２０２０年度

青山学院大学審査学位論文

指導教員　Peter, J. Robinson　教授

難易度の異なる学習タスクにおける第二言語語彙知識の保持と転移

Second Language Vocabulary Acquisition: Retention and Transfer Under Easy and Difficult Task Conditions

文学研究科

英米文学専攻

田邉　正行

SECOND LANGUAGE VOCABULARY ACQUISITION:

RETENTION AND TRANSFER

UNDER EASY AND DIFFICULT TASK CONDITIONS

A Dissertation

Submitted in Partial Fulfilment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of English and American Literature,

Graduate School of Literature

Aoyama Gakuin University

Tokyo, Japan

By Masayuki TANABE

October 30th, 2020

ABSTRACT

SECOND LANGUAGE VOCABULARY ACQUISITION:
RETENTION AND TRANSFER
UNDER EASY AND DIFFICULT TASK CONDITIONS

Masayuki TANABE

A major challenge in education, including language learning and teaching, is how to promote the retention of knowledge and skills provided in a classroom and their successful reproduction in practical situations. The durability and transferability of acquired knowledge and skills are fundamentally deserving of attention when considering the fact that there is a certain period of disuse after learning, and that situations of learning and subsequent performance are, in most cases, different. The depth-of-processing framework (Craik & Lockhart, 1972) in psychology argues that an input processed at a deep level will be retained longer than that processed at a shallow level because the strength of memory traces is associated with the level of input processing. The desirable difficulty framework (Bjork, 1994) in second-language acquisition holds that post-training retention and transfer will become maximized when an optimal level of difficulty for a learner is provided during training. Therefore, the phrase "Easy and Difficult Task Conditions" in the title of the present research refers to difficulties learners have in successful task completion during training that are influenced by various task characteristics. Schneider, Healy, and Bourne (2002) demonstrated, by operationalizing levels of task difficulty with translation directions in paired-associate word learning, that foreign word knowledge learned in a difficult

training condition (recalling words in L2 from L1 cues) was more retained and transferable in the long run than that learned in an easy training condition (recalling words in L1 from L2 cues). Further, they argued that any type of difficulty during training benefits the retention and subsequent transfer of acquired knowledge. Therefore, the present study attempted to replicate Schneider *et al.* (2002) in two ways. In a direct replication (Study 1), Japanese students learned a list of Indonesian–Japanese word pairs. Half the participants were trained by recalling Japanese words from Indonesian cue words (easy training condition), and the other half recalled Indonesian words from Japanese cue words (difficult training condition). Results replicated the patterns that Schneider *et al.* found, in which learning in the easy training condition enhanced immediate learning but the rate of retention a week after the learning session was better for the difficult training condition. A conceptual replication (Study 2) operationalized an ESL environment, in which an entire class is conducted in a foreign language with no translations involved. In the easy training condition, Japanese students were pretrained on English-like nonwords as targets, and then read English passages containing target nonwords with marginal glosses for their meaning. In the difficult training condition, pretrained participants read the English passages without such glosses. In contrast to the results of Study 1, the difficult training condition outperformed the easy training condition in the rate of learning. Furthermore, the rates of subsequent retention and transfer were comparable to each other. A separate small-scale investigation into perceived difficulty of the present experimental tasks confirmed that participants considered the difficult training conditions of the present tasks more difficult than the easy training conditions. These results suggested that task difficulty brought about by manipulating the translation direction facilitates the subsequent retention and transfer of acquired second-language vocabulary knowledge. Meanwhile, a task-difficulty factor other than the translation direction had marginal effects on the retention and transfer, suggesting that the effects of task difficulty in monolingual second-language classes might be limited. Results also suggested that existing knowledge contributes to the retention of newly acquired knowledge. Successful subsequent recall of learned L2 words could be influenced by retention of L2 forms to a greater degree than retention of conceptual associations between L2 and L1.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

A major challenge in education, including language learning and teaching, is how to promote retention of knowledge and skills provided in a classroom and their successful reproduction in practical situations. The durability and transferability of acquired knowledge and skills are fundamentally deserving of attention when considering the fact that there is a certain period of disuse after learning, and that situations of learning and subsequent performance are, in most cases, different. In studies of second-language acquisition, this issue has been discussed in relation to the introduction of difficulty into various learning tasks since Schmidt and Bjork's theorizing of "desirable difficulty" (Schmidt & Bjork, 1992; Bjork, 1994a, 1994b). The desirable difficulty framework holds that post-training retention and transfer will become maximized when an optimal level of difficulty for a learner is provided during training. Therefore, the phrase "Easy and Difficult Task Conditions" in the title of the present research refers to difficulties learners have in successful task completion during training that are influenced by various task characteristics. Findings in experimental psychology suggest that effortful input processing during training promotes the retention of leaned knowledge and skills because durable memory traces are generated by the elaborate input processing (Craik & Lockhart, 1972). For example, materials given in active learning tasks, such as those where learners need to produce answers for missing parts of the material without reference to correct answers, enhance subsequent recall of the material compared to when simply reading the whole material passively, which is known as the "generation effect" (Crutcher & Healy, 1989; Slamecka & Graf, 1978). Another example is that, in learning that involves repeated practice, material is retained longer when a rest interval is given between training sets than when it is provided successively without intervals. This is referred to as the "spacing effect" (Bahrick *et al.*, 1993; Bird, 2010; Karpicke & Roediger, 2007). The present paper investigates task difficulty, one of those paradigms that leads to elaborate processing for the retention of inputs. Healy, Schneider, and Bourne (2012) mentioned that difficulty in learning refers to "interference" that leads the learner to make an incorrect response. A difficult learning task hinders the rate of learning in the short run but facilitates the retention and transferability of what is learned in the long run. An easy task for learners is not necessarily good in terms of the long-term efficiency.

Schneider, Healy, and Bourne (2002) demonstrated, by operationalizing levels of task difficulty with translation directions in paired-associate word learning, that foreign word knowledge learned in a difficult training condition (e.g., translating a word from a first language to a second language) was better retained and more transferable in the long run than that learned in an easy training condition (e.g., translating a word from a second language to a first language). Furthermore, they argued that any type of difficulty during training benefits the retention and subsequent transfer of acquired knowledge. However, a closer looking at their results suggested that an easy task and a difficult task might each have different effects on the retention and transfer, namely, learning in an easy condition facilitates the retention rather than the transfer, whereas learning in a difficult condition promotes the transfer rather than the retention. Therefore, the first part of the present research (Study 1) involved a direct replication of Schneider *et al.*'s (2002) work by using a different set of languages and different learner populations. The second part of the research (Study 2), a conceptual replication, extended Study 1 to both theoretically challenging and pedagogically applicable usages with three changes. First, the implementation of task difficulty was made on post-learning training tasks rather than at the time of encoding. Second, task difficulty was operationalized by employing a “here-and-now vs. there-and-then” task condition (Robinson, 1995, 2001a) in the same second language (i.e. learning of English-like nonwords with or without English definitions), rather than by altering translation directions between two languages. Finally, the transferability of word knowledge was tested on a picture naming task, a method considered to involve more practically desirable use of language than merely translating words from one language to another. These extensions dealt with further questions such as:

(a) while the depth-of-processing framework (Craik & Lockhart, 1972) predicts that the retention of inputs can only be determined at the time of encoding, do levels of difficulty also matter on a practice phase immediately after learning?

(b) does a type of difficulty that is different from those Schneider *et al.* referred to also enhance the retention and subsequent transfer as they argued? and

(c) with respect to foreign vocabulary learning, can the framework of task difficulty be readily applied to typical ESL (English as a second language) classes, such as those conducted only in English language where translation into students’ first language is less likely?

In addition to these two main studies, a small-scaled study to look into perceived difficulty of the present experimental tasks will be addressed.

The dissertation starts by reviewing classic human memory studies that are still influential today (Chapter 1), then narrows down to skill acquisition theories in psychology including empirical investigations on task difficulty (Chapter 2), and then illustrates a lack of "difficulty" aspects as indices among contemporary second-language vocabulary acquisition frameworks (Chapter 3). Following these review chapters, a direct replication of Schneider *et al.* (Chapter 4), a conceptual replication (Chapter 5), and an additional investigation of perceived difficulty of the present tasks (Chapter 6) are reported. Theoretical explanations of findings of the present research are made in the general discussion chapter, as well as pedagogical implications, limitations, and suggestions for future studies, including conclusions of the research (Chapter 7).

Chapter I

TWO WAYS OF LOOKING AT THE HUMAN MEMORY SYSTEM

1.0 Introduction

In this introductory chapter, theories of how human memory is maintained are reviewed. In psychology, two major views explaining the functions of the human memory system exist. Theorists of the multiple storage view assume that the human memory system consists of several stores, each of which is characterized by its unique function. For example, when asking why a person forgot information soon after processing it, these researchers may contend that the information was only stored in a temporary buffer. In contrast, those who hold a processing view argue that types of processes for input are responsible for determining the degree of retention, rather than types of storage and so forth. Therefore, such scholars may assert that the person did not process the information at a level such that they could generate a firm memory trace.

1.1 Multiple Storage Theories

Stimulated by early computer architectures in the 1960s, psychologists theorized that the human memory system is similar to ways in which computers function (Foster, 2009). A three-stage model of the human memory system was first proposed. Later, this simple model was developed by many researchers who proposed various modifications. It is not an exaggeration to say that the multiple storage model laid a foundation for current theories of the human memory system.

1.1.1 Sensory Register, Short- and Long-term Stores

One of the contemporary ideas about how information is stored in the brain entails a view that human memory consists of several subsystems that are responsible for a particular function. A structural model of the memory system, proposed by Atkinson and Shiffrin (1968), assumes three stores: a sensory register, a short-term store, and a long-term store. In their model, an external stimulus first enters the sensory register. The entry of a stimulus to this storage is automatic regardless of the subject's attention to input (Craik & Lockhart, 1972). Here, the model supposes that stimuli are either visual or auditory. The sensory register holds input transiently and is copied to the short-term store if attention is present. In the short-term store, the transferred information cannot not be deleted by the control processes of the subject but decays after a certain period of time passes. Retention in the sensory register is estimated to be 0.25 to 2 seconds (Craik & Lockhart, 1972). Retention in the short-term store can be 5 to 20 seconds (Craik & Lockhart, 1972) or 15 to 30 seconds (Atkinson & Shiffrin, 1968). The short-term store deals with auditory information. Therefore, for example, a word presented in a visual mode enters the sensory register as visual information but is phonemically encoded when it is sent to the short-term store. In Atkinson and Shiffrin's (1968) model of the memory system, the short-term store is referred to as tht "auditory-verbal-linguistic store." The information passed into the short-term store can be transferred to the long-term store by conscious rehearsal processes of the subject (Waugh & Norman, 1965). With its limited-capacity characteristics (Miller, 1956), the short-term store quickly loses unrehearsed information because subsequent stimuli override the working information in this storage (Waugh & Norman, 1965). Once information enters the long-term store, it can be held permanently. Whereas the capacity of the short-term store is limited, the long-term store has an unlimited capacity. Major differences between the long-term store and the other two involve the durability, specificity of modalities of information, and retrieval processes involved. Information in the sensory register and the short-term store is eventually lost completely, whereas information in the long-term store lasts a long time, possibly a whole lifetime. Modalities of information are limited in the sensory register and the short-term store, but various formats are possible in the long-term store. It is hypothesized that a literal copy of an input can reside in the sensory register in both a visual mode and an auditory mode depending on the type of stimulus and whether or not phonemic coding is needed for the information to be retained in the short-term store. In contrast, the long-term store is considered to be largely semantic (Baddeley, 1966), and both visual and auditory modalities are possible, as evidenced by our ability to remember in these modes.

Search processes are another characteristic of the long-term store. Information can be retrieved from the long-term store and sent back to the short-term store, which is in the opposite direction to the normal information flow for acquisition. This process mainly occurs under the control of the subject, as seen in a cued recall paradigm. It is assumed that search processes in the sensory register are not likely to occur because of a very quick rate of information decay. Controlled retrieval from the short-term store is also difficult to identify because, even if the subject uttered a given stimulus perfectly, we cannot distinguish whether it is the result of the subject's rehearsal capability or a pure retrieval of unrehearsed input. However, Atkinson and Shiffrin (1968) mentioned that retrieval from the short-term store is theoretically not impossible, noting that, in such cases, a search for a particular piece of information should be completed quickly and efficiently due to the temporary nature of the store. Craik and Lockhart (1972) summarized the major characteristics of these stores (see Table 1.1), making it possible to distinguish them.

Table 1.1

Commonly Accepted Differences between the Three Stages of the Verbal Memory.

	Sensory registers	Short-term store	Long-term store
Entry of information	Pre-attentive	Requires attention	Rehearsal
Maintenance of information	Not possible	Continued attention Rehearsal	Repetition Organization
Format of information	Literal copy of input	Phonemic Probably visual Possibly semantic	Largely semantic Some auditory and visual
Capacity	Large	Small	No known limit
Information loss	Decay	Displacement Possibly decay	Probably no loss Loss of accessibility or discriminability by interference
Trace duration	1/4-2 seconds	Up to 30 seconds	Minutes to years
Retrieval	Readout	Probably automatic Items in consciousness Temporal/phonemic cues	Retrieval cues Possibly search process

(Adopted from Craik and Lockhart, 1972, p. 672)

1.1.2 Procedural, Semantic, and Episodic Memories

In Tulving's ternary classification of the human memory system (Tulving, 1985), procedural memory is considered to be the most primitive memory system, which non-human organisms may also possess. This memory holds knowledge of connections between stimuli and responses, including those involving motor and simple cognitive skills to adopt to the environment (Schacter *et al.*, 2000; Surprenant & Neath, 2009). Knowledge in this memory is also associated with the capability to react to external and internal stimuli by sensing them without consciousness (Tulving, 1985). Procedural memory can be anatomically distinguished from the other two memory systems due to its ability to function normally with damage to the medial temporal lobe structures (Schacter *et al.*, 2000), a part of the brain essential for establishing long-term memory of facts and events (Squire & Zola-Morgan, 1991). It is also theoretically differentiated from the other two, in which knowledge represented in procedural memory (a) can be expressed in a fixed way in which responses based on this system are largely restricted in the format determined at the time of learning; (b) relies heavily on gradual, incremental learning which involves repetitions; and (c) has a “non-declarative” (Squire, 1987, 2004) or “anoetic” (Tulving, 1985) nature that the subject often cannot verbalize, meaning that it is difficult to transfer to others.

Semantic memory encompasses general knowledge, including facts, concepts, and vocabulary (Tulving, 1983). It is a “mental thesaurus” (Tulving, 1984), which is necessary for us to construct the world mentally. It is argued that semantic memory plays a central role in the use of language with its contents in relation to the knowledge of linguistic symbols and their meanings, relationships among them, and rules for how to handle them (Tulving, 1984). Unlike procedural memory, representations in semantic memory are largely “declarative” (Squire, 1987; 2004) or “noetic” (Tulving, 1985), requiring conscious retrieval processes. Furthermore, knowledge in this system is stored in separation from learning episodes, in which only declarative parts of associations of related components are available (Surprenant & Neath, 2009). Semantic memory is associated with knowing “that” in contrast with a major characteristic of procedural memory, knowing “how.” Therefore, for example, people use semantic memory to answer questions such as “What is the capital of France?” “How many days are there in the week?” or “What direction would you travel in if you were flying from London to Johannesburg?” (Foster, 2009, p. 40), whereas procedural memory is responsible for skills such as riding a bicycle or reading (Schacter, *et al.*, 2000, p. 636).

Episodic memory handles contextual information about events or episodes experienced by the subject. Only this system allows subjects to “travel back” (Tulving, 2002, p. 2) through their personal past mentally in subjective time. The major characteristics of this system are marked by “pastness,” “autonoetic” (self-knowing) consciousness (Tulving, 1985), or “re-experience” (Wheeler, 2000) of a past moment, often accompanied with feelings at that time. According to Tulving (2002), episodic memory is, in evolutionary terms, a late-acquired memory system that only humankind may have. Although the episodic memory system is a part of long-term memory, knowledge in this type of memory can be lost by interference or brain damage relatively easily compared to the other two, more basic, memory systems. Animals or infants may not have very developed episodic functions (Tulving, 1985). Scholars believe that episodic memory works together with semantic memory in most cases. For example, when you are asked about specific events or episodes such as the color of shirt a person wore yesterday, the respondent must rely on the recollection of a past episode about their clothing yesterday. Such memory traces are available in episodic memory. In addition, the respondent has also retrieved the meanings of the terms “color” and “yesterday” without relying on the conscious recollection of those words. Some researchers have questioned how episodic memory can be dissociated from semantic memory. Furthermore, it is still unclear how episodic memories can be incorporated into the semantic memory system over time (Foster, 2009). Wheeler (2000) pointed out that it remains difficult to identify, for example, whether a correct response of a student taking a test was due to recall from semantic memory or the episodic recollection of the moment the material was studied.

1.1.3 Working Memory

Working memory is a mental mechanism for the temporary processing and storage of internal and external information, as proposed by Baddeley and Hitch (1974), as a revision of Atkinson and Shiffrin's version of the short-term store (Baddeley, 2000; Gathercole & Baddeley, 1993). Working memory constitutes a multicomponent system rather than a single unitary system for the storages that Atkinson and Shiffrin conceptualized, which consists of a main control component called the "central executive" and two modality-specific registers, the "phonological loop" and "visuospatial sketchpad," as slave systems under the central control. The central executive is responsible for regulating activities in the working memory system, as well as interacting with other devices, such as the long-term memory (Gathercole & Baddeley, 1993). Baddeley and Hitch (1974) hypothesized that this master device has a finite capacity and requires attention for activities to carry out. Therefore, as with the earlier concept of the short-term store, the amount of information that working memory can deal with at a time is limited.

The phonological loop is one of the two subsidiary functions of the working memory system, and it handles verbal input. This function consists of two components, the "phonological store" and "articulatory rehearsal processes." The phonological store holds phonologically coded information from inputs for a very limited time, but the fading information in the phonological store can be refreshed and retained by articulatory rehearsal processes. Non-auditory inputs, such as written words or meaningful pictorial images, are converted into phonological codes by the rehearsal processes so that they can also be retained in the phonological store. Spoken inputs enter the phonological store directly without any encoding. The presence of the phonological loop is supported by the articulatory suppression effect (Foster, 2009; Gathercole & Baddeley, 1993), in which input is less likely to be held when the subject is memorizing material while continuously pronouncing words unrelated to it, such as "one, two, three, four."

The visuospatial sketchpad, the other slave function, deals with the processing; and brief storage of; visual and spatial information. Studies on the visuospatial sketchpad have been relatively scarce compared to those focused on the phonological loop. One possible reason for the lack of research on this system could be that researchers assume that this device does not play a central role in major cognitive developments such as language acquisition (Gathercole & Baddeley, 1993). Baddeley's (1990, 2000; Gathercole & Baddeley, 1993) initial conceptualization of the visuospatial sketchpad originated from observations by Brooks (1967), who found that a conflict in visual processing invited inaccurate short-term memory recollection.

In Brooks' task, participants were first shown a blank four-by-four matrix with one of the squares marked as the starting cell. The matrix was soon taken away. Participants were encouraged to hold the matrix mentally. Then, they were given the following instructions either only verbally or both verbally and via a sentence written on a card: "In the starting square, put a 1." "In the next square to the right, put a 2." "In the next square up, put a 3." The instructions continued until the participants were asked to put an 8 in a square of the matrix. The participants' task was simply to repeat the messages verbally immediately after each message was given. The results showed that the participants in the listening only condition repeated the messages more accurately than those in the listening and reading condition did. This finding suggested the existence of a visual short-term memory, which could be explained by the way in which the simultaneous processing of visual inputs from reading and picturing mental images depleted the mental resources responsible for visual processing. Later, Baddeley replicated this task and obtained similar results (Baddeley, *et al.*, 1975).

1.1.4 Implicit and Explicit Memories

Whereas the retrieval of information from memory is commonly assumed to involve conscious or explicit operations under traditional memory tests (e.g., free recall, cued recall, or recognition;), there also be another type of memory assumed from seemingly non-conscious performance on more perception-oriented tests, such as priming tasks (Graf & Schacter, 1985; Schacter, 1987; Schacter, 1992). Graf and Schacter (1985) contrasted these two types of memory as conceptual terms, the "implicit memory" and "explicit memory." Implicit memory is revealed when stimuli during a task facilitates access to particular pieces of information in memory without any intentional recollection. Explicit memory, by contrast, refers to the intentional or conscious remembering of stored information. Moreover, the findings on normal priming performance by amnesic patients struggling with recall tasks suggest that these two memories rely on different underlying systems (Diamond & Rozin, 1984; Graf & Shimamura, 1985; Graf *et al.*, 1984; Shimamura, 1986; Shimamura & Squire, 1984, 1987). For example, Graf, Squire, and Mandler (1984) presented amnesic participants with a list of words taken from a contemporary English dictionary while being asked whether the words in the list shared any of the same vowels, as well as to evaluate the pleasantness of each word on a 5-point Likert scale. Immediately after the list presentation, participants were asked to recall the words they saw in any order (i.e., free recall). Following the free recall task, they then received a list of cues composed of the first three letters of each word in the previous list (e.g., DEF____ for DEFEND) and were asked to fill in the following letters based on the first word that came to mind (i.e., primed recall). A comparable number of alcoholic patients performed the same tasks as control subjects. The results revealed that amnesic participants performed the primed recall task at a level similar to the control subjects, but their performance on the free recall task was significantly impaired. Typically, this line of research has found relatively intact performance among amnesic subjects on word-stem priming despite their deficits in recall performance. This finding suggests that memory holds decontextualized semantic-lexicographic information and contextualized episodic information separately. Therefore, a functional dissociation between non-conscious implicit memory and conscious explicit memory is supported.

1.1.5 Perceptual Representation System

The term perceptual representation system (PRS) also refer to implicit memory (Schacter, 1992). Although the great majority of investigations into implicit memory have been conducted on a priming paradigm based on written words, researchers have also found similar priming effects in other domains, such as visual object identification or auditory recognition. According to Schacter (1992), as a part of implicit memory, PRS is considered to have three subsystems, each associated with (a) a visual word form, (b) a structural description, and (c) an auditory word form. While these subsystems are responsible for the perceptions in different modalities of input, they share common characteristics as they operate at a pre-semantic level and are involved in non-conscious expressions of memory. Here, the term "pre-semantic" refers to a type of processing in which semantic access is not necessary. The term "non-conscious" refers to a type of retrieval without relying on a particular episode of past experience.

One of the assumptions that PRS operates at a pre-semantic level originated from a case study about a brain-damaged patient who could read printed words fluently but did not comprehend them. Sartori, Masterson, and Job (1987) tested the language performance of a brain-damaged English housewife, aged 68, by using a simple reading task, a synonym-matching task, and a naming task. In a reading task, the subject, was able to read aloud several newspaper articles with an average of 80% accuracy. However, her accuracy on a synonym-matching task, in which she needed to associate words from the articles with synonyms, was found to be at the level of chance. Moreover, an object-naming task revealed that she had difficulty naming common objects, such as a key or a watch, in spite of normal responses via verbatim repetition about those words.

In this way, case studies looking at verbal abilities of brain-damaged patients (e.g., the case above;), together with priming effects among amnesic subjects on word-stem completion tasks, where subjects respond to three-letter stems with the first word that comes to mind, led Schacter (1992) to hypothesize that PRS operates at a pre-semantic level and that there must be an underlying mechanism responsible for the recognition of written words, which can be referred to as the visual word form subsystem of PRS.

The second subsystem, the structural description, was hypothesized in priming effects in the domain of visual object processing. An object decision task devised by Schacter, Cooper, Delaney, Peterson, and Tharan (1991), for example, illustrates pre-semantic operations for visual objects. Schacter *et al.* (1991) used 20 drawings of unfamiliar three-dimensional objects. Among them, half of the objects were structurally sufficient and could exist in a three-dimensional (3D) body whereas the other half could not. In an initial task, 80 university students saw the objects one by one, judging whether each object faced the left or the light. This left-right decision was, however, in fact designed to focus the students on studying the objects closely: The main purpose was just to expose the objects to the participants. Half of the participants saw each object four times, and the other half observed them only once. After the study phase, participants were presented with the studied objects, as well as several unstudied objects, very briefly to decide whether each object was possible as a proper 3D object. An analysis of the accuracy of their responses revealed that the participants judged studied possible objects more accurately than non-studied possible objects regardless of the number of exposures, indicating the presence of priming. However, this was not the case for the “impossible” objects. The decision accuracy was significantly poorer for “impossible” objects than for “possible” objects. Moreover, the number of exposures did not contribute the judgments on the “impossible” objects more accurate. These differential effects of priming on “possible” and “impossible” visual objects suggested that a subsystem for visual objects was present in PRS. Regarding the failure to obtain evidence for priming on “impossible” objects, Schacter (1992) later explained that “the failure to observe priming of impossible objects has been replicated many times, and may indicate that it is difficult to form an internal representation of the global structure of an impossible object” (p. 248).

The third PRS subsystem, the auditory word form, is assumed based on neuropsychological evidence on people who reported having “word meaning deafness” (Kohn & Friedman, 1985). Such patients typically show difficulties understanding spoken words while being able to repeat them, or understand them when written. For example, in Franklin, Turner, Lambon-Ralph, Morris, and Bailey’s (1996) case study with a patient they called Dr. O, the subject received a list of words to repeat and verbally define. Out of 100 presented words, Dr. O was not able to give an appropriate definition for 55 of the words he heard. He answered, for instance, "woman is the grown up of the male species" for the word woman, "science, science is the disease of our generation" for the word science, and "soul, soul, soul, I don't know what it is; I should know that" for the word soul (Franklin *et al.*, 1996, p. 1150), although he could successfully repeat those words individually. The words were written down for him when he made erroneous responses, and then he was able to produce an appropriate verbal definition. To check his basic linguistic functions, the researchers also gave Dr. O a phoneme discrimination task in which he had to repeat in response to 40 consonant-vowel-consonant (CVC) real (e.g., buck-bud) and nonword (e.g., peb-beb) minimal pairs, as well as a picture-word decision task in which he was to decide whether the auditory presented word was a correct name for the picture presented visually. Dr. O's performance on these tasks was considered to be within the normal range in terms of those of average elderly participants. Therefore, Franklin *et al.* (1996) concluded that Dr. O's word deafness did not result from the early stages of auditory processes but was associated with the mappings between spoken lexical forms and meaning, which could be characterized as word meaning deafness.

The auditory word form subsystem of PRS is also supported by auditory priming tasks using healthy participants, although the amount of such research appears to be limited. In one of those studies, Shacter and Church (1992) presented a list of 48 familiar words to university students aurally. In the presentation of the words, half of the participants were asked to judge the pitch of each word (high, medium-high, medium-low, and low), which was in the non-semantic encoding condition, and the other half were instructed to make category judgments (animals, food, places, and occupations), which was in the semantic encoding condition. After the studying phase, participants' memory was tested by giving half of the presented words together with non-studied novel words, but this time, the latter half of each word was masked with white noise, which was a degraded presentation. Half of the students in each encoding condition were involved in an auditory identification task, and their task was to write down the very first word that came to mind in response to the stimulus degraded by white noise. This step served as a pre-semantic priming task. The other half of the students in each encoding condition participated in an auditory recognition task in which they were to make a yes-no decision on whether the degraded stimulus was a part of previously presented words. This step served as explicit recognition, and the participants had to rely on their memory traces. As a result of comparisons in response accuracy on each task between the two encoding conditions, the researchers found that semantically encoded participants outperformed non-semantically encoded participants on the explicit recognition task, whereas there was no significant difference on the auditory identification task. According to the researchers, these results implied that auditory priming relied less on semantic processes than on explicit retrieval and thus could be viewed as support for the presence of the auditory word form subsystem in PRS.

1.2 Processing Theories

Hyde and Jenkins (1973) found that the probabilities of subsequent recall of words were significantly different when subjects were asked, (a) to rate the words for pleasantness, (b) to estimate the frequency of each word in the English language, (c) to detect the occurrence of the letters "e" and "g," (d) to decide the part of speech of each word, and (e) to decide whether each word could come after a definite article (i.e., "the"), at the time of encoding. Their findings prompted them to theorize a processing-oriented framework for the human memory system because early storage theories did not have enough explanatory power for these phenomena. Processing theories regarding the human memory system assume that types of processes performed during learning are responsible for retention of input and that similarities in processing between encoding and retrieval influence successful recall.

1.2.1 Counterarguments to the Storage View

In contrast to multi-storage explanations of the human memory system, processing theorists argue that the type of processing that a subject uses at the time of encoding is important for durable memory, apart from structures per se. An influential modern account of this view is the levels-of-processing framework (Craik & Lockhart, 1972), also known as the depth-of-processing framework. Memory researchers have noted that this framework is one of the most popular approaches (Bower, 1977, p. 442) in the human memory literature, "enshrined [in] virtually all introductory human memory textbooks" (Roediger, 2008, p. 230), and it has had an "overwhelming impact on the field" (Roediger *et al.*, 2002, p. 319). In establishing their new processing-oriented approach, Craik and Lockhart (1972) asked several questions about the theoretical characteristics of human memory storage, especially those for the short-term store.

First, Craik and Lockhart (1972) claimed that any underlining mechanisms for a limited-capacity nature of the short-term store were unexplained. In multi-store models, the short-term store was said to be limited in its capacity, but they asserted that it was unclear what the "capacity" referred to exactly—a storage capacity, a processing capacity, or an interplay between these two. They also pointed out an incongruity between theories and empirical data on the short-term store, in which the number of items the memory can deal with at a time was hypothesized to be two to nine (Miller, 1956), but subjects were typically capable of reproducing 20 or so words accurately when the words in a memory-span test formed a meaningful sentence. These arguments could suggest a crucial theoretical flaw in the short-term store, given that a major feature of this store is its definition as having a limited capacity.

Second, they mentioned issues regarding the modality of codes in the short-term store. Initially, proponents of the storage view argued that the short-term store and the long-term store could be theoretically dissociable from one another because the modalities of information of each of these stores dealt with were basically different. (Conrad, 1964; Baddeley, 1966). Information in the short-term store was largely phonemic, whereas that held in the long-term store was basically semantic. However, subsequent research suggested a possibility that a mode of information may vary in the short-term store. Kroll, Parks, Parkinson, Bieber, and Johnson (1970), for example, demonstrated that input could sometimes be held visually in the short-term store even when stimuli were presented verbally. Furthermore, Shulman (1972) discovered that the short-term store could deal with semantic information depending on the paradigm used. Citing these counterexamples to the initial theorization of the storage view, Craik and Lockhart (1972) warned theorists of the storage view that they should not argue that coding in the short-term store might, in fact, be flexible because such a position would cancel their fundamental theoretical grounds, in which memory storages were distinguished by their unique characteristics. Rather, they emphasized, the coding issue should be viewed in terms of the processing demands imposed by the material to be remembered. That is, coding in an acoustic form may be adequate in some cases, but processing at a semantic level may be needed or advantageous in other cases.

Third, Craik and Lockhart (1972) focused on issues related to the forgetting characteristics of the short-term store. They insisted that the rate of decay of information in a particular store should be invariant under different experimental settings provided that types of memory stores were to be distinguished by the retention rate. However, researchers revealed that short-term retention in paired-associate learning tends to be greater than that in free recall paradigms (Craik & Lockhart, 1972, p.674). Furthermore, they pointed out the fact that different studies yielded different estimates on the short-term retention of visual iconic stimuli, suggesting unexpected methodological effects on these estimates, which might be due to the different paradigms or materials employed in those studies. Moreover, they posed a basic question regarding the storage view, noting that we can remember pictures, music, and faces long after the learning period. In the storage view, only the sensory register dealt with such nonverbal information. Consequently, it was, in fact, difficult to draw a line between the sensory register and other memory storages—even when only looking at an initial phase of input processing.

In summary, Craik and Lockhart (1972) argued that the storage view did not provide a satisfactory rationale for its grounds, although they acknowledged that mental operations such as attention and rehearsal were necessary for explanations of the human memory phenomena. Their idea was that the same external stimulus could be analyzed and coded multiple times at different levels. For example, a word may be analyzed in terms of visual and phonological as well as semantic features, and the retention of those different representations may vary depending on their characteristics. Therefore, to explain the rate of retention, it is useful to focus on how deeply the stimulus is processed, rather than the characteristics attributed to memory stores.

1.2.2 The Levels of Processing (Depth of Processing) Framework

In the levels-of-processing framework (Craik & Lockhart, 1972), it is assumed that an external stimulus (i.e., input) is rapidly analyzed by perception through a number of stages each at a deeper level than the previous stage. The earlier stages are thought to consist of analyses of surface features such as lines, brightness, pitch, and loudness, whereas the later stages are more involved with analyzing semantic features and matching the input against existing knowledge from past learning experiences. Therefore, cognitive or semantic analyses, such as pattern recognition or the extraction of meaning, are performed at the later stages of the analysis. These successive stages are referred to as "depth of processing" (Craik & Lockhart, 1972, p. 675). It is also assumed that memory traces are generated as byproducts of input processing. Therefore, analyses at deeper levels are associated more with elaborate and durable traces than those at shallower levels. At a deep level that involves cognitive processing, the subject can handle the material with learned rules and knowledge, thus realizing more elaborate coding. Highly familiar stimuli can arrive at a deep level more rapidly and are better retained than unfamiliar stimuli due to the assistance of existing knowledge. Therefore, the speed of processing itself does not necessarily determine retention, although peripheral factors, such as the amount of attention paid to the stimulus or time available for processing, influence the depth to which information is processed, as well as its retention in the initial learning process.

In this way, the levels-of-processing framework (Craik & Lockhart, 1972) emphasizes that retention is a function of depth of processing at the time of encoding. Memory traces are byproducts of input processing and materials are processed in an increasingly elaborate manner as perceptual analysis proceeds. Therefore, a deeper analysis leads to more elaborate and persistent memory traces than shallower analyses do. In the following sections, we will look at the depth-of-processing effects replicated in various research paradigms (Craik & Tulving, 1975; Fay, *et al.*, 2005; Jacoby & Craik, 1979; Kane & Anderson, 1978; Morris *et al.*, 1977; Seamon & Murray, 1976).

1.2.2.1 Effects of Decision Difficulty on Retention—Jacoby and Craik (1979)

Jacoby and Craik (1979) investigated whether initial decision difficulty influenced subsequent performance on cued recall. In their Experiment 1, participants were given cards, each with one word printed on one side (the "focus word") and two words printed on the other side. The participants' task was to study the focus word and then pick one of the two words on the other side that was related more to the focus word. The two words on the reverse side had semantically high or low levels of associations with the focus word or were not related to the focus word at all. For example, the two words could be both highly related to the focus word, one highly associated and one lowly associated, or one highly associated and one unrelated. As a result, six patterns were possible as a whole: high–high, high–low, high–unrelated, low–low, low–unrelated, and unrelated–unrelated.

The researchers assumed that a decision to pick a word was difficult in the high–high, low–low, and unrelated–unrelated cases and that it would be the easiest in the high–unrelated case. The words picked by the participants were referred to as target words. In a subsequent memory test, each participant was asked to recall the word he or she picked from a focus word that was given as a cue. The results showed positive correlations between the levels of decision difficulty and cued recall scores: Target words in the high–high condition were most accurately recalled, followed by those in the high–low condition and then the high–unrelated condition. Similarly, target words in the low–low condition were recalled more accurately than those in the low–unrelated condition. However, unlike the initial prediction, the mean recall rate was higher for the high–high condition than for the low–low condition, and target words in the unrelated–

unrelated condition were the least recalled among all conditions. From these results, the researchers suggested that retention might depend on both the depth of encoding and the effectiveness of retrieval operations. In this sense, less meaningful items, such as those in the low–low condition and the unrelated–unrelated condition, might be disadvantageous in terms of the ease of their reconstruction at the time of recall.

In Experiment 2, they further investigated the effects of initial decision difficulty on subsequent recognition. In this experiment, decision difficulty was manipulated by asking a category question (e.g., “Is the word a type of tree?” or “Is the word a type of cloth?”) then presenting a target word. There were 48 target words with three different levels of similarity to a particular category. Based on normative lists for general categories used in psychological experiments, one-third of the target words were taken from the top third of the normative lists (i.e., highly normative), one-third was from the middle third, and the remaining one-third was from the bottom third of the normative lists (i.e., less normative). The researchers postulated that category decisions would be easiest for high-ranking exemplars, more difficult for middle exemplars, and most difficult for low-ranking exemplars. In the target word presentation, decision latency was measured for each target word. After studying all 48 target words with category questions, participants received a recognition sheet with the 48 target words mixed randomly with 144 distractor words that had not been presented in the studying phase. Their task was to check, exactly, the 48 items they studied. The decision latencies increased systematically from the high to low exemplars, suggesting, arguably, that decision difficulty increased from high to low. Moreover, recognition scores rose systematically from the high-ranking exemplars to low-ranking exemplars. Together with the findings in Experiments 1 and 2, Jacoby and Craik (1979) concluded that greater decision difficulty is associated with higher levels of retention.

1.2.2.2 Effects in Word-stem and Word-fragment Completion Paradigms—Fay, Isingrini and Clarys (2005)

Fay, Isingrini, and Clarys (2005) examined the effects of depth-of-processing manipulations on cued recall in two types of word completion paradigms. The paradigms employed consisted of word-fragment completion, in which participants produced a word with a stem (e.g., sol_ _ _ _ for soldier), and word-fragment completion, in which participants produced a word with some letters presented randomly (e.g., _o_ _i_r for soldier). In the initial learning phase, participants were given three word lists of 15 items. Each word list was presented in different learning conditions.

In the perceptual study condition, participants read a word while counting the enclosed spaces within the letters of the word (e.g., so[]ld[]i[]er). This condition was assumed to be performed solely on the basis of the visual pattern presented by the word and without lexical processing or conceptual processing, which was considered to be the shallowest. The lexical decision condition, another shallow processing condition, was expected to engage lexical processing, but only at a minimal level of conceptual processing. In this condition, nonwords were added to the word list, and participants were told to decide if each item was a real word or not. Finally, in the semantic study condition, participants were instructed to rate the pleasantness of the presented word on a 5-point Likert scale (i.e., very unpleasant, unpleasant, neutral, pleasant, and very pleasant). This condition was assumed to require both lexical and conceptual processing, and thus the deepest processing.

After learning 45 target words (i.e., 15 in each study condition), an explicit memory test was administered. In this test, an answer sheet that consisted of 45 letter cues (i.e., fragments or stems) corresponding to the studied words was provided. Participants were told to complete the cues without guessing. As a result, the researchers found “a standard depth-of-processing effect” (p. 795) in the word-stem paradigm. Recall performance with semantic encoding (i.e., pleasantness rating) was better than that with lexical encoding (i.e., word–nonword lexical decision). Moreover, recall performance with lexical encoding was better than that in the perceptual study condition. However, in the word-fragment paradigm, recall performance in the lexical encoding condition and the semantic encoding condition did not show a significant difference, while a slight advantage of the lexical encoding condition over the perceptual encoding condition in recall probability was found. Therefore, the researchers concluded that different processes might be involved in the word-fragment completion paradigm and the word-stem completion paradigm. The researchers suggested that a word-stem completion task required the semantic components of words more than the word-fragment completion tasks did. Because a word-fragment completion task required the subject to fill in letters in a randomized fashion, time largely spent searching for the word rather than for processing it semantically.

1.2.2.3 Effects of the Meaningfulness of Single Words—Seamon and Murray (1976)

The depth-of-processing framework was also confirmed by Seamon and Murray (1976) who investigated the probabilities of the subsequent recall of English words with varying levels of meaningfulness. The depth-of-processing framework maintains that semantic processing is carried out at a deep level in input processing. Following this hypothesis, in this study, the researchers assumed that the meaningfulness of words would affect the performance of free recall. Highly meaningful words could always be processed more deeply than less meaningful words, thus, they are retained well.

In their experiment, American undergraduates participated in incidental learning tasks involving binary decision making about words. The material included 60 English nouns taken from the word norm list devised by Paivio *et al.* (1968). The list provided indices of concreteness, imageability, and meaningfulness for 925 English nouns. For example, one of the most meaningful words in this list, “FRIEND”, was rated 9.12 on a questionnaire. The examples “INFANT” and “PERMISSION” each had ratings of 7.20 and 4.68, respectively. Half of the stimulus words were high in meaningfulness (M = 7.11), and the other half were low in this dimension (M = 4.70). Participants were equally divided into three groups: two incidental learning groups and a control group. The two incidental learning groups consisted of the structural decision condition and the semantic decision condition. The same word list, containing 30 meaningful and 30 less meaningful words, was used for all conditions. Participants were given words with different levels of meaningfulness (i.e., high and low) from the list interchangeably one by one in a fixed order.

Participants in the structural decision condition were told to attend to the position of their lips while subvocally repeating each word. Decisions should be made regarding whether their lips were touching at either the beginning (e.g., MANE) or end (e.g., TOMB) of a word, at both the beginning and the end (e.g., BOMB), or neither at the beginning nor the end (e.g., CLOCK). Participants in the semantic decision condition were instructed to focus on the meaning of each word and to decide whether the word was a general (e.g., TOOL) or a specific (e.g., HAMMER) instance of a semantic category. Control participants saw the stimulus words with a notification that a free recall test would be given later. For the two incidental learning groups, response times were recorded trial by trial. Moreover, immediately after the last trial, all participants were asked to write down as many of the presented words as possible in any order within three minutes.

The results indicated that participants in the semantic group and the control group recalled highly meaningful words more than those that were less meaningful. Meaningfulness did not affect recall in the structural condition. Furthermore, an examination of decision latencies revealed that the semantic group responded significantly more quickly than the structural group, which was consistent with the claim made by Craik and Lockhart (1972): Processing time is not important, and highly meaningful items will be processed faster and retained better than less meaningful items. Therefore, the researchers concluded that the meaningfulness of words could be a primary control variable for depth-of-processing effects.

1.2.2.4 Effects of the Meaningfulness of Words in Sentences—Kane and Anderson (1978)

The depth-of-processing effect was also investigated in a sentence-completion paradigm created by Kane and Anderson (1978). They compared the recall probabilities of words in sentences among participants who guessed the last word of a sentence and others who simply read a whole sentence. Two types of sentences were prepared for the experiment. One was a sentence frame in which everyone could guess the same last word to complete the sentence, and in the other one people could come up with various words to complete the sentence. What they called a "determined" sentence stem was an incomplete sentence, such as "The executive sat behind his large oak _____," in which people would commonly insert "desk". In contrast, an "undetermined" sentence stem, for example, could be "The executive went to the shop for a new _____," and many different words could be possible for the final position, but inserting the word "desk" was unlikely.

In their experiment, American undergraduates were divided into two task conditions—sentence completion and reading only. Each participant was given one of the following three types of lists: (a) a determined list consisting of 12 determined sentence stems, (b) an undetermined list with 12 undetermined sentence stems, and (c) a mixed list containing 6 determined and 6 undetermined sentence stems. Participants in the sentence completion condition saw a sentence with the subject noun underlined and a blank in the place of the last word (e.g., "The dove is a symbol of _____"; "The dove appeared when the magician said _____"). Their task was to read the sentence frame aloud and guess the word the experimenter would come up with to complete the sentence. After guessing, a completed sentence was presented, and participants read the correct last word aloud. For this group, the researchers expected that undetermined sentence stems would induce errors during learning because the correct words were almost always different from what the participant initially thought of, whereas this was less likely to happen for determined sentence stems. Therefore, undetermined sentence stems would hinder an elaborate encoding of words. Participants in the reading-only group saw a completed sentence and simply read it aloud once, which corresponded to a typical shallow processing condition in the depth-of-processing norm.

After completing the 12 sentences in a list, an immediate recall test was administered in both a forward test and a backward test. In the forward test, participants received a stack of index cards with the subject noun from one sentence typed on each card. Then, they read the word aloud and orally reported the last word of that sentence. In the backward test, participants were presented with the last word of the sentence on a set of index cards and then asked to report the subject noun of the sentence orally. Participants completed the two types of tests in a randomized order.

The results of an immediate posttest showed that the sentence-completion group, overall, recalled significantly more words than the reading-only group. Interactions between the types of tasks and types of lists were not significant, indicating that the sentence-completion group was always better than the reading-only group. Moreover, analyses on a follow-up posttest a week later revealed that recall performance on undetermined responses was significantly poorer than that on determined responses, as the researchers expected. Consequently, the researchers concluded that tasks requiring students to construct meaningful representations are advisable, especially in classes using verbal materials.

1.2.2.5 An Extensive Review on the Depth-of-Processing Framework—Craik and Tulving (1975)

Craik and Tulving (1975) inspected the depth-of-processing framework through 10 experiments. It is viewed as one of the most comprehensible papers about the effects of depth of processing on the retention of words addressing this issue from multifaceted angles.

First, Craik and Tulving examined the effects of the depth-of-processing framework (Craik & Lockhart, 1972) in a general way to determine whether words processed at a deep level are retained longer than shallowly processed words. A total of 20 college students participated in the first experiment. The encoding phase involved perceptual yes-no judgments on five types of questions, followed by a recognition test. The five levels were, from the shallowest to the deepest 1) presence (e.g., "Is there a word present?"), 2) structural (e.g., "Is the word in capital letters?"), 3) phonemic (e.g., "Does the word rhyme with ____?"), 4) category (e.g., "Is the word in the category ____?"), and 5) sentence (e.g., "Would the word fit the sentence ______?"). Each participant received 40 trials; four words applied to each of the 10 conditions (i.e., 5 levels x yes-no). On each trial, the question was read to the participant, and then the word was given. Participant responded by saying yes or no so that a voice key stopped a millisecond timer to measure the response latency. After exposure to all the 40 words, participants were unexpectedly given a sheet with the 40 original words plus 40 distractors and asked to check the words they had seen in the trials. The researchers summarized the results in four points. First, the rate of hits on the recognition posttest increased systematically from below 20% recognized for the words with the structure questions to over 95% correct for the words with the sentence questions. Second, the rate of recognition was better overall for the words associated with the yes responses than the no responses. In addition, this superiority increased with depth of processing; the gap was smaller for those concerning structural characteristics (e.g., upper case or lower case) and greater for those requiring semantic processing. Third, response latency increased systematically as questions required deeper processing. Finally, there were no significant time differences between yes and no responses.

The second experiment was concerned with a replication of the findings of Experiment 1 because the researchers suspected that the high accuracy in the sentence condition in Experiment 1 would cause a ceiling effect. The experiment procedure was basically the same as in Experiment 1, but the distractor choices in the subsequent recognition test were increased to 120, from 40 in the previous experiment. Moreover, the processing levels in the coding phase were reduced to three, questions concerning types (i.e., uppercase or lowercase), rhyme questions, and sentence questions in which participants were given a sentence frame with a word missing. Experiment 2 clearly replicated the results of Experiment 1. The accuracy associated with the sentence question decreased to 81% due to the increase in the distractor choices but was still best among the three processing levels. The accuracy also increased systematically from case decisions (i.e., the shallowest encoding) to sentence-matching decisions (i.e., the deepest encoding). However, the other characteristics of the findings in Experiment 1 remained unchanged.

Experiments 3 and 4 examined the generalizability of the findings in Experiments 1 and 2, in which the researcher sought to determine whether the same pattern of recall would be obtained under incidental and intentional learning conditions. The same three levels of encoding (i.e., case, rhyme, and sentence) were employed in both experiments. The participants of Experiment 3 were informed that the task would involve perceptual decision making before learning, but there was no mention of the subsequent recall task. The participants of Experiment 4 were, in contrast, told that there was a final recall task. As a result, the patterns in the previous experiments were replicated in both Experiment 3 and 4. Only the intentional participants in Experiment 4 outperformed the incidental participants in Experiment 3. The findings, to this point, indicated that (a) different encoding questions were associated with different response times, which could be interpreted as meaning that semantic questions induced an analysis of the presented word at a deep level; (b) positive (i.e., yes) and negative (i.e., no) responses were equally fast; (c) the rate of recognition increased as the encoding question dealt with more semantic features of the word; and (d) words given a positive response were associated with higher recognition performance, and these characteristics did not change whether the type of learning was incidental or intentional.

The previous findings supported the depth-of-processing hypothesis; however, since the response times increased as encoding questions became more complex, there was still a possibility that learning at a deep level promoted superior retention simply because the learners were exposed to the material for a longer time. Therefore, Experiment 5 investigated this issue by providing participants either a complex structural task (i.e., a complex task based on shallow processing) or a simple semantic task (i.e., a simple task based on deep processing). The researchers hypothesized that the complex structural task would take longer to complete than the simple semantic task but that the simple semantic task would result in better memory performance. The complex structural task involved a non-semantic yes-no decision about the pronunciation of a word. In this task, participants were presented a phonetic pattern, such as CCVVC, and then a word such as “brain” was given. The participants’ task was to judge whether the given word matched the phonetic pattern presented in advance. The simple semantic task included a sentence task that had been used in the previous series of experiments. In this task, a short sentence with one word missing was presented (e.g., “Near the bed she kept a ____”), and then a word was given (e.g., “clock”). The participants’ task was to decide whether the word would fit the sentence. A total of 24 American undergraduate students participated in this experiment, and they studied 40 target words in either of the two conditions. They were told that the task was a perceptual reaction time study but were not informed of the subsequent recognition test that included the 40 targets mixed with 80 distractors. The results supported the researchers’ hypothesis. Non-semantic consonant- vowel decisions were significantly slower (1.7 sec) than sentence decisions (0.85 sec), but subsequent recognition was significantly greater for sentence decisions (75%) than consonant-vowel decisions (53%). The researchers also showed that “yes” responses were recognized better than “no” responses at a significant level. Experiment 5, therefore, confirmed that the qualitative nature of a task, but not the amount of processing time, influenced memory performance.

Although qualitatively enhanced encoding was found to be important for retention in Experiment 5, it remained unclear why positive responses had resulted in better retention than negative responses. According to the theory, a word processed at the same level in depth should be retained to the same extent regardless of the types of responses. Experiment 6, therefore, compared the retention of words between positive and negative responses at the same encoding level. In this experiment, 16 American undergraduates were to make yes-no judgments regarding questions (e.g., “Taller than a man?”) about words (e.g., “steeple”—yes; “child”—no). As in the previous experiments, participants were told that this was a perceptual reaction time study and were not informed of the final retention test. Questions were formed in eight dimensions; size, length, width, height, weight, temperature, sharpness, and value. Each participant was asked questions about two of those dimensions. For each dimension, four questions yielded a positive response, and four yielded a negative response when participants were asked about a total of 16 target words. For half of the participants, the question was reversed in sense in such a way that “Taller than a man?” became “Shorter than a man?” so that a correct response was altered. After completing 16 question trials, participants were asked to recall the target words. The results showed a different tendency from the previous findings. Words given “yes” responses were recalled at a probability of .36, while words given “no” responses were recalled at a probability of .39, indicating that words in the two conditions were equally well retained. These results made the researchers hypothesize that it was not the type of response given to the presented word, rather, the richness or elaborateness was responsible for the subsequent retention, and negative decisions in Experiment 1 though Experiment 4 might be improperly formed in terms of those associated with rather poor encodings of the presented words.

Experiment 7 focused on elaborate encoding in response to the “unexpected” results of Experiment 6, which showed that positive and negative responses in encoding yielded an equal level of retention in contrast to superior retention rates associated with positive responses that were consistently found in Experiments 1–4. To manipulate the levels of elaboration in encoding, in Experiment 7, three types of sentence frames that varied in terms of sentence complexity were used to encode questions: simple (e.g., “She cooked the ____”), medium (e.g., “The ripe ____ tasted delicious”), and complex (e.g., “The small lady angrily picked up the red ____”). In total, 20 participants were given a set of 20 sentence frames and target words that were randomized with respect to their complexity levels and types of responses. In each encoding trial, the participant saw a sentence frame and gave a yes-no decision in response to the word given afterward. The retention of words was measured on a subsequent surprise free recall test followed by a cued recall task in which participants recalled words while looking at a sentence frame. The results were twofold. First, words associated with positive responses were well retained overall, but the gap in the retention between positive and negative responses was small in free recall, whereas there was a fourfold difference in cued recall. Second, in both free recall and cued recall, the rate of word retention increased systematically from words associated with simple to complex sentence frames. These results indicated that the type of recall might influence the superior retention associated with positive responses, and more importantly, more complex or elaborate sentence frames could lead to higher retention. The researchers argued that these results did not meet their original expectation that memory performance would be determined only by nominal perceptual levels of input analysis. In the present experiment, semantic processing for the target words were necessary in all conditions, but in fact, there were differences in performance at different complexity levels. Regarding the results, the researcher explained that the learners could not form a meaningful image of the complete sentence if a presented word did not fit the sentence frame, leading to poor retention because less meaningful memory traces decay easily.

Experiment 8 concerned encoding elaboration from another viewpoint. For poor performance in retention based on shallower encoding, especially those associated with case questions or rhyme questions in the previous experiments, the researchers suspected that such poor performance, in fact, happened because similar questions (e.g., “Is the word in capital letters?”) had been given each time. In contrast to semantic questions, each of which consists of a variety of words, the question patterns in both the case questions and rhyme questions were limited. Similar encoding questions for many target words might not be as effective as retrieval cues because each of them was difficult to identify in episodic memory as a single encoded event. Therefore, in this experiment, each group received a different number of trials about different question types: Group 1 received 4 case questions, 16 rhyme questions, and 40 category questions. Group 2 received 16, 40, and 4, respectively. Finally, group 3 received 40, 4, and 16, respectively. The prediction was that the subsequent recognition performance would be enhanced with a smaller set size, especially for the case-level questions. It was revealed, however, that recognition accuracy with case encoding was not changed across different set sizes, which was also the case for the deepest category encoding. Only the intermediate rhyme encoding showed a gradual increase in subsequent recognition by the manipulation of set size. Overall, the recognition rate attributed to the minimum set size (i.e., 4 sets) on case encoding was lower than that of the maximum set size (i.e., 40 sets) on rhyme encoding (.43 vs. .60), and the recognition rate based on the minimum set size on rhyme encoding did not show a remarkable advantage to that of the minimum set size on category encoding (.80 vs. .79). Therefore, it was concluded that reducing the number of target words did not contribute to the identification of a particular encoding event in episodic memory and that the encoding type played a robust role in the subsequent retention.

Experiment 9 addressed the ecological validity of the theory, and the researchers looked into whether the depth-of-processing effect could be found in a classroom environment. In the encoding phase, 60 words were presented, one at a time, through a projection screen for 1 second with a 5-second interval. All students saw the same words in the same order, but different students were asked different questions about each word. The questions were given in booklets, which consisted of six versions in combination with three encoding types (i.e., case, rhyme, and category) and two response types (i.e., yes and no). Therefore, for example, if the presented word was "SPEECH", one student would be asked, "Is the word in capital letters?"; a second student, "Is the word in small in print?"; a third, "Does the word rhyme with EACH?"; a forth, "Does the word rhyme with TENSE?"; a fifth, "Is the word a form of communication?"; and a sixth, "Is the word something to wear?" Students were notified that a recognition test would be conducted after seeing all words, and the researchers emphasized that their main task was to remember the words. Therefore, the encoding was intentional. A subsequent recognition test included 60 target words and 120 distractor words in print in a randomized order, and the students were asked to circle all the words they had seen in the presentation. Even though the experimental conditions were largely different from those employed in the previous lab-based experiments, the results showed a typical depth-of-processing effect: Words with rhyme encoding were remembered more than those with case encoding, and words with category encoding were well recognized than those with rhyme encoding, each at a significant level. Consequently, this classroom demonstration demonstrated the robustness of the basic phenomenon in question.

The final examination, Experiment 10, further investigated the generalizability of the depth-of-processing effect by manipulating motivational aspects. The researchers suspected that poor memory performance at a shallow level of encoding would result from the fact that subjects were poorly motivated and thus simply did not rehearse the presented words. Therefore, in this experiment, participants were offered rewards based on the results. The procedure was the same as in Experiment 9, but participants were informed that they would be paid according to the number of words they recognized correctly. Participants were divided into three groups. The group 1 participants were notified that they would be paid 1 cent, 6 cents, and 3 cents for case, rhyme, and category words, respectively. The group 2 participants were paid 3 cents, 1 cent, and 6 cents, respectively. Finally, the group 3 participants were paid 6 cents, 3 cents, and 1 cent, respectively. However, the results revealed a common pattern of the depth-of-processing effect, indicating that the differential reward manipulation had no effect. The result itself might be a mere confirmation of the robustness of the phenomenon, but it provided an educational implication that even very motivated learners could fail in learning when the materials are not processed at a sufficient level.

Through 10 well-designed experiments, Craik and Tulving (1975) revealed the characteristics of the depth-of-processing effect in great detail. Theoretical support for the original Craik and Lockhart's (1972) paper can be summarized in that greater memory performance with a deep level of encoding was robust as it was replicated in a series of experiments in the present research, and it was not disturbed by the level of motivation. Meanwhile, some of the results suggest that revisions might be needed to refine the original theory. For example, the complexity of encoding contexts was found to influence the subsequent retention even at the same level of encoding. Furthermore, the retention rates associated with positive (i.e., yes) responses were better than those with negative (i.e., no) responses. These results imply that levels of elaboration and meaningfulness should be incorporated into the original theory.

1.2.2.6 Effects of Nonverbal Materials—Baddeley and Hitch (2017)

The depth-of-processing framework has been researched predominantly by using verbal stimuli, and evidence has largely been brought by the retention of printed words. Meanwhile, early critics of the framework argued that people might employ different strategies in processing materials with different modalities. Arguably, human semantic processing involves, in fact, more complex phenomena than theories assume (Baddeley, 1978). The focal points in Baddeley and Hitch's (2017) paper involved comparing the effect sizes of the depth-of-processing phenomena between verbal and nonverbal materials and between different categories in the same medium, rather than manipulating processing levels.

In Experiment 1, 20 university students processed pictures of doors and concrete words either deeply or shallowly for a post-recognition test. Participants were given 24 pictures of doors and 24 concrete words consisting of animal names (e.g., a lion or a leopard) and household objects (e.g., a plate or a saucer). Half of the door pictures were predominantly brown, and half of the concrete nouns were printed mostly in brown. In the deep processing condition, participants judged whether they considered each door or word pleasant or unpleasant. In the shallow processing condition, participants judged whether each door or word was predominantly brown. The stimuli were presented, one at a time, on a PC screen for 1 second and another 2 seconds were allowed for immediate judgment. Immediately after the encoding phase, participants' recognition memory was measured. The task was to choose the learned item as "old" among four items, including a learned item and three distractor items. The results revealed a typical depth-of-processing effect for both materials: Deeply encoded items were significantly more recognized than shallowly encoded items on the posttest. However, the overall recognition probability was different between pictures and words, with a significant advantage for word materials. Moreover, the effect size was significantly larger for word stimuli than for visual stimuli; namely, the gap of recognition between deep and shallow processing was much larger for word stimuli.

The second experiment involved processing door pictures and two lists of concrete words composed of the names of people and occupations processed either deeply or shallowly. Twenty-four participants saw 30 door pictures, 30 names of people chosen from a telephone directory, and 30 words related to occupations. The stimuli were presented at a rate of 1 second each, followed by an interval of 1.5 seconds for judgment. The doors were presented in full color. The verbal stimuli were also presented in either red, green, brown, blue, or black. The deep encoding required a dichotomous judgment in pleasantness, in the shallow encoding condition, participants were asked to name the color of the presented item. A post-recognition test in a four-choice format was given immediately after the presentation phase. The results showed, again, a positive depth-of-processing effect for all types of stimuli. In addition, the overall recognition performance was significantly greater for occupations than for names, even though they were both verbal stimuli. A post-hoc comparison revealed the recognition probability for occupations outperformed that for names and doors at a significant level, whereas names and doors were equally recognized. As a result, this experiment indicated the degree to which the depth-of-processing effect might differ according to categories—even in the same modality.

Finally, Experiment 3, which consisted of three sub-experiments, compared one type of verbal stimulus and two types of visual stimuli. In Experiment 3a, pictures of doors and clocks, as well as food menus, were presented in a set of 30 each. Each item was presented for 1 second for encoding, and another 2 seconds were given for judgment. The deep judgment involved a verbal response (i.e., pleasant or unpleasant;), while the shallow judgment concerned size. In the shallow condition, for the door items, participants were asked to judge whether the presented door was larger or smaller than a typical domestic door. For the clocks, they were asked to judge whether the presented clock was larger or smaller than a typical mantel clock. For the food menus, they were asked whether the presented menu consisted of more than three words. In Experiment 3b, pictures of doors and phones, as well as subway station names, were presented in the same design as in Experiment 3a. However, different judgments were required in the shallow condition. In this experiment, participants were asked to report the dominant color of each stimulus. The results revealed the same pattern for these two experiments. Significant effects of depth of processing were found for all stimulus types. Nevertheless, verbal stimuli were significantly more susceptible than visual items. In contrast, there was no difference in visual materials. Experiment 3c was also conducted using the same design but with door pictures, scenes of nature, and names of unfamiliar animals. Judgments in the shallow processing involved reporting whether the stimulus had one predominant color or multiple colors. This experiment yielded a different pattern from the two previous experiments. Although the depth-of-processing effect was consistent for all stimulus types, there was no difference among these stimulus types, even though different modalities were employed.

In this way, Baddeley and Hitch (2017) demonstrated that, in addition to verbal materials, depth-of-processing effects can also occur with nonverbal visual materials. Furthermore, a post-hoc comparison of the effect sizes between verbal and nonverbal stimuli indicated that the effects were considerably more varied for verbal materials than those found with visual nonverbal materials. While the effects were relatively stable and modest across all the visual materials studied, ranging from .30 to .90 in terms of Cohen's *d*, verbal stimuli related to animal names and food menus were much larger in effect size (1.50 and 1.40, respectively;), than those related to unfamiliar animal names and people's names (.50 and .70, respectively) (Baddeley and Hitch, 2017, Fig. 7, p. 10).

How could such variation between different material types and modalities be accounted for? The researchers explained the observed data pattern with the theory of affordance (Gibson, 1986; Norman, 1988). The term "affordance," coined originally by Gibson, a cognitive psychologist, refers to values or meanings of things in the environment that all animals, including human beings, can perceive. In particular, Gibson mentions an interaction between the surface features of things and animals' spontaneous behaviors regarding the observed object. For example, a path affords an animal a chance to walk on it. A simple chair provides people a chance to sit, whereas an extremely innovative chair might not; rather, a step of a staircase or the edge of a fountain would more likely do so (i.e., sitting). Animals, as observers of things, can use the chair in as many ways as they can imagine. From this viewpoint, nonverbal materials could only be processed with superficial cues, such as size and shape, suggesting that levels of depth or elaborateness in processing are relatively level. Verbal materials, as opposed to nonverbal materials, could be processed in greater variations according to the degree or usefulness of the elaboration the word affords. Therefore, a familiar animal names, such as "leopard" will afford not only a visual image but also other characteristics, such as fierce, meat-eating, and agile, as well as any familiar product names related to the idea "leopard," whereas an unfamiliar animal name will not offer such a range of affordance. The richness of cues available in processing is a critical factor for elaborate encoding. In their discussion, Baddeley and Hitch (2017) noted that the depth-of-processing effect could generally be interpreted in terms of the concept of affordance; and that remarkable effects that are typically observed on printed materials could be viewed as the result of the effect capitalizing on the possibilities those materials offer.

1.2.3 The Transfer-Appropriate Processing Framework

While the depth-of-processing framework (Craik & Lockhart, 1972) emphasizes deep semantic access to input for durable memory traces, Bransford *et al.* (1979) argued that the nature of a test should also be taken into account for useful memory traces. The concept of transfer-appropriate processing (Bransford *et al.*, 1979; Morris *et al.*, 1977) is typically viewed as a complementary notion of the depth-of-processing framework and concerns the acquisition-test interaction. Moreover, successful post-learning retrieval is more likely if learning and testing situations share similar cognitive processes (Lightbown, 2013). A similar view also appears in the encoding specificity principle (Tulving & Thomson, 1973). This framework could complement the others, but it entails a much broader concept, including the generalization and understanding of the acquisition of knowledge and use, with the arguments on at least three shortcomings that the original depth-of-processing framework may possess (Baddeley, 1979).

First, transfer-appropriate processing raises a question regarding the claim that deep semantic processing yields better retention based on a counterexample found in the work of Morris, Bransford, and Franks (1977). In this experiment, 32 American university students processed 32 target words embedded in sentences in two conditions: semantic and rhyme. All target words were common five-letter words. In the semantic condition, students judged whether the combination of the sentence frame and the target words made sense. For each trial, the experimenter first presented a sentence frame orally (e.g., “The ____ had a silver engine”). The word “blank” was said to indicate the missing target word. After a 2-second pause, the target word, in the form of either a yes or no response (e.g., “train” or “eagle”), was verbally presented. In the rhyme condition, students judged whether the target word rhymed with the provided word. As in the semantic condition, a sentence frame was given orally (e.g., “____ rhymes with legal”), followed by the presentation of the target word (e.g., “eagle” or “peach”). The presentation mode (i.e., semantic or rhyme) and the type of answer (i.e., “yes” or “no”) were counterbalanced for each participant. After the encoding phase, an immediate recognition test was conducted in two ways. A semantic-orienting test was performed for 16 participants, and the other 16 participated in a rhyme-orienting test. In the semantic test, served as a standard test, the 32 target words mixed with 32 additional distractor words were randomly presented, and participants judged each word as either a new or an old item. In contrast, in the rhyme test, all 64 items to be tested were new. A “yes” response was to be made for a word rhyming with one of the target words presented in the encoding phase. Therefore, the memory assessed in this test was related to the abstract characteristics of input, such as sound patterns, that could be transferred from rhyme encoding, rather than nominal knowledge. The results partly supported the claim of the depth-of-processing framework. In the semantic recognition test, the semantic acquisition mode was significantly superior to the rhyme acquisition mode. However, in the rhyme recognition test, the memory performance based on the rhyme acquisition mode was significantly better than that in the semantic acquisition mode. This pattern was persistent in a delayed posttest performed a day after learning. Therefore, the researchers argued that the depth-of-processing framework lacks an adequate account of the nature of testing.

Second, transfer-appropriate processing suggests that the expertise of the learner might influence the ease with which deep processing is achieved. Although many proponents of the depth-of-processing framework have interpreted deep processing as elaborate processing (Craik & Tulving, 1975; Jacoby & Craik, 1979), transfer-appropriate processing stresses that elaborate processing for a certain type of input is possible if the learner has sufficient background knowledge. As seen in the aforementioned experiment (Morris *et al.*, 1977), rhyme processing, defined as a shallow processing in the depth-of-processing framework, could take advantage of rhyme testing. This finding could be interpreted as demonstrating that memory traces from shallow processing could also be durable to a certain extent, and thus there were virtually no inferior memory traces at processing levels. Taking these accounts into consideration, the researchers suggested, for example, that experienced poets and linguistics experts could process certain types of auditory input efficiently at a level at which they were dealing with semantically processed information. A key claim here is that deep processing brought through expertise must from that used by advocates of the depth-of-processing view.

The third claim is concerned with more general constraints on learning and memory that are clearly out of the scope of the depth-of-processing framework. The concept of transfer-appropriate processing is often viewed as synonymous with the encoding specificity principle. Bransford *et al.* (1979) distinguished between them in terms of the type of learning that the two positions focus on. Most of Bransford *et al.*'s paper discussed contrasts between the depth-of-processing framework and their original transfer-appropriate processing view, which is mainly concerned with remembering learned episodes. According to Bransford *et al.* (1979), it is this issue that the encoding specificity principle mainly deals with. In contrast, the transfer-appropriate processing framework includes a broader range of issues than just the processes necessary for better remembering, such as the further refinement of knowledge through interactions between encoding and testing to facilitate the transfer to new situations. Bransford *et al.* argued that, because the concept of transfer-appropriate processing offers more general perspective on learning and memory, the encoding specificity principle could be taken as a special subset of the transfer-appropriate processing framework.

Chapter II

SKILL ACQUISITION THEORIES

2.0 Introduction

Studies of skill acquisition are concerned with differences in knowledge between novices and experts, as well as with how novices gain expertise in a specific domain. Research interests extend over a variety of domains including cognitive and motor skills. Foreign language acquisition is one of many research topics concerning cognitive skill acquisition studies. This chapter reviews theories of skill acquisition that have been discussed mainly in the field of psychology, with a particular emphasis on those associated with language learning. While most of the present chapter deals with particular techniques for gaining advanced performance, it starts by introducing a widely held belief that no one can reach a professional level of expertise without prolonged intentional training, known as the theory of deliberate practice (Ericsson, 2006; Ericsson *et al.*, 1993).

2.1 Deliberate Practice

Ericsson (2006) and Ericsson, Krampe, and Tesch–Romer (1993) presented a life-long perspