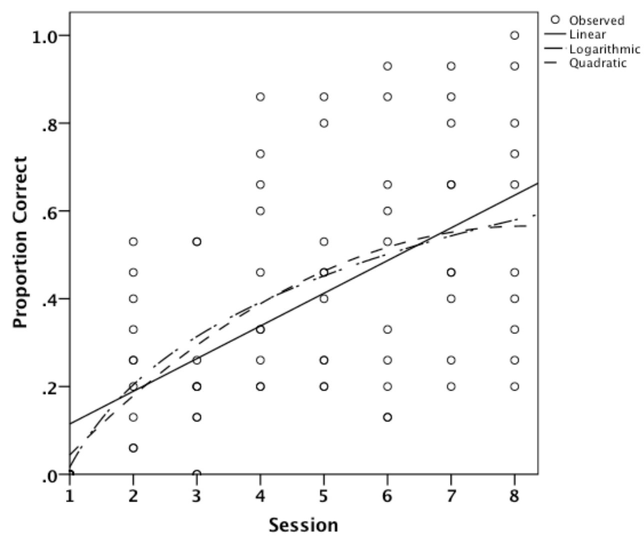
Figure **4-30**: Learning trends for the duration time of the Inverted Pendulum task for the distributed practice group (N=10).

Table **4-37**: Statistical output of curve estimation for the duration time of the Inverted Pendulum task for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.01	78	.69	.42
Logarithmic	.01	78	.82	.37
Quadratic	.01	77	.40	.67
Power	.04	78	3.34	.07
Exponential	.04	78	3.62	.06

4.2.4.3 Hybrid1 Group

Figure 4-31 shows the estimated learning trends for the proportion correct answers of the Japanese vocabulary test for the Hybrid1 practice group. The linear, logarithmic, and quadratic models are significant (p < .05) to the estimated learning trend, and shown in table 4-38.



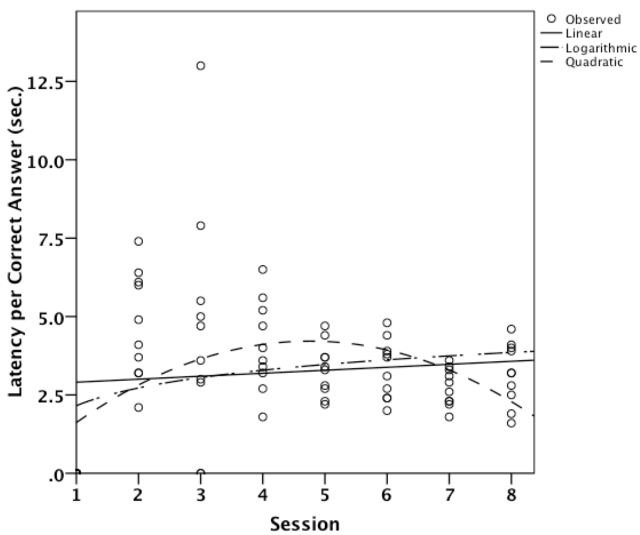
Linear: $y = 0.074x + 0.04,\ R^2 = 0.37$
Logarithmic: $y = 0.271\ln(x) + 0.016,\ R^2 = 0.40$
Quadratic: $y = -0.01x^2 + 0.165x - 0.111,\ R^2 = 0.39$

Figure **4-31**: Learning trends for the proportion correct answers of the Japanese vocabulary test for the Hybrid1 practice group (N=10).

Table **4-38**: Statistical output of curve estimation for the proportion correct answers of the Japanese vocabulary test for the Hybrid1 practice group.

Model	R^2	df	F	p
Linear	.37	78	44.79	.00
Logarithmic	.40	78	51.59	.00
Quadratic	.39	77	24.78	.00

Figure 4-32 shows the estimated learning trends for the latency per correct answer of the Japanese vocabulary test for the Hybrid1 practice group. The logarithmic and quadratic models are significant ($p < .05$) to the estimated learning trend, however, the linear model is not significant ($p < .05$), and shown in table 4-39.



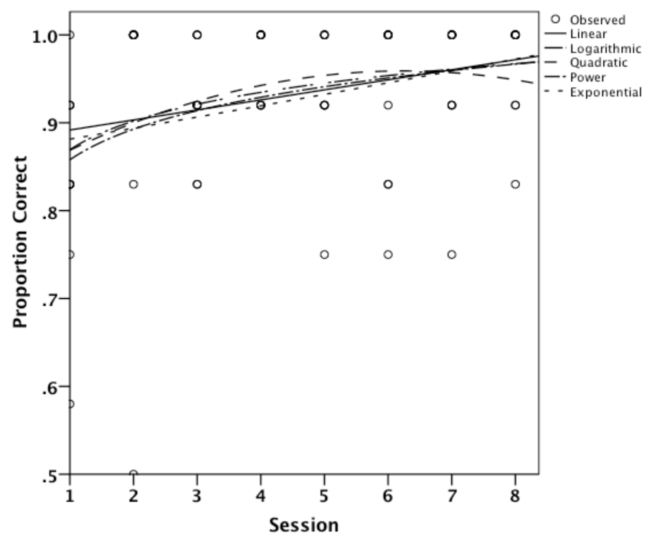
Linear: $y = 0.095x + 2.811,\ R^2 = 0.01$
Logarithmic: $y = 0.816\ln(x) + 2.157,\ R^2 = 0.07$
Quadratic: $y = -0.184x^2 + 1.749x + 0.054,\ R^2 = 0.17$

Figure **4-32**: Learning trends for the latency per correct answer of the Japanese vocabulary test for the Hybrid1 practice group (N=10).

Table **4-39**: Statistical output of curve estimation for the latency per correct answer of the Japanese vocabulary test for the Hybrid1 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.01	78	.84	.36
Logarithmic	.07	78	5.41	.02
Quadratic	.17	77	7.89	.00

Figure 4-33 shows the estimated learning trends for the proportion correct answer of the Permutation problem-solving task for the Hybrid1 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant (p < .05) to the estimated learning trend and shown in table 4-40.



Linear: $y = 0.011x + 0.88,\ R^2 = 0.07$

Logarithmic: $y = 0.047\ln(x) + 0.869,\ R^2 = 0.10$

Quadratic: $y = -0.003x^2 + 0.041x + 0.831\ \ R^2 = 0.01$

Power: $y = x^{0.058} + 0.858,\ R^2 = 0.10$

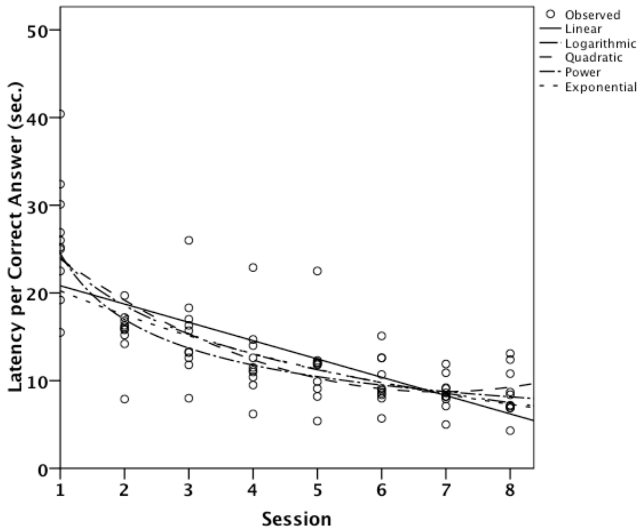
Exponential: $y = 0.869e^{0.014x},\ R^2 = 0.07$

Figure **4-33**: Learning trends for the proportion correct answers of the Permutation problem-solving task for the Hybrid1 practice group (N=10).

Table **4-40**: Statistical output of curve estimation for the proportion correct answer of the Permutation problem-solving task for the distributed practice group.

Model	R^2	*df*	*F*	*p*
Linear	.07	78	6.25	.01
Logarithmic	.10	78	9.03	.00
Quadratic	.01	77	4.21	.02
Power	.10	78	8.72	.00
Exponential	.07	78	6.12	.02

Figure 4-34 shows the estimated learning trends for the latency per correct answer of the Permutation problem-solving task for the Hybrid1 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-41.



Linear: $y = -2.087x + 22.908,\ R^2 = 0.51$

Logarithmic: $y = -7.99\ln(x) + 24.108,\ R^2 = 0.41$

Quadratic: $y = 0.436x^2 - 6.011x + 29.448,\ R^2 = 0.59$

Power: $y = x^{-0.529} + 24.517,\ R^2 = 0.59$

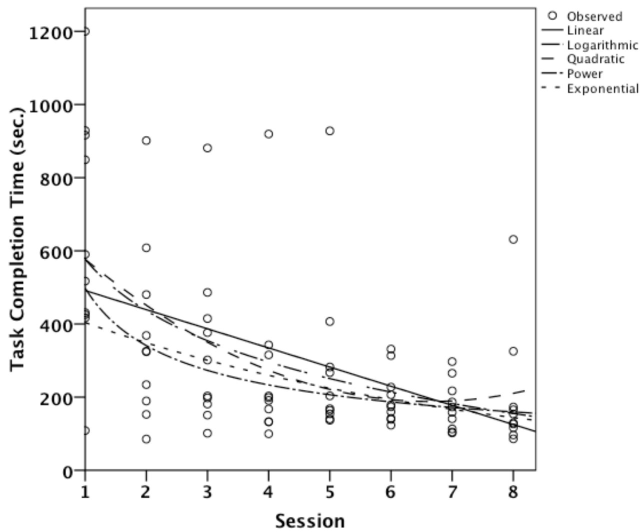
Exponential: $y = 23.408e^{-0.146x},\ R^2 = 0.54$

Figure **4-34**: Learning trends for the latency per correct answer of the Permutation problem-solving task for the Hybrid1 practice group (N=10).

Table **4-41**: Statistical output of curve estimation for the latency per correct answer of the Permutation problem-solving task for the Hybrid1 practice group.

Model	R^2	df	F	p
Linear	.51	78	79.71	.00
Logarithmic	.61	78	122.44	.00
Quadratic	.59	77	56.24	.00
Power	.59	78	113.57	.00
Exponential	.54	78	91.13	.00

Figure 4-35 shows the estimated learning trends for the task completion time of the Tower of Hanoi puzzle for the Hybrid1 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-42.



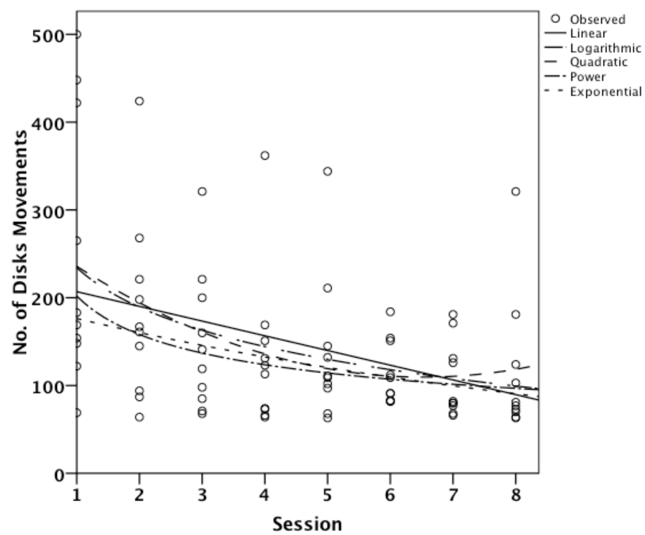
Linear: $y = -52.323x + 543.729,\ R^2 = 0.23$
Logarithmic: $y = -202.793\ln(x) + 577.093,\ R^2 = 0.29$
Quadratic: $y = 12.2x^2 - 162.127x + 726.736,\ R^2 = 0.29$
Power: $y = x^{-0.547} + 498.035,\ R^2 = 0.29$
Exponential: $y = 467.919e^{-0.147x},\ R^2 = 0.26$

Figure **4-35**: Learning trends for the task completion time of the Tower of Hanoi Puzzle for the Hybrid1 practice group (N=10).

Table **4-42**: Statistical output of curve estimation for the task completion time of the Tower of Hanoi puzzle for the Hybrid1 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.23	78	23.88	.00
Logarithmic	.29	78	31.91	.00
Quadratic	.29	77	15.37	.00
Power	.29	78	32.34	.00
Exponential	.26	78	27.06	.00

Figure 4-36 shows the estimated learning trends for the number of disk movements of the Tower of Hanoi puzzle for the Hybrid1 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-43.



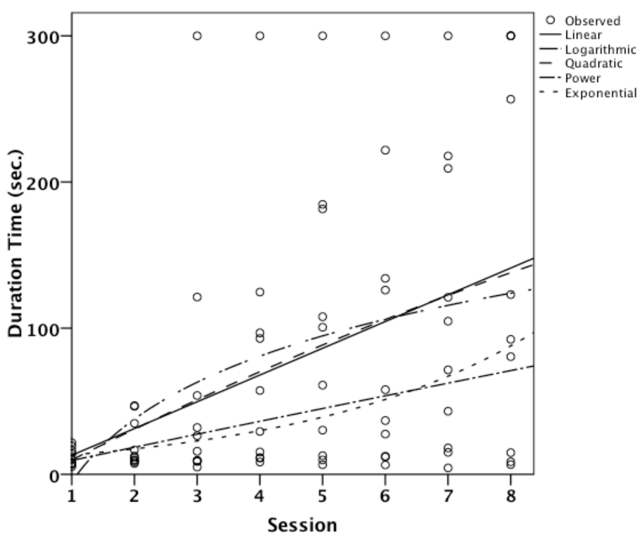
Linear: $y = -16.756x + 223.639,\ R^2 = 0.16$
Logarithmic: $y = -64.854\ln(x) + 234.207,\ R^2 = 0.20$
Quadratic: $y = 4.132x^2 - 53.94x + 285.612,\ R^2 = 0.20$
Power: $y = x^{-0.354} + 201.999,\ R^2 = 0.19$
Exponential: $y = 193.835e^{-0.095x},\ R^2 = 0.17$

Figure **4-36**: Learning trends for the number of disk movements of the Tower of Hanoi Puzzle for the Hybrid1 practice group (N=10).

Table **4-43**: Statistical output of curve estimation for the number of disk movements of the Tower of Hanoi puzzle for the Hybrid1 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.16	78	14.65	.00
Logarithmic	.20	78	18.94	.00
Quadratic	.20	77	9.42	.00
Power	.19	78	18.19	.00
Exponential	.17	78	15.47	.00

Figure 4-37 shows the estimated learning trends for the duration time of the Inverted Pendulum task for the Hybrid1 practice group. The linear, logarithmic, quadratic, power and exponential models are significant ($p < .05$) to the estimated learning trends, shown in table 4-44.



Linear: $y = 18.3x - 5.243,\ R^2 = 0.20$

Logarithmic: $y = 62.105\ln(x) - 5.219,\ R^2 = 0.19$

Quadratic: $y = -0.452x^2 + 22.366x - 12.021,\ R^2 = 0.20$

Power: $y = x^{0.968} + 9.487,\ R^2 = 0.23$

Exponential: $y = 10.183e^{0.269x},\ R^2 = 0.22$

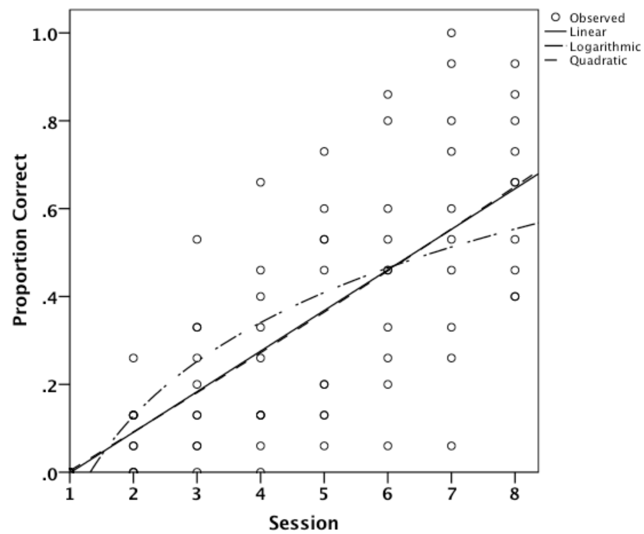
Figure **4-37**: Learning trends for the duration time of the Inverted Pendulum task for the Hybrid1 practice group (N=10).

Table **4-44**: Statistical output of curve estimation for the duration time of the Inverted Pendulum task for the Hybrid1 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.20	78	19.13	.00
Logarithmic	.19	78	17.95	.00
Quadratic	.20	77	9.47	.60
Power	.23	78	23.75	.00
Exponential	.22	78	21.91	.00

4.2.4.4 Hybrid2 Group

Figure 4-38 shows the estimated learning trends for the proportion correct answers of the Japanese vocabulary test for the Hybrid2 practice group. The linear, logarithmic, and quadratic models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-45.



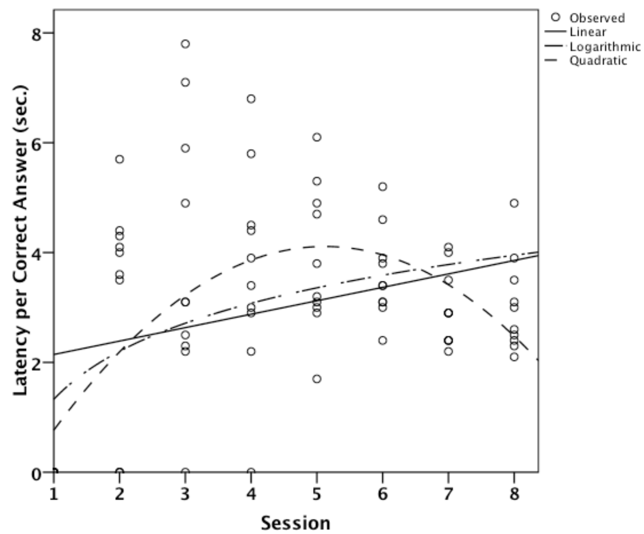
Linear: $y = 0.092x - 0.094,\ R^2 = 0.55$
Logarithmic: $y = 0.307\ln(x) - 0.086,\ R^2 = 0.50$
Quadratic: $y = 0.001x^2 + 0.086x - 0.083,\ R^2 = 0.55$

Figure **4-38**: Learning trends for the proportion correct answers of the Japanese vocabulary test for the Hybrid2 practice group (N=10).

Table **4-45**: Statistical output of curve estimation for the proportion correct answers of the Japanese vocabulary test for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.55	78	95.34	.00
Logarithmic	.50	78	78.70	.00
Quadratic	.55	77	47.09	.00

Figure 4-39 shows the estimated learning trends for the latency per correct answer of the Japanese vocabulary test for the Hybrid2 practice group. The linear, logarithmic and quadratic models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-46.



Linear: $y = 0.245x + 1.9,\ R^2 = 0.09$

Logarithmic: $y = 1.261\ln(x) + 1.328,\ R^2 = 0.20$

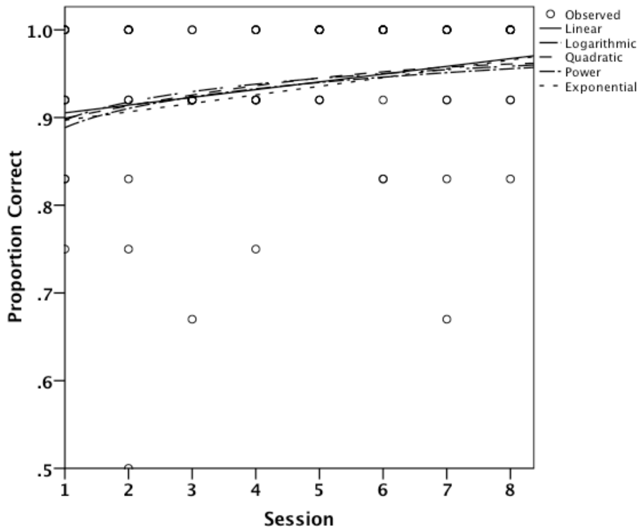
Quadratic: $y = -0.197x^2 + 2.018x - 1.056,\ R^2 = 0.33$

Figure **4-39**: Learning trends for the latency per correct answer of the Japanese vocabulary test for the Hybrid2 practice group (N=10).

Table **4-46**: Statistical output of curve estimation for the latency per correct answer of the Japanese vocabulary test for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.09	78	7.86	.00
Logarithmic	.20	78	19.59	.00
Quadratic	.33	77	18.89	.00

Figure 4-40 shows the estimated learning trends for the proportion correct answer of the Permutation problem-solving task for the Hybrid2 practice group. The linear model is marginally significant ($p < .05$) to the estimated learning trend; however, logarithmic, quadratic, power, and exponential models are not significant ($p < .05$), and shown in table 4-47.



Linear: $y = 0.009x + 0.897,\ R^2 = 0.05$

Logarithmic: $y = 0.03\ln(x) + 0.897,\ R^2 = 0.04$

Quadratic: $y = -0.001x^2 + 0.017x + 0.883\ R^2 = 0.05$

Power: $y = x^{0.035} + 0.888,\ R^2 = 0.04$

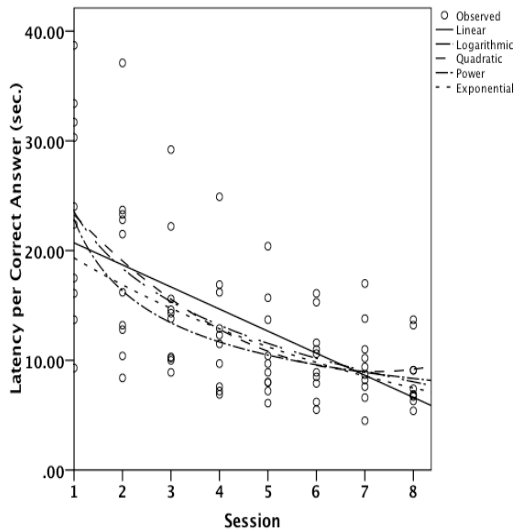
Exponential: $y = 0.888e^{0.011x},\ R^2 = 0.04$

Figure **4-40**: Learning trends for the proportion correct answers of the Permutation problem-solving task for the Hybrid2 practice group (*N*=10).

Table **4-47**: Statistical output of curve estimation for the proportion correct answer of the Permutation problem-solving task for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*P*
Linear	.05	78	3.88	.05
Logarithmic	.04	78	3.56	.06
Quadratic	.05	77	1.99	.14
Power	.04	78	3.28	.07
Exponential	.04	78	3.60	.06

Figure 4-41 shows the estimated learning trends for the latency per correct answer of the Permutation problem-solving task for the Hybrid2 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, and shown in table 4-48.



Linear: $y = -2.012x + 22.716,\ R^2 = 0.37$

Logarithmic: $y = -7.457\ln(x) + 23.547,\ R^2 = 0.42$

Quadratic: $y = 0.366x^2 - 5.308x + 28.209,\ R^2 = 0.42$

Power: $y = x^{-0.484} + 22.846,\ R^2 = 0.42$

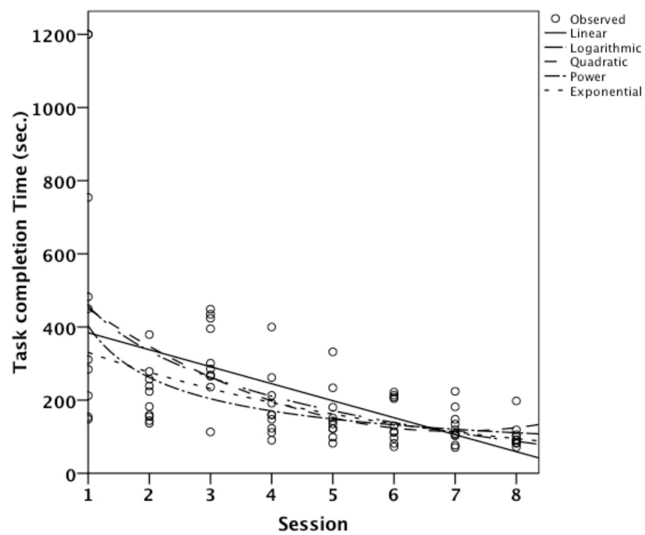
Exponential: $y = 22.133e^{-0.136x},\ R^2 = 0.40$

Figure **4-41**: Learning trends for the latency per correct answer of the Permutation problem-solving task for the Hybrid2 practice group (N=10).

Table **4-48**: Statistical output of curve estimation for the latency per correct answer of the Permutation problem-solving task for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.37	78	46.39	.00
Logarithmic	.42	78	57.06	.00
Quadratic	.42	77	28.15	.00
Power	.42	78	56.82	.00
Exponential	.40	78	52.17	.00

Figure 4-42 shows the estimated learning trends for the task completion time of the Tower of Hanoi puzzle for the Hybrid2 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-49.



Linear: $y = -46.456x + 430.709,\ R^2 = 0.29$

Logarithmic: $y = -177.963\ln(x) + 457.562,\ R^2 = 0.35$

Quadratic: $y = 9.405x^2 - 131.103x + 571.788,\ R^2 = 0.34$

Power: $y = x^{-0.621} + 403.308,\ R^2 = 0.44$

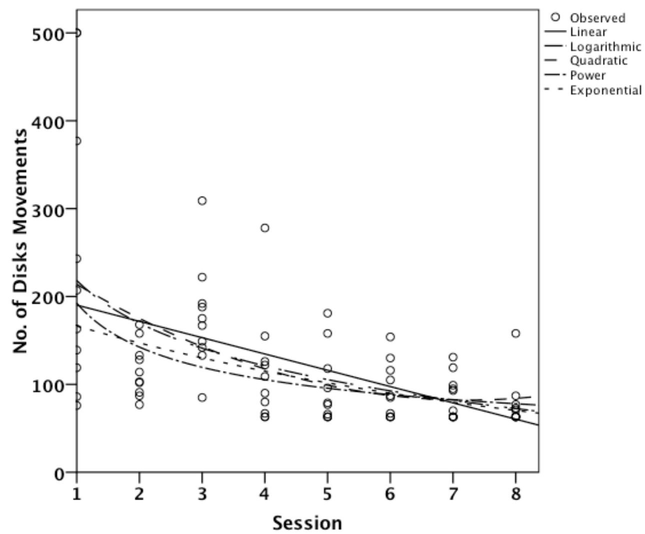
Exponential: $y = 395.021e^{-0.178x},\ R^2 = 0.44$

Figure **4-42**: Learning trends for the task completion time of the Tower of Hanoi Puzzle for the Hybrid2 practice group (N=10).

Table **4-49**: Statistical output of curve estimation for the task completion time of the Tower of Hanoi puzzle for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.29	78	32.03	.00
Logarithmic	.35	78	42.43	.00
Quadratic	.34	77	19.73	.00
Power	.44	78	62.29	.00
Exponential	.44	78	62.28	.00

Figure 4-43 shows the estimated learning trends for the number of disk movements of the Tower of Hanoi puzzle for the Hybrid2 practice group. The linear, logarithmic, quadratic, power, and exponential models are significant ($p < .05$) to the estimated learning trend, shown in table 4-50.



Linear: $y = -18.55x + 208.775,\ R^2 = 0.25$

Logarithmic: $y = -70.1\ln(x) + 218.223,\ R^2 = 0.30$

Quadratic: $y = 3.389x^2 - 49.054x + 259.614,\ R^2 = 0.29$

Power: $y = x^{-0.435} + 192.505,\ R^2 = 0.33$

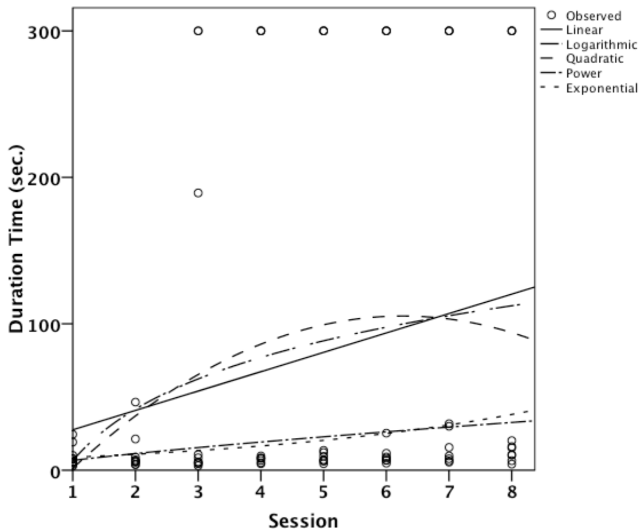
Exponential: $y = 188.140e^{-0.123x},\ R^2 = 0.32$

Figure **4-43**: Learning trends for the number of disk movements of the Tower of Hanoi Puzzle for the Hybrid2 practice group (N=10).

Table **4-50**: Statistical output of curve estimation for the number of disk movements of the Tower of Hanoi puzzle for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.25	78	26.38	.00
Logarithmic	.30	78	33.05	.00
Quadratic	.29	77	15.46	.00
Power	.33	78	38.04	.00
Exponential	.32	78	36.37	.00

Figure 4-44 shows the estimated learning trends for the duration time of the BalanceMe Game for the Hybrid2 practice group. The linear, logarithmic, quadratic, power and exponential models are significant ($p < .05$) to the estimated learning trends, shown in table 4-51.



Linear: $y = 13.234x + 14.371,\ R^2 = 0.07$
Logarithmic: $y = 50.927\ln(x) + 6.416,\ R^2 = 0.08$
Quadratic: $y = -3.755x^2 + 47.026x - 41.95,\ R^2 = 0.09$
Power: $y = x^{0.76} + 6.716,\ R^2 = 0.10$
Exponential: $y = 7.215e^{0.208x},\ R^2 = 0.09$

Figure **4-44**: Learning trends for the duration time of the Inverted Pendulum task for the Hybrid2 practice group (N=10).

Table **4-51**: Statistical output of curve estimation for the duration time of the Inverted Pendulum task for the Hybrid2 practice group.

Model	R^2	*df*	*F*	*p*
Linear	.07	78	5.39	.02
Logarithmic	.08	78	6.68	.01
Quadratic	.09	77	3.59	.03
Power	.10	78	8.70	.00
Exponential	.09	78	7.82	.00

4.2.5 Summary of Results

All of the participants (N=40) were divided into four groups and performed four tasks according to the training schedule of each group. There were eight learning sessions and one-retention test for each schedule. Participants showed improved performance with practice and knowledge degradation at the retention test. The average performances of all tasks with respect to the practice schedule were analyzed. All of the tasks except the Japanese vocabulary test showed little knowledge degradation at the retention test. These results indicate that the tasks that mainly depend on the procedural knowledge or perceptual-motor skill did not be influenced by the memory decay, because the required knowledge for these tasks might be proceduralized in the memory of participants during the entire learning sessions. Thus, the rest period between the last learning session and the retention test could not influence to the memory decay for these tasks. These results also show that the decay of declarative memory knowledge (Japanese vocabularies in my experiment) may be more strongly influenced by rest period compared to the other knowledge types.

To examine whether the participants in each group were randomly divided or not, the performance test at the first learning session was conducted using the analysis of variance. The results showed that there was no significant difference among groups in all tasks, so I could assume the participants in this experiment could not influence the performance during the whole learning sessions and the retention test.

I also analyzed the improved performance rate that is the performance differences between the last learning session and the retention test (see equation 4.1). The improved performance rates of each group with respect to each task were compared through analyses of variance and t-tests. For the Japanese vocabulary test, there was no significant difference among the groups. These results were somewhat different from the previous studies that showed the distributed practice schedule is better than the massed practice schedule at a retention test. So, I conducted analysis of variance again without the outlier in the massed practice group. The results still showed there was no significant difference among the groups. To compare between the two groups, I also analyzed the results using t-tests. The results of t-test showed that there was significant difference between the distributed and massed groups in accuracy, $t(18) = -2.227$, $*p < .05$, however, there was no significant different between the other groups.

The Permutation problem-solving task and Tower of Hanoi puzzle resulted in no significant difference among the groups. There results indicate that the different practice schedules may not be influenced to the performance for solving these kinds of tasks.

The Inverted Pendulum task, however, showed significant different among groups on duration time, $F(3,36) = 3.348$, $*p < .05$. The t-tests resulted in significant difference between the distributed and the Hybrid1 groups, $t(18) = 02.566$, $*p < .05$. These results showed that the task that requires perceptual-motor skill produce better learning and retention performance with a massed or somewhat early massed way rather than a distributed way.

To see the learning trends for each group with respect to the tasks, a series of learning graphs are presented. By doing this estimation, I can figure out the learning trends of each training schedule and each task, thus it enables us to estimate the performance of further learning sessions. From the results of these estimation, I found that some of the graphs follow the Power law of learning (Ritter & Schooler, 2001), and others has a linear relationship.

Chapter 5

ACT-R model

In this chapter, I provide an overview of the ACT-R architecture that I used to make cognitive models for representing the human behavior in my experiments. The explanation of ACT-R models and their considerations are also provided.

5.1 Overview of the ACT-R Architecture

ACT-R (Anderson, 2007; Anderson et al., 2004; Anderson & Lebiere, 1998; Anderson, Matessa, & Lebiere, 1997), which stands for Adaptive Character of Thought – Rational, is a cognitive architecture that contains theories about how human cognition works. The history of ACT-R that is presented in figure 5-1, can be found in a book, *How can the human mind occur in the physical universe?* (Anderson, 2007). It started from Human Associative Memory (Anderson & Bower, 1973) that is a founding theory of human's declarative memory that includes spreading activation and retrieval of information from memory. Anderson developed the ACT theory (Anderson, 1976) by combining the declarative system of HAM theory and Newell's symbolic representation system. He also proposed the subsymbolic system by extending symbolic representation. The ACT* system that contained the modern ACT-R theory was proposed, and it included goal-oriented processing (goal module) and production learning mechanism (production compilation). ACT-R came into being in 1993 with Lisp programming language. The "R" denoted the influence of rational analysis (Anderson, 1993). Furthermore, ACT-R included perceptual and motor capabilities, inspired from the EPIC architecture (Meyer & Kieras, 1997), to represent the perceptual-motor component of human (Byrne & Anderson, 1998), and the perceptual-motor module enabled ACT-R models to represent tasks that require human's perceptual-motor components. The current version of ACT-R is continuously evolving and expanding its research areas to the brain imaging with fMRI (Qin et al., 2004). It can be a candidate unified theory of cognition (Newell, 1990), by validating human's cognition process (e.g. learning and forgetting) and human behaviors in various tasks (Pew & Mavor, 1998).

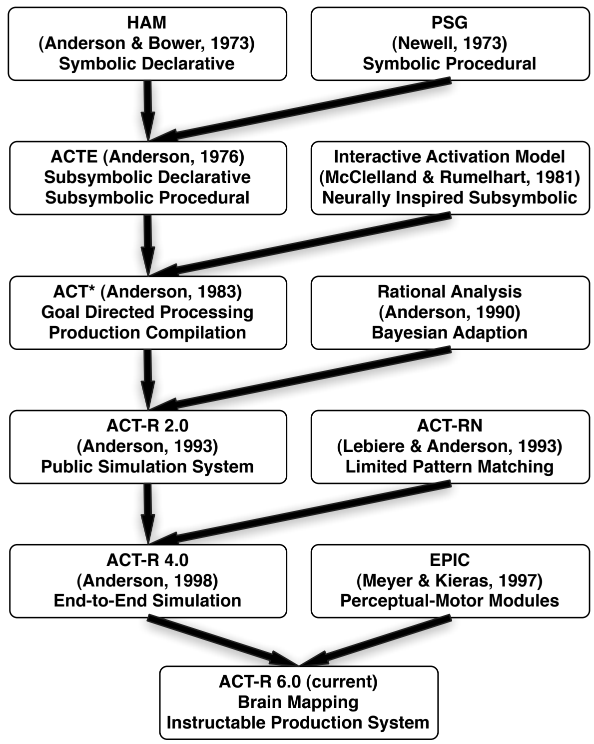


Figure **5-1**: The history of ACT-R, taken from Anderson (2007).

Figure 5-2 shows a schematic view of the ACT-R architecture. ACT-R consists of seven modules; goal module, declarative module, imaginal module, motor module, visual modules, aural module and vocal module, and one production system. Each module has each corresponding buffer that enables a model to make a request from the production system and to hold a chunk as a result of such a request.

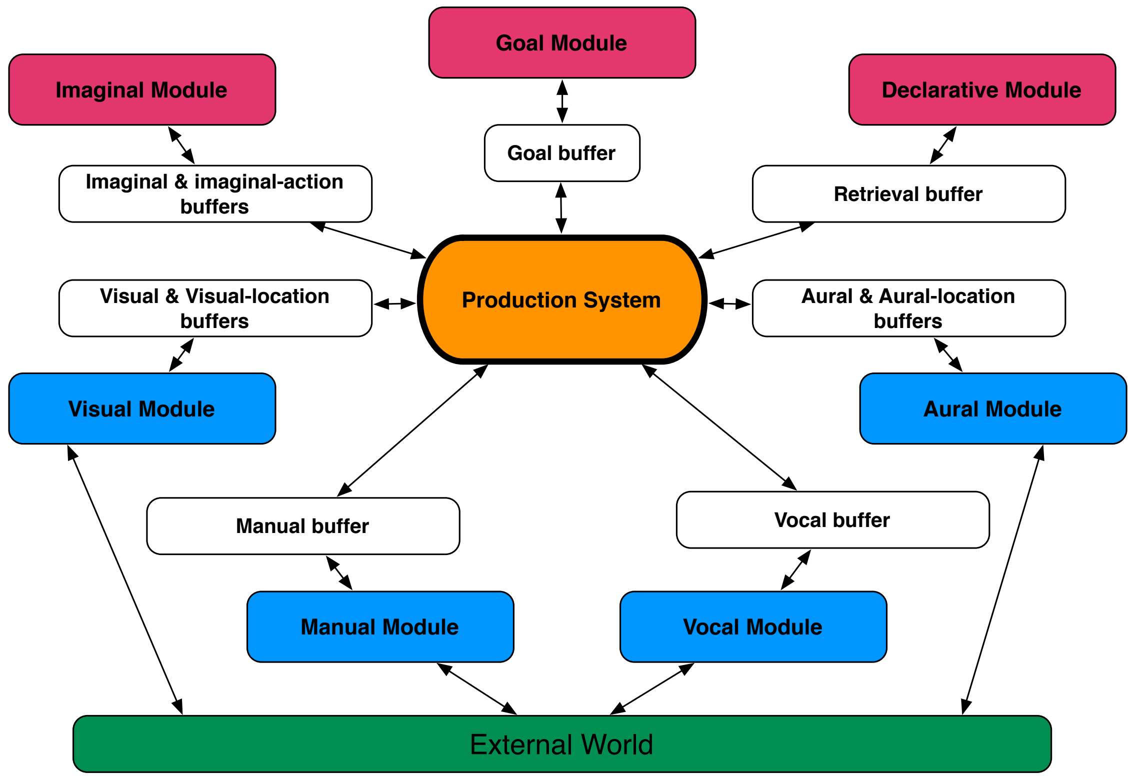


Figure **5-2**: A schematic view of the ACT-R architecture, taken from Anderson (2007) and Byrne (2001).

The production system that is a rule-based system communicates with buffers of each module, so it could generate and coordinate behaviors according to the functions of each module. It represents cognition as a sequence of recognize-decide-act cycles (Gunzelmann, Gross, Gluck, & Dinges, 2009). On each cycle, all productions are compared to the current state that is represented by the contents of buffers. Among the productions that match the current statue, one production is selected by the utility function and then fired. The results of this process might modify the current state. Finally, it begins to recognize again for next cycle.

The goal module enables a model to maintain the current task state and to hold relevant information for the current task. It also serves as a source of activation for retrievals by default.

The declarative module represents the process of retrieving facts from human memory. It stores chunks that are generated by models. The retrieval process of chunks from the declarative memory depends on many factors that affect the accuracy and speed with which a chunk can be retrieved, and this process is based on research of human memory performance.

The imaginal module has two kinds of buffers, one is an imaginal buffer, and the other is imaginal-action buffer. The imaginal buffer typically is used to create and hold task relevant information. It operates similar to the goal buffer, but there is a time cost (0.2 sec) to create and manipulate the chunks.

The rest of four modules belong to the perceptual-motor modules in ACT-R (Byrne, 2001; Byrne & Anderson, 1998). They enable models to interact with an external world. ACT-R models could attend to visual and aural stimuli using visual and aural modules (perceptual). Motor modules (manual and vocal) enable ACT-R models to send outputs to the world. In my thesis, I describe visual and motor modules, because I only used these two modules to develop models.

The visual module provides a model with information about what can be seen in the current environment. It has two subsystems, a “where” system and a “what” system. Each system has its own buffer and accepts specific requests from the production system. The “where” system generally provides the location information according to conditions, such as x-coordinate, y-coordinate, color, etc. The production system could request these conditions to let a model to see. The “what” system takes request through the visual buffer. It stores contents in chunks according to the location which have been found using the where system.

The motor module that is based on EPIC’s Manual Motor Processor (Kieras & Meyer, 1996) serves as a model’s hands. It enables a model to operate a virtual keyboard and mouse with two-dimensional layout (Figure 5-3).

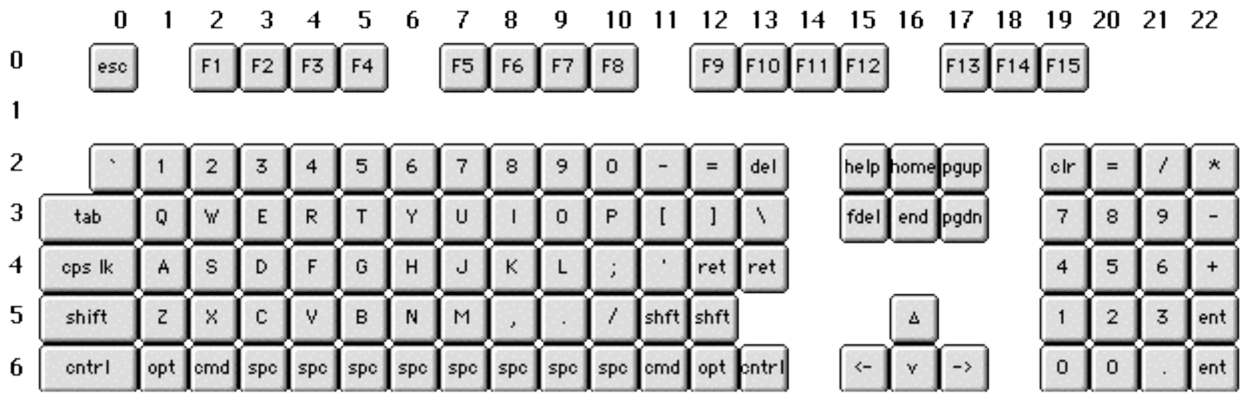


Figure **5-3**: A virtual keyboard in ACT-R (taken from ACT-R 6.0 manual).

For the assumption of a virtual mouse, it has one button and is controlled by a model's right hand. Its location is (28 2) relative to the virtual keyboard layout. The starting positions of virtual keyboard are F, (4 4) for the left hand, and J, (7 4) for the right hand. Before using a keyboard and mouse, model's hands should be located its default position using (hand-to-home) or (hand-to-mouse) requests. Further information for the ACT-R modules are provided by ACT-R's official website at http://act-r.psy.cmu.edu.

5.1.1 Declarative Memory Learning and Forgetting in ACT-R

In this section, I present ACT-R's declarative memory learning and forgetting mechanisms. The declarative memory learning mechanism is based on base-level learning equation that has a decay value to represent a forgetting mechanism. The features of the equations for declarative memory learning and forgetting consist of sub-symbolic construct of the ACT-R architecture.

5.1.1.1 Activation of Chunks and Base-Level Learning

Every chunk in ACT-R's declarative memory has associated with it a numerical value that is called activation. The activation of a chunk *i* is represented as A_i. It consists of the base-level learning, and a noise component.

$$A_i = B_i + \varepsilon \qquad \text{Equation 5.1}$$

The base-level activation for a chunk *i* is:

$$B_i = \ln(\sum_{j=1}^{n} t_j^{-d}) \qquad \text{Equation 5.2}$$

n: the number of presentations for chunk *i*
t_j: the time since the *jth* presentation
d: the decay parameter which is set using the :bll (base-level learning) parameter

The base-level equation describes a process in which each time an item is presented there is an increase in its base-level activation, which decays away as a power function of time since that presentation.

5.1.1.2 Optimized Learning

To calculate the base-level activation of each chunk is computationally expensive as we can see in the equation 5.2. To reduce this computational cost, ACT-R provides an optimized learning equation that assumes the presentations follow uniform distribution over the time since the item was created.

$$B_i = \ln(\frac{n}{1-d}) - d \times \ln(L) \qquad \text{Equation 5.3}$$

n: the number of presentations for chunk *i*
L: the lifetime of chunk *i* (the time since its creation)
d: the decay parameter

5.1.1.3 Recall Probability

When a model makes a retrieval request and a matching chunk exists, that chunk will be retrieved only if its activation value exceeds the retrieval threshold, τ. The probability of retrieval is represented by the recall probability in equation 5.4.

$$P(A_i) = \frac{1}{1 + e^{\frac{\tau - A_i}{s}}} \qquad \text{Equation 5.4}$$

From equation 5.4, the activation of chunk, A_i, tends higher, the recall probability approaches 1, whereas, as τ tends higher, the probability decreases. When τ equals A_i, the probability is 0.5. The noise parameter, *s*, controls the sensitivity of recall to changes in activation.

5.1.1.4 Retrieval Latency

The activation of chunk determines retrieval latency. When a model requests retrieval of a chunk, the time taken until the chunk is available in the retrieval buffer is calculated using equation 5.5.

$$Time = Fe^{-A_i} \qquad \text{Equation 5.5}$$

A: the activation of the chunk that is retrieved
F: the latency factor parameter

5.1.2 Procedural Memory Learning and Forgetting in ACT-R

ACT-R has procedural memory learning mechanisms that include production compilation (Taatgen & Lee, 2003) and utility learning. However, there is no mechanism that predicts forgetting for procedural memory in the ACT-R architecture.

5.1.2.1 Production Compilation

Production compilation (Taatgen & Lee, 2003) is a mechanism that represents procedural learning in ACT-R by collapsing two productions into a single production. To determine whether two productions can be combined into one production, compatible usage between the two productions for all buffers is checked. Compatible usage depends on the "compilation type" of the buffer, and there are four kinds of compilation type in ACT-R. Table 5-1 shows compilation types for each buffer.

The motor style buffers, such as manual and vocal buffers, never hold a chunk. To avoid jamming, the production compilation process rarely occurs between these buffers. For example, if the first production makes a request to a motor buffer in the action side, then to compose it with a second production that also makes request of that buffer is impossible.

Table **5-1**: Compilation types with respect to buffers in ACT-R.

Buffer Name	Compilation Type
goal	Goal
imaginal	Goal
retrieval	Retrieval
aural	Perceptual
aural-location	Perceptual
visual	Perceptual
visual-location	Perceptual
manual	Motor
vocal	Motor

The perceptual style buffers that include aural, aural-location, visual, and visual-location hold chunks generated by their modules, and these modules could interact with the external world, that is, conditions and actions in productions could be changed by the external world. So, the production compilation process between these buffers rarely occurs.

The retrieval style buffer only has the retrieval buffer. Because the retrieval buffer is an internal buffer, it is easy to predict and thus gives chances to compile between productions. For example, if the action of the first production requests retrieval and the condition of the second production tests the outcome of this retrieval request. In this case, the request and test processes could be deleted, and these processes could be replaced by any variables in the retrieval process. Figure 5-4 shows the example of the production compilation process in retrieval style buffer.

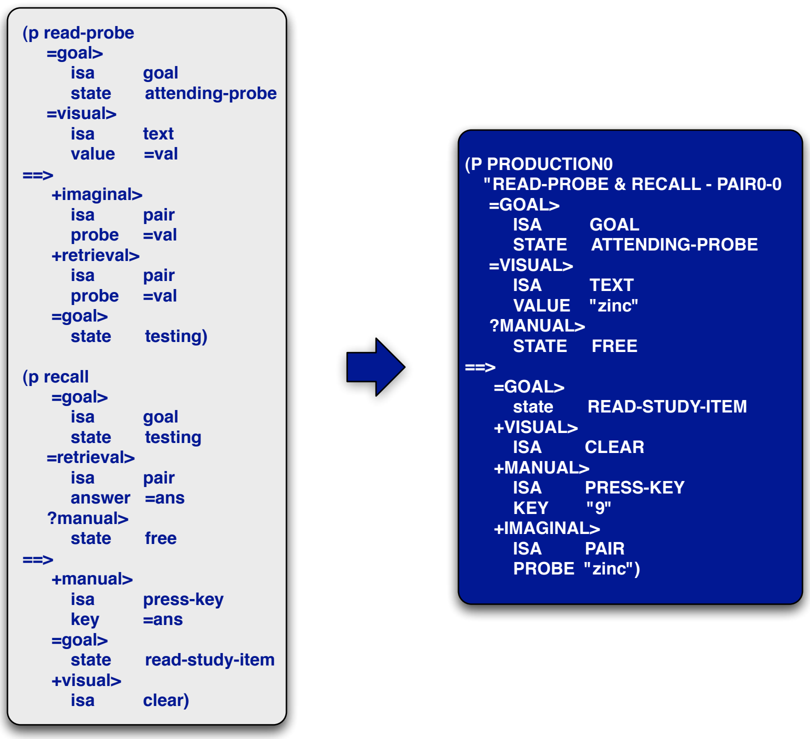


Figure **5-4**: An example of the production compilation process in the retrieval style buffer.

The goal and imaginal buffers are also internal buffers, so their production compilation processes are similar to the retrieval style buffer's process. When the first production does not make a request, two productions could be combined. However, when the first production makes a request, the production compilation could be happened only if the second production does not make a request.

5.1.2.2 Utility Learning

In this section, I present the process of how the created rules are selected. Every production in ACT-R has a utility value, and a new production that is created by production compilation also has its own value. When two productions, new production and old production, satisfy the specific condition, so both of them could apply, they are compared by their utility values. The old production has larger value than the new one at the first time, however, whenever the new one is recreated, its utility value is updated with a reward, and finally, the new one could be greater than the old one, so in this case new one could apply in that condition. Consequently, the new production that is created by production compilation stands for the learning process of procedural memory, and when the new production applied, it leads to decrease the overall task completion time, so it could show the learning effect for a task. Equation 5.6 shows the utility learning equation. If $U_i(n\text{-}1)$ is the utility of a production *i* after its $n\text{-}1^{st}$ application and $R_i(n)$ is the reward the production receives for its *nth* application, then its utility $U_i(n)$ after its *nth* application could be calculated by this equation.

$$U_i(n) = U_i(n-1) + \alpha[R_i(n) - U_i(n-1)] \qquad \text{Equation 5.6}$$

The α indicates the learning rated, and it could be changed by adjusting the :alpha parameter in the ACT-R architecture. According to this equation, the utility value of each production will be changed until it matched the average reward that the production receives.

5.2 The Model Predictions

The Japanese vocabulary test was modeled among the four tasks using the ACT-R 6.0 architecture. The current version of ACT-R supports two kinds of base-level learning equations, the original base-level learning equation (equation 3.2) and the revised based-level equation

(equation 3.5). So, two ACT-R models were developed based on the two equations with respect to the four practice schedules.

The experiment environment for the ACT-R models is different from the one for the human participants. As I described in figure 4-1, the experiment environment for the participants is a web-based experiment. However, ACT-R models could not interact with web browsers directly, so I developed the experiment environment for the ACT-R models by using the Lisp language and an ACT-R environment. Figure 5-5 shows the ACT-R models' experiment environment.

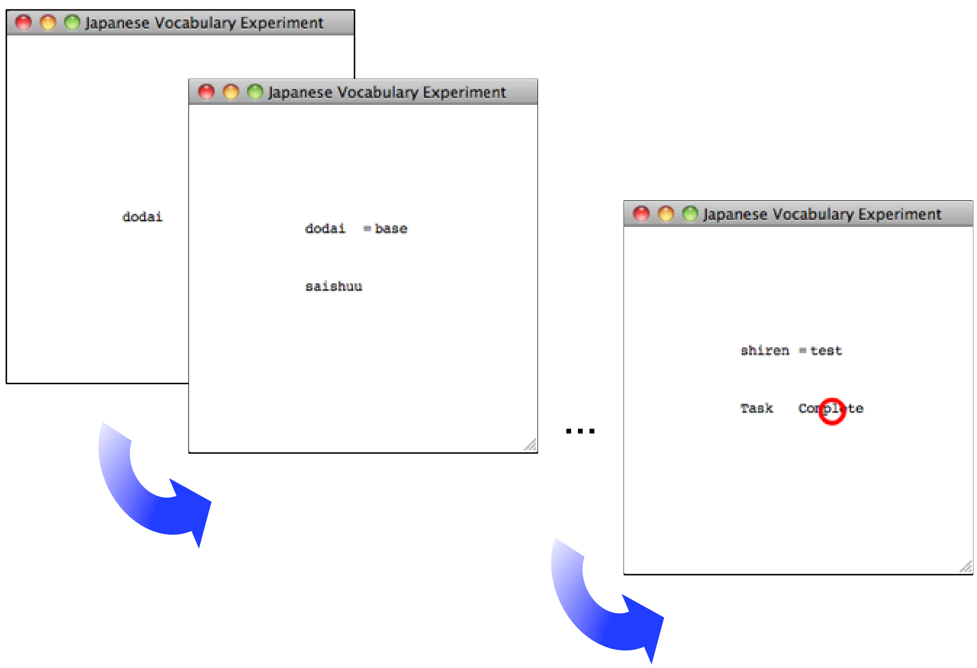


Figure **5-5**: Experiment environment for ACT-R models.

Like human participants, when a model sees a problem through its visual module, it tries to find an answer. If it succeeds to find and retrieve, it types what it retrieves with the enter key through its manual module to go on to the next page. If it fails to find or retrieve the answer, it just hits the enter key and the next problem is presented. Each page has the answer to the previous problem except for the first page, and a model could store these pairs of problems and answers using its imaginal buffer. Finally, when a model sees the “task complete” sign, it terminates the task.

The other consideration for making a model is a time factor. Anderson, Finchman, and Douglass introduced psychological time in modeling long term period of practice (1999). In their study, participants received as many as 240 trials of practice distributed over intervals as long as 400 days. To fit the human data with ACT-R models, they used and assumed a psychological time factor that indicates the psychological time slowed after the experiment session.

Pavlik and Anderson (2003, 2005) also used the psychological time factor when validating practice and forgetting effects on vocabulary memory. They introduced a scaling parameter, *h*, which is 0.025 in their recall equation. So, I used the same psychological factor when modeling Japanese vocabulary test models, therefore, the one-day rest period is 2,160 sec. in the model simulation.

I ran two models for 100 times (Ritter, Schoelles, Quigley, & Klein, in press) with respect to the four practice schedules, and the average proportion correct and latency were gathered. Figure 5-6 shows the proportion correct of the model based on the original base-level equation and the model based on the revised base-level equation for the massed practice schedule of the Japanese vocabulary test.

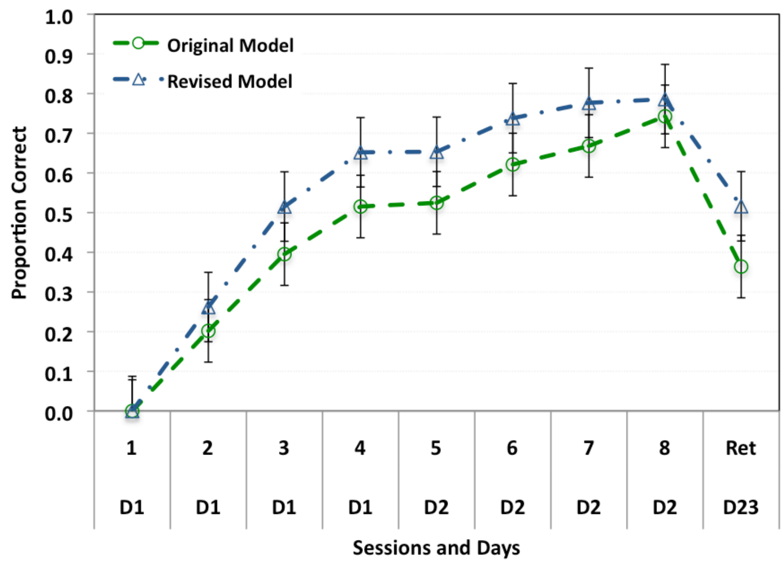


Figure **5-6**: The proportion correct of the original and revised models for the massed practice schedule of the Japanese vocabulary test.

The proportion correct increased with practice in both models during the entire learning sessions. However, There is no increase or little bit between increases between the fourth and fifth sessions, because between those sessions there was one-day rest interval. Like as human participants, memory decay occurs during the rest period. Both models also forget a lot between the last learning session and the retention session.

Figure 5-7 shows the proportion correct of the two models that follow the distributed practice schedule for the Japanese vocabulary test. As we see, there are big differences between the two models in performance prediction. The model based on the original equation shows less increasing between the sessions; however, the model based on the revised equation shows relatively sharp increase during the first four sessions. Both models show the memory decay between the fourth and fifth learning sessions and between the last learning session and the retention session.

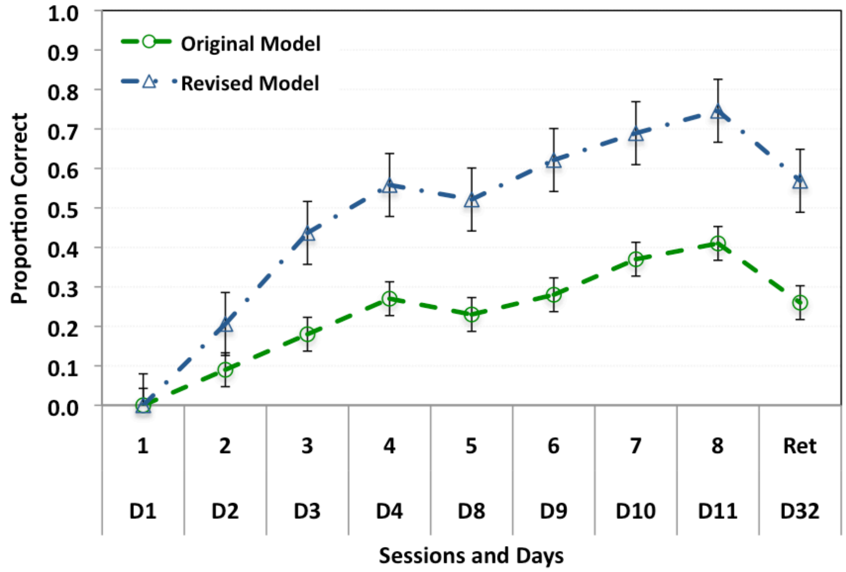


Figure **5-6**: The proportion correct of the original and revised models for the distributed practice schedule of the Japanese vocabulary test.

Figure 5-8 shows the proportion correct of the two models that follow the Hybrid1 practice schedule for the Japanese vocabulary test. The proportion correct of revised-equation based model shows relatively sharp increase during the first five sessions, however, the original-equation base model shows relatively less increasing during the entire learning sessions. Both models show memory decay between the last learning session and the retention session.

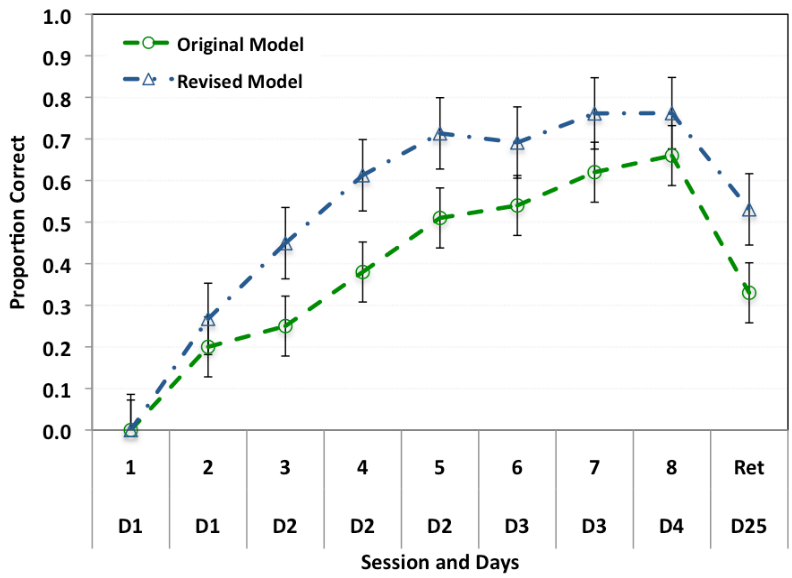


Figure **5-8**: The proportion correct of the original and revised models for the Hybrid1 practice schedule of the Japanese vocabulary test.

Figure 5-9 shows the proportion correct of the two models that follow the Hybrid2 practice schedule for the Japanese vocabulary test. The proportion correct of revised-equation base model shows relatively sharp increase during the first three sessions, and almost no increase between the third and the fourth session (4-day rest period). After the fourth session, the proportion correct increases again until the last learning session. However, the original-equation base model shows little increase during the first six sessions, and reaches around .4 at the last learning session. Both models show the memory decay between the last learning session and the retention session.

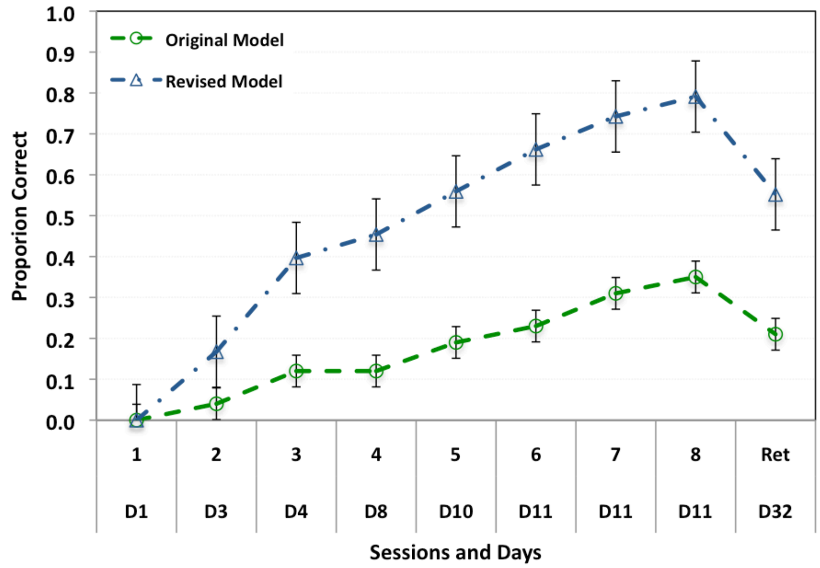


Figure **5-9**: The proportion correct of the original and revised models for the Hybrid2 practice schedule of the Japanese vocabulary test.

Figure 5-10 shows the results of the model based on the original equation for the Japanese vocabulary test with respect to the all practice schedule. The original-equation base model predicts the massed practice schedule is the most effective schedule among all the schedules. The Hybrid1 practice schedule ranks second place among the schedules. The distributed and the Hybrid2 practice schedules are almost identical performance at the last learning session and the retention session. Thus, we can assume that the original-equation base model cannot predict the spacing effects of human mind as I mentioned in chapter 3, so this model predicts massed or early massed practice schedule produce better performance than the distributed practice schedule.

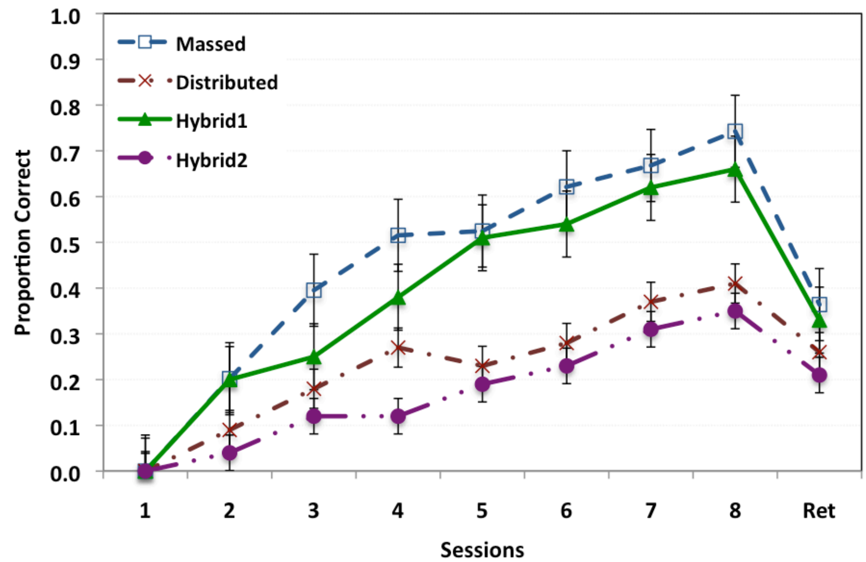


Figure **5-10**: The proportion correct of the model with the original base-level learning equation for the Japanese vocabulary test with respect to the four practice schedules.

Figure 5-11 shows the proportion correct of the model based on the revised base-level equation for the Japanese vocabulary test with respect to the all practice schedules. This model predicts the Hybrid2 practice schedule is the most effective schedule, and the Hybrid1 practice schedule is the least effective schedule in the retention session. However, the performances among the practice schedules are almost identical at the retention session. So, we may assume that the model based on the revised base-level learning equation may not be able to predict the performances of different schedules with long term period, although it can predict the spacing effects of human mind. To explore this, I present the comparison between the human data and the models prediction in the next chapter.

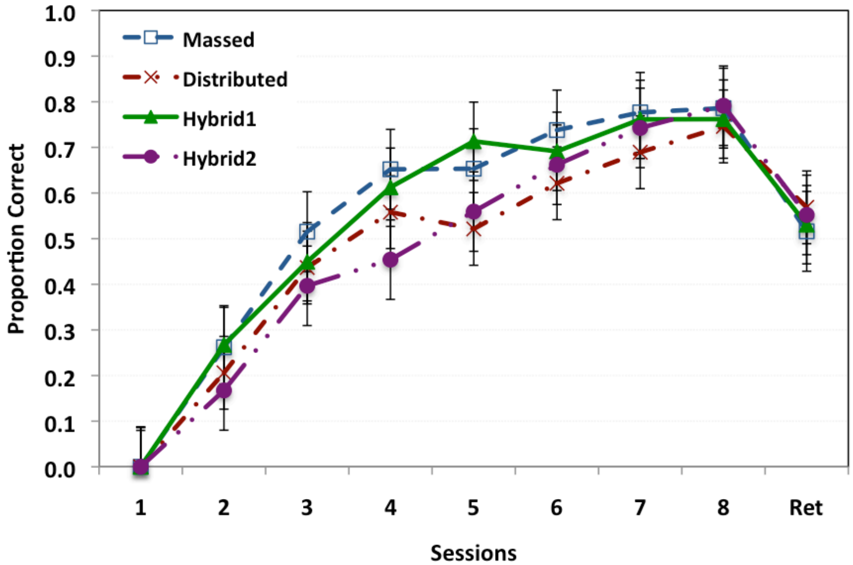


Figure **5-11**: The proportion correct of the model with the revised base-level learning equation for the Japanese vocabulary test with respect to the four practice schedules.

5.3 Summary of the Model

The two kinds of Japanese vocabulary test models, the model based on the original base-level learning and the model based on the revised base-level learning were developed by using ACT-R 6.0. Both models were run for 100 times with psychological time base, and the average proportion correct of each model with respect to the four practice schedules was gathered, and the models predictions were compared. The model based on the original base-level learning equation showed relatively slow learning effect during the entire learning sessions and low performance at the retention session, however, the one based on the revised base-level learning equation showed quick learning effect during the entire learning sessions and high performance at the retention session.

The results of the model based on the original base-level learning equation showed that the massed practice schedule is the most effective one among the four practice schedules at the retention session. It indicated that the original base-level learning equation could not predict the spacing effect of human mind, and this is the reason the needs of the revised base-level learning equation (Pavlik & Anderson, 2005).

The model based on the revised base-level learning equation showed that the Hybrid2 practice schedule is the most effective practice schedule among all the schedules at the retention session, and it also showed that the distributed practice schedule is better than the massed and the Hybrid1 practice schedules. It indicated that the revised base-level learning equation supports the spacing effect of human mind. However, we need to figure out that the model prediction fits with the human data for the model validation. In the next chapter, I present the comparison between the human data and the models.

Chapter 6

Comparison of the Human data with the ACT-R Models

I presented the empirical data in chapter 4, and the models prediction in chapter 5. In this chapter, I validate the models' predictions by comparing the human data to the models with respect to the four practice schedules.

6.1 Comparison of the Human Data with the Models

Figure 6-1 shows the proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test of the massed practice schedule.

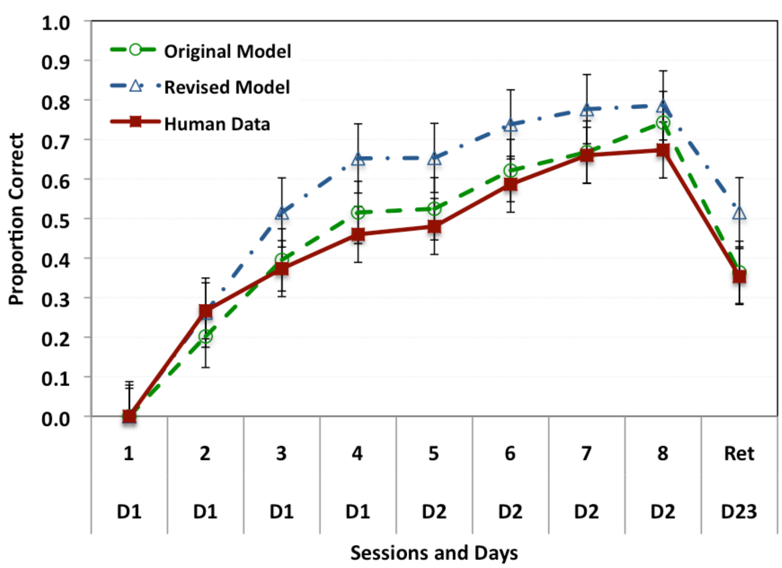


Figure **6-1**: The massed practice schedule: The proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test.

As we see in the graph, the original-equation base model predicts the human data more closely than the revised-equation base model in the massed practice schedule. However, the revised-equation base model also predicts the learning and forgetting trends of the human data for the massed practice schedule. To validate models with the human data, r^2 and RMSD (root mean square deviation) for both models were calculated and the results are presented in table 6-1.

Table **6-1**: The prediction results of the original-equation base model and the revised-equation base model for the Japanese vocabulary test of the massed practice schedule.

ACT-R Model	Model	R^2	RMSD
Original-equation base Model	Linear	.98	0.03
	Quadratic	.98	0.03
Revised-equation base Model	Linear	.96	0.04
	Quadratic	.97	0.04

The original-equation base model predicts the human data with R^2=0.98 and *RMSD*=0.03 through the linear and quadratic models. The revised-equation base model predicts the human data with R^2=0.96 and *RMSD*=0.04 through the linear model and R^2=0.97 and *RMSD*=0.04 through the quadratic model. Thus, both of the models predict the learning performance of the massed practice schedule very closely.

Figure 6-2 shows the proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test of the distributed practice schedule.

As we see the figure 6-2, the human data of the distributed practice schedule is located between the original-equation base model and the revised-equation base model. However, the revised one predicts the human data more closely than the original one. To validate models with the human data, r^2 and RMSD (root mean square deviation) for both models were calculated and the results are presented in table 6-2.

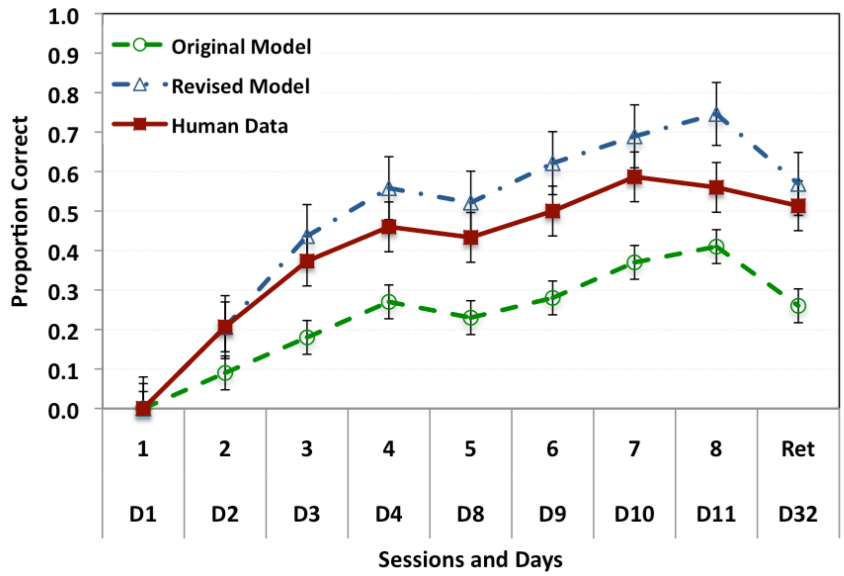


Figure **6-2**: The distributed practice schedule: The proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test

Table **6-2**: The prediction results of the original-equation base model and the revised-equation base model for the Japanese vocabulary test of the distributed practice schedule.

ACT-R Model	Model	R^2	RMSD
Original-equation base Model	Linear	.92	0.05
	Quadratic	.98	0.03
Revised-equation base Model	Linear	.98	0.03
	Quadratic	.99	0.03

The original-equation base model predicts the human data with R^2=0.92 and *RMSD*=0.05 through the linear model, and R^2=0.98 and *RMSD*=0.03 through the quadratic model. The revised-equation base model predicts the human data with R^2=0.98 and *RMSD*=0.03 through the linear model and R^2=0.99 and *RMSD*=0.03 through the quadratic model. Thus, both of the models predict the learning performance of the distributed practice schedule very closely.

Figure 6-3 shows the proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test of the Hybrid1 practice schedule.

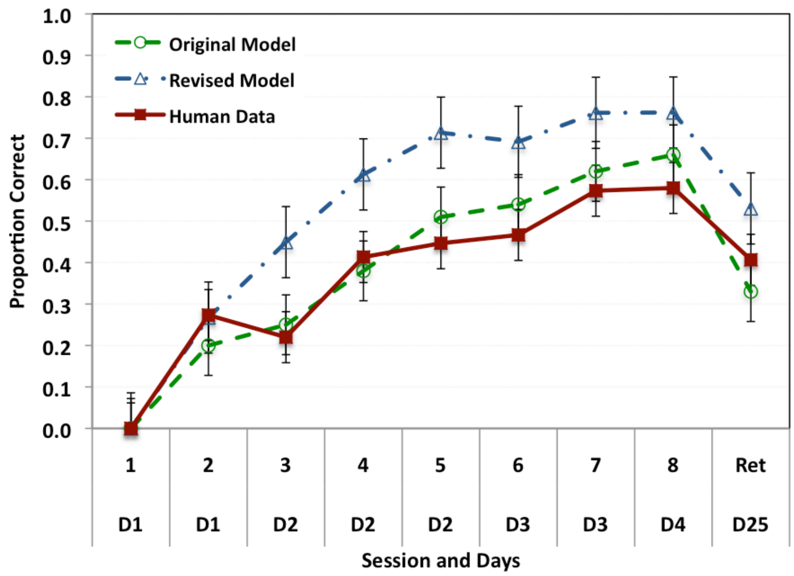


Figure **6-3**: The Hybrid1 practice schedule: The proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test.

As we see the figure 6-3, the original-equation base model predicts the human data more closely than the revised-equation base model in the massed practice schedule. However, the revised-equation base model also predicts the learning and forgetting trends of the human data for the Hybrid1 practice schedule. To validate models with the human data, r^2 and RMSD (root mean square deviation) for both models were calculated and the results are presented in table 6-3.

The original-equation base model predicts the human data with R^2=0.96 and *RMSD*=0.04 through the linear model, and R^2=0.97 and *RMSD*=0.03 through the quadratic model. The revised-equation base model predicts the human data with R^2=0.92 and *RMSD*=0.06 through the linear and quadratic models. Thus, we can conclude both of the models predict the learning performance of the distributed practice schedule very closely.

Table **6-3**: The prediction results of the original-equation base model and the revised-equation base model for the Japanese vocabulary test of the Hybrid1 practice schedule.

ACT-R Model	Model	R^2	RMSD
Original-equation base Model	Linear	.96	0.04
	Quadratic	.97	0.03
Revised-equation base Model	Linear	.92	0.06
	Quadratic	.92	0.06

Figure 6-4 shows the proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test of the Hybrid2 practice schedule.

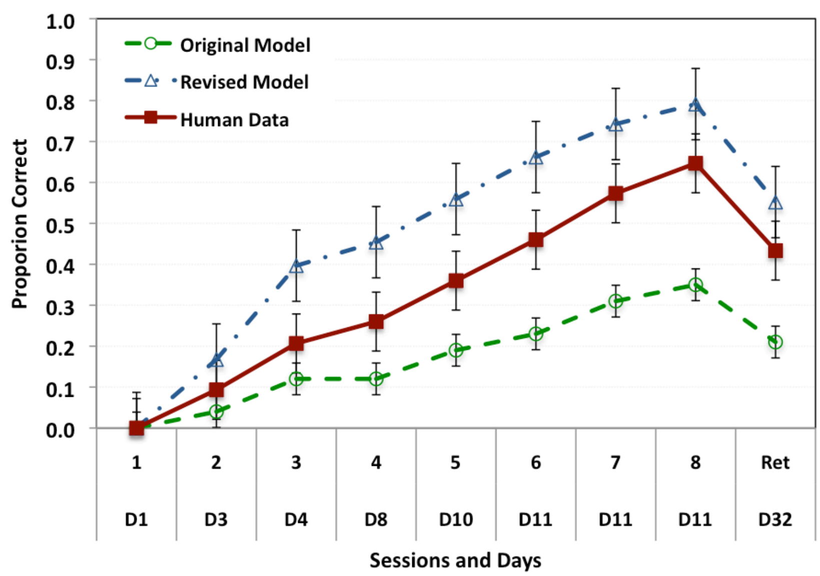


Figure **6-4**: The Hybrid2 practice schedule: The proportion correct of the human data, the model based on the original base-level learning equation, and the model based on the revised base-level learning equation for the Japanese vocabulary test.

As we see the figure 6-4, the human data of the Hybrid2 practice schedule is located between the original-equation base model and the revised-equation base model. However, the original one predicts the learning trends of the human data more closely than the revised one. To validate models with the human data, r^2 and RMSD (root mean square deviation) for both models were calculated and the results are presented in table 6-4.

Table **6-4**: The prediction results of the original-equation base model and the revised-equation base model for the Japanese vocabulary test of the Hybrid2 practice schedule.

ACT-R Model	Model	R^2	RMSD
Original-equation base Model	Linear	.97	0.04
	Quadratic	.99	0.03
Revised-equation base Model	Linear	.94	0.05
	Quadratic	.99	0.03

The original-equation base model predicts the human data with R^2=0.97 and *RMSD*=0.04 through the linear model, and R^2=0.99 and *RMSD*=0.03 through the quadratic model. The revised-equation base model predicts the human data with R^2=0.94 and *RMSD*=0.05 through the linear model and R^2=0.99 and *RMSD*=0.03 through the quadratic model. Thus, we can conclude both of the models predict the learning performance of the distributed practice schedule very closely.

6.2 Summary of Comparison

Two ACT-R models with respect to the four practice schedules were validated against the human data of the Japanese Vocabulary test. Both models predict the learning performance of each trial with respect to the four practice schedules very closely, and the prediction results with R^2 and *RMSD* are presented in table 6-1, 6-2, 6-3, and 6-4. The original-equation base model predicts the human data with massed (4-4-0-0-0-0-0-0) and somewhat massed (2-3-2-1-0-0-0-0) practice schedules more closely than the revised one, however, the revised-equation base model predicts the human data with distributed (1-1-1-1-1-1-1-1) and somewhat distributed (1-0-1-1-1-0-1-3) practice schedules more closely than the original one.

These results are consistent with the arguments of Pavlik and Anderson (2003, 2005) that the original base level learning equation cannot predict the spacing effect of human. However, the model based on the revised base-level learning equation could not predict the human data exactly either, so a new base-level learning equation is still needed to represent this kind of long-term learning and forgetting tasks.

The revised base-level learning equation by Lebiere and Best (2009) that I described in equation 3.6 can predict the long-term learning and forgetting processes of human, because this equation considers to balance long-term reinforcement and short-term inhibition. Using this equation, we can predict the recall probabilities at any specific time as I showed in chapter 3, and then the best practice schedule can be tested through the psychological experiment. Finally, we can predict the performance of candidate schedules using ACT-R models.

The other consideration for revising the base-level learning equation is about sleep deprivation. There is an extensive literature for the negative cognitive consequences associated with less-than-adequate amounts of sleep (Dinges, Rogers, & Baynard, 2005; Durmer & Dinges, 2005; Gunzelmann et al., 2009), however, most of the cognitive architectures do not reflect the effect of sleep deprivation in their architectures. The approach to represent this effect was just changing the associated parameters to operate less effectively or efficiently in existing mechanisms (Gunzelmann, Cluck, Price, Van Dongen, & Dinges, 2007; Gunzelmann, Gluck, Van Dongen, O'Conner, & Dinges, 2005; Gunzelmann et al., 2009). It is good start to predict the effect of sleep deprivation by adjusting ACT-R parameters, however, the adequate amount of sleep is the one of the most important factors in performance of human cognition, this factor should not be presented as a parameters, but be located in the basic theory of the ACT-R cognitive architecture. This can be presented with the decay function (the decay value is a constant in the current version of ACT-R 6.0) like as Pavlik and Anderson did in their revised base-level learning equation, or it also can be presented as an added equation like as Lebiere and

Best did in their revised equation. Whichever we choose, it can provide more accurate prediction of human behaviors. I remain a new equation as a future work.

Chapter 7

Conclusions and Discussions

In this chapter, I present summaries of findings, contributions, insights, and future works for this dissertation. The goal of my dissertation is exploring new training paradigms that could produce better performance than a distributed practice schedule at retention. To achieve this goal, I have approached in the three ways. First, I investigated the learning and forgetting theories of the ACT-R cognitive architecture that is called base-level learning equation, and I could find the best practice schedule that is supported by the theory. Second, this practice schedule was investigated with the massed and distributed practice schedules through the psychological experiment in a laboratory setting. Finally, ACT-R cognitive models were made to predict human behavior and compared with the results of the human data. By following these steps, I can provide theoretical and practical contributions in investigating practice schedules, and these are presented in next section.

The results of the prediction by the ACT-R theories showed that the hybrid practice schedule (1-0-1-1-1-0-1-3) could produce 2.5% better performance than the distributed practice schedule (1-1-1-1-1-1-1-1) at the retention test in my experiment setting (8 learning session across two weeks and retention test after 21 days). I also found that the distributed practice schedule ranks at the 754th and the massed practice schedule ranks ate the 6,103th among the 6,345 candidate practice schedules.

Among the 6,345 candidate practice schedules, four practice schedules include distributed (1-1-1-1-1-1-1-1), semi-massed (4-4-0-0-0-0-0-0), Hybrid1 (1-0-1-1-1-0-1-3), and Hybrid2 (2-3-2-1-0-0-0-0), were used in the experiment. Three kinds of knowledge types, declarative, procedure, and perceptual-motor skill, were investigated with four kinds of tasks, the

Japanese-English vocabulary test, the Tower of Hanoi test, the simple Permutation problem-solving test, and the iPodtouch based inverted pendulum test.

The results of human data showed that the distributed practice schedule is better than the massed practice schedule in the Japanese vocabulary test (declarative knowledge), however, there are no significant differences among the other practice schedules in this task. There are no significant differences among the four practice schedules in the two procedural tasks (Permutation problem-solving task and the Tower of Hanoi puzzle), and I could find that the knowledge of all of the participants in each group was proceduralized in their memory during the entire learning sessions.

There is significant difference between the Hybrid1 and the distributed practice schedules in the perceptual-motor skill task. The participants in the Hybrid1 and massed practice schedules showed better performance than the distributed and the Hybrid2 practice schedules, and these results indicated that massed and somewhat early massed practice schedules could help the knowledge proceduralized in this kind of task. Table 7-1 shows the results of the human data that I found in my experiment.

Table **7-1**: The effective practice schedules with respect to knowledge types.

Knowledge Types	Tasks	Effective Practice Schedules
Declarative Knowledge	The Japanese Vocabulary Task	D > H2 > H1 > M
Procedural Knowledge	The Permutation Problem-Solving Task	M ≈ D ≈ H1 ≈ H2
	The Tower of Hanoi Task	M ≈ D ≈ H1 ≈ H2
Perceptual-Motor Skill	Inverted Pendulum Task	H1 > M > H2 > D

Note: *D* stands for the distributed practice schedule, *M* stands for massed practice schedule, *H1* stands for the Hybrid1 practice schedule, and *H2* stands for Hybrid2 practice schedule.

Two kinds of the ACT-R models, the model based on the original base-level learning equation and the model based on the revised base-level learning equation, were developed to predict the human behavior of the Japanese vocabulary test. Those models were simulated for 100 trials, and the average proportion correct were gathered and compared with the human data. Both models followed the learning and forgetting trends of the human data, however, they could not predict the human data exactly.

7.1 Contributions

The results of this study could shift a training paradigm of how to design training programs in the fields of education, industry, and military with respect to the knowledge types of tasks. The detailed contributions are presented below.

The first contribution of this dissertation is to provide a new paradigm of training different from the widely used method in most training. Most of the previous studies have focused on massed practice and distributed practice schedules, but this dissertation shows new hybrid training paradigm could exist, and it can produce better performance in knowledge and skills acquisition and retention. Furthermore, the later researchers who explore training schedules for better retention could consider testing a hybrid practice schedule in their research.

The second contribution of this dissertation is to provide a theory of how to investigate practice schedules that are supported by the theory. In this dissertation, I used the base-level learning equation of the ACT-R cognitive architecture (original), and its extension (revised), and I could predict the activation strength at each learning session and the retention session. Throughout the results of the prediction, I could find that the best practice schedule is not a distributed schedule, but a hybrid practice schedule.

The third contribution of this dissertation is to explore the most efficient training schedule for three kinds of knowledge types. The previous studies investigated two practice schedules with the specific knowledge type, such as declarative or procedural knowledge, or perceptual-motor skill. However in this study, I investigated the three knowledge types with the four kinds of practice schedules, and from the results of these experiments, most efficient practice regimen with respect to the specific knowledge type was provided.

The fourth contribution of this dissertation is to examine the ability of the ACT-R cognitive architecture. The most of the previous ACT-R models have focused on microscopic psychological tasks, however, model for learning, forgetting, and retaining aspects of human mind with long-term duration has not been verified. By comparing the ACT-R models with the human data, I could show that models could predict the learning trends of the human data, however, they could not predict the human data exactly or somewhat exactly. These results indicate that ACT-R needs some extensions to predict long-term learning and forgetting process of humans.

The fifth contribution of this dissertation is to start for examining the theory of skill retention (Kim et al., in press). According to the theory of skill retention, the forms of forgetting are different in each learning stage, so learning should occur with different manners and degrees with respect to the learning stages. The hybrid practice schedules that I used in this study might be the schedules that could be explored to test this theory.

7.2 Future Work

Although this dissertation gives insight the needs of paradigm shifting for the training, there still remain several things to investigate as future works. I present the list of future work in this section.

I tested four practice schedules in this study. According to Pavlik and Anderson's revised learning equation showed in chapter 3, the Hybrid2 practice schedule provides the best performance at retention. However, the results of the human data were not consistent with equation (The distributed practice schedule is the best and the Hybrid2 practice schedule is the second). Although the sample size ($N = 40$) of the experiment is not enough, this result is almost identical to the findings of the recent study (Cepeda et al., 2008) that showed the Pavlik and Anderson's equation could not predict their human data either. That is, the revised base-level learning equation is not enough to predict various learning interval and retention interval, so it still needs to be revised as I mentioned in chapter 3. The revised equation of Lebiere and Best (2009) that have not been validated with empirical data may produce better prediction for long-term learning and forgetting processes of human than the revised equation of Pavlik and Anderson (2003, 2004, 2005). The recall probabilities of the entire candidate practice schedules (6,435) will be explored by using the equation, and the results will be compared with the human data and the results of Pavlik and Anderson's revised base-level learning equation.

The Predict Performance Equation (Jastrzembski et al., 2010) will be also explored to predict the best practice schedule among the candidates practice schedules, and the results will be compared with the human data.

Between the two hybrid schedules, the Hybrid2 schedule is the best, and the Hybrid1 schedule ranks 5,532 among the 6,435 candidate schedules. The reason I tested the Hybrid1 practice schedule is not by theory, but by the recent work (Kim et al., in press) and my thought that some tasks require the procedural knowledge or perceptual-motor skills should be trained in a massed or somewhat massed way to get better performance. However, we can generate numerous similar practice schedules, such as 3-2-1-1-0-0-0-0 or 2-4-1-1-0-0-0-0, and among these possible schedules we cannot choose the best one without any theoretical supports. Thus, we need theories to predict the performance of procedural or perceptual-motor tasks.

Among the four tasks, I made ACT-R models for one task, the Japanese vocabulary task. The inverted pendulum type task could not be modeled because of the limitation of the current ACT-R's perceptual-motor component. The other tasks, the Permutation problem-solving task and the Tower of Hanoi puzzle, will be modeled and compared with the human data.

The number of participants (*N* = 40) and some outliers in each practice schedule may result in failing significant differences among the schedules. For example, the proportion correct in the Japanese vocabulary test was not significant difference among the practice schedules, however, when I excluded the outlier in the massed practice group, I found the significant difference between the distributed and the massed group. Thus, I will recruit more participants to get better results.

References

Anderson, J. R. (1976). *Language, memory, and thought*: Lawrence Erlbaum.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (2007). *How can the human mind exist in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review, 111*(4), 1036-1060.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Hillsdale: Erlbaum.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory & Cognition, 25*(5), 1120-1136.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction, 12*(4), 439-460.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86*, 124-140.

Atkinson, R. C. (1972). Optimizing the learning of second-language vocabulary. *Journal of Experimental Psychology, 96*, 124-129.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296-308.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 316-321.

Bahrick, H. P., & Phelps, E. (1987). Retention of Spanish vocabulary over 8 years. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(2), 344-349.

Bloom, K. C., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*(4), 245-248.

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies, 55*(1), 41-84.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Calfee, R. C. (1985). *Experimental methods in psychology*. New York, NY: Holt, Rinehart and Winston.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing Distributed Practice. *Experimental Psychology (formerly" Zeitschrift f‚r Experimentelle Psychologie"), 56*(4), 236-246.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*(3), 354.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal redgeline of optimal retention. *Psychological Science, 19*(11), 1095-1102.

Chong, R. S. (2004). Architectural explorations for modeling procedural skill decay. In M. Lovett, C. Schunn, C. Lebiere & P. Munro (Eds.), In *Proceedings of the Sixth International Conference on Cognitive Modeling*. Mahwah, NJ: Erlbaum.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*: Lawrence Erlbaum.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33*(4), 497-505.

Dinges, D. F., Rogers, N. L., & Baynard, M. D. (2005). Chronic sleep deprivation. In M. H. Kryger, T. Roth & W. C. Dement (Eds.), In *Principles and practice of sleep medicine* (pp. 67-76). Philadelphia: W. B. Saunders.

Durmer, J. S., & Dinges, D. F. (2005). Neurocognitive consequences of sleep deprivation. *Seminars in Neurology, 25*(1), 117-129.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York: Dover.

Fishman, E. J., Keller, L., & Atkinson, R. C. (1968). Massed versus distributed practice in computerized spelling drills. *Journal of Educational Psychology, 59*(4), 290-296.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling amplitude of movement. *Journal of Experimental Psychology, 47*(6), 381-391.

Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (Vol. 47, pp. 381-391). New York: Academic Press.

Fleishman, E. A., Quaintance, M. K., & Broedling, L. A. (1984). *Taxonomies of human performance: The description of human tasks*: Academic Press.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., Smith, C. D., & Satava, R. M. (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.

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.

Gray, W. D., & Altmann, E. M. (2001). Cognitive modeling and human-computer interaction. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors*. New York: Taylor & Francis, Ltd.

Gunzelmann, G., Cluck, K. A., Price, S., Van Dongen, H. P. A., & Dinges, D. F. (2007). Decreased arousal as a result of sleep deprivation: The unraveling of cognitive control. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 243-253). New York, NY: Oxford University Press.

Gunzelmann, G., Gluck, K. A., Van Dongen, H. P. A., O'Conner, R. M., & Dinges, D. F. (2005). A neurobehaviorally inspired ACT-R model of sleep deprivation: Decreased performance in psychomotor vigilance. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), In *Proceedings of the Twenty-Seventh Annual Meeting of the Cognitive Science Society* (pp. 857-862). Mahwah, NJ: Erlbaum.

Gunzelmann, G., Gross, J. B., Gluck, K. A., & Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive Science, 33*(5), 880-910.

Jastrzembski, T. S., Addis, K., Krusmark, M., Gluck, K. A., & Rodgers, S. (2010). Prediction intervals for performance prediction. In D. D. Salvucci & G. Gunzelmann (Eds.), In *Proceedings of 10th International Conference on Cognitive Modeling* (pp. 109-114). Philadelphia.

Jastrzembski, T. S., & Gluck, K. A. (2009). A formal comparison of model variants for performance prediction. In D. P. A. Howes, R. Cooper (Ed.) In *Proceedings of the 9th International Conference of Cognitive Modeling*. Manchester, UK.

Jastrzembski, T. S., Gluck, K. A., & Gunzelmann, G. (2006). Knowledge tracing and prediction of future trainee performance. In *Proceedings of the 2006 Interservice/Industry Training, Simulation, and Education Conference* (pp. 1498-1508). Orlando, FL: NTSA.

Jastrzembski, T. S., Gluck, K. A., Rodgers, S., & Krusmark, M. (2009). The predictive performance optimizer: Mathematical modeling for performance prediction. In *Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation* (pp. 141-142). Sundance, UT.

Kieras, D. E., & Meyer, D. E. (1996). The EPIC architecture: Principles of operation. *Unpublished manuscript from ftp://ftp.eecs.umich.edu/people/kieras/EPICarch.ps*.

Kim, J. W. (2008). *Procedural skills: From learning to forgetting.* Department of Industrial and Manufacturing Engineering, The Pennsylvania State University, University Park, PA.

Kim, J. W., Koubek, R. J., & Ritter, F. E. (2007). Investigation of procedural skills degradation from different modalities. In R. L. Lewis, T. A. Polk & J. L. Laird (Eds.), In *Proceedings of 7th International Conference on Cognitive Modeling* (pp. 255-260). Triesta, Italy: Oxford, UK: Taylor & Francis/Psychology Press.

Kim, J. W., Ritter, F. E., & Koubek, R. J. (in press). An intergrated theory for imporved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*.

Lebiere, C., & Best, B. J. (2009). Balancing long-term reinforcement and short-term inhibition. In *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Mackay, S., Morgan, P., Datta, V., Chang, A., & Darzi, A. (2002). Practice distribution in procedural skills training. *Surgical endoscopy, 16*(6), 957-961.

Martin-Emerson, R., & Wickens, C. D. (1992). The vertical visual field and implications for the head-up display. In (pp. 1408-1412). Human Factors and Ergonomics Society.

Meyer, D. E., & Kieras, D. (1997). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*(1), 3-65.

Moulton, C. A. E., Dubrowski, A., MacRae, H., Graham, B., Grober, E., & Reznick, R. (2006). Teaching surgical skills: What kind of practice makes perfect?: A randomized, controlled trial. *Annals of Surgery, 244*(3), 400.

Newell, A. (1990). *Unified Theories of Cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-51). Hillsdale, NJ: Erlbaum.

Ohlsson, S. (2011). *Deep learning: How the mind overrides experience*. New York, NY: Cambridge Univ 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.* Unpublished PhD thesis, Department of Industrial Engineering, The Pennsylvania State University, University Park.

Paik, J., Kim, J. W., & Ritter, F. E. (2009). A preliminary ACT-R compilier in Herbal. In *Proceedings of 9th International Conference on Cognitive Modeling* (pp. 466-467). Manchester, England.

Paik, J., Kim, J. W., Ritter, F. E., Morgan, J. H., Haynes, S. R., & Cohen, M. A. (2010a). Building large learning models with Herbal. In D. D. Salvucci & G. Gunzelmann (Eds.), In *Proceedings of 10th International Conference on Cognitive Modeling* (pp. 187-192). Philadelphia.

Paik, J., Kim, J. W., Ritter, F. E., Morgan, J. H., Haynes, S. R., & Cohen, M. A. (2010b). Building user models faster with Herbal. In *Human Computer Interaction Consortium (HCIC) Workshop 2010*. [unrefereed workshop paper/presentation].

Pavlik, P. I. (2007). Understanding and applying the dynamics of test practice and study practice. *Instructional Science, 35*(5), 407-441.

Pavlik, P. I., & Anderson, J. R. (2003). An ACT-R model of the spacing effect. In *Proceedings of Fifth International Conference on Cognitive Modeling* (pp. 177-182). Bamberg, Germany.

Pavlik, P. I., & Anderson, J. R. (2004). An ACT-R model of memory applied to finding the optimal schedule of practice. In *Proceedings of Sixth International Conference on Cognitive Modeling* (pp. 376-377). Pittsburgh, PA USA.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29*(4), 559-586.

Pavlik, P. I., & Anderson, J. R. (2008). Using a model to compute the optimal schedule of practice. *Journal of Experimental Psychology: Applied, 14*(2), 101-117.

Pew, R. W., & Mavor, A. S. (Eds.). (1998). *Modeling human and organizational behavior: Application to military simulations*. Washington, DC: National Academy Press. books.nap.edu/catalog/6173.html.

Proctor, R. W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage Publications, Inc.

Qin, Y., Carter, C. S., Silk, E. M., Stenger, V. A., Fissell, K., Goode, A., & Anderson, J. R. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences of the United States of America, 101*(15), 5686.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93.

Raaijmakers, J. G. W. (2003). Spacing and repetition effects in human memory: Application of the SAM model. *Cognitive Science, 27*(3), 431-452.

Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*: Elsevier Science Inc. New York, NY, USA.

Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (in press). Determining the number of model runs: Treating cognitive models as theories by not sampling their behavior. In S. Narayanan & L. Rothrock (Eds.), *Human-in-the-loop simulations: Method and Practice*: Springer.

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In *International encyclopedia of the social and behavioral sciences* (Vol. 13, pp. 8602-8605). Amsterdam: Pergamon.

Ritter, F. E., & Wood, A. B. (2005). Dismal: A spreadsheet for sequential data analysis and HCI experimentation. *Behavior Research Methods, 37*(1), 71-81.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20*(9), 1209-1224.

Rumelhart, D. E. (1967). The effects of interpresentation intervals on performance in a continuous paired-associate task. *Institute for Mathematical Studies in the Social Sciences, Stanford University*.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say broke? A model of learning the past tense without feedback. *Cognition, 86*(2), 123-155.

Taatgen, N. A., & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors, 45*(1), 61-76.

Training Magazine. (2009). Gagues & drivers: Industry report. *Training Magazine,* 16-34.

Vearrier, L. A., Langan, J., Shumway-Cook, A., & Woollacott, M. (2005). An intensive massed practice approach to retraining balance post-stroke. *Gait & Posture, 22*(2), 154-163.

Wilson, M. S., Middlebrook, A., Sutton, C., Stone, R., & McCloy, R. F. (1997). MIST VR: a virtual reality trainer for laparoscopic surgery assesses performance. *Annals of the Royal College of Surgeons of England, 79*(6), 403.

Young, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology, 8*(1), 58-81.

Zbrodoff, N. J. (1995). Why is 9+ 7 harder than 2+ 3? Strength and interference as explanations of the problem-size effect. *Memory and Cognition, 23*, 689-689.

Zhao, C., Paik, J., Morgan, J. H., & Ritter, F. E. (2010a). Validating a High Level Behavioral Representation Language (HERBAL): A Docking Study for ACT-R. In *First International Conference on Biologically Inspired Cognitive Architectures*. Virginia.

Zhao, C., Paik, J., Morgan, J. H., & Ritter, F. E. (2010b). Validating a high level behavioral representation language: Docking study between Herbal and ACT-R. *International Journal of Machine Consciousness, 2*(2), 383.

Appendix A

The Experiment Materials

The Japanese Vocabularies and meanings (Pavlik & Anderson, 2005)

chitai	area
ushiro	back
ginkou	bank
dodai	base
keta	beam
itamae	cook
nendai	date
sore	fear
fujo	help
saishuu	last
itonami	life
yuubin	mail
keiro	path
amari	rest
shiren	test

The Simple Permutation Problems (Rohrer & Taylor, 2006)

abccc	20
abcccc	30
aabbbbb	21
abbcc	30
aaabbb	20
aabbb	15
aabb	6
abbccc	60
abccccc	42
aabbbbbb	28
abbcccc	105
abcccccc	56

VITA

JAEHYON PAIK

Department of Industrial & Manufacturing Engineering
The Pennsylvania State University
319 IST Building
University Park, PA 16802
Office: +1 (814) 865-4455
Mobile: +1 (814) 876-0984
jaehyon.paik@psu.edu

Research Interests

I am interested in modeling human behavior using cognitive architectures, such as ACT-R and Soar, to improve human-computer interaction. I am also interested in investigating an optimal training schedule to increase performance at retention using cognitive theories and models.

Education

Ph.D.	The Pennsylvania State University, 2011 (Industrial and Manufacturing Engineering)
M.S.	Ajou University, 2000 (Industrial Engineering)
B.S.	Ajou University, 1998 (Industrial Engineering)

Professional Experience

In Academia

Research Assistant

- Project: *Pathway-based Integration of Relational Networks and Textual Information for Detecting Motivation and Intent of WMD-related Threats*, sponsored by Defense Threat Reduction Agency (DTRA), Sep. 2009 ~ Present
- Project: *Implementation of the ACT-R compiler in the Herbal Environment,* sponsored by Office of Naval Research (ONR), Jan. 2009 ~ Present
- Project: *School Like Mine,* sponsored by Regional Educational Laboratory Mid-Atlantic (RELMA), Sep. 2007 ~ Dec. 2009

Lecturer

- Department of Computer Science, Kyonggi University, South Korea

In Industry

Research & Development Consultant

SmartNet Technology, South Korea, Mar. 2002 ~ May. 2005
Ablex Inc., South Korea, Jun 2000 ~ Feb. 2002 (2011)(2011)(2011)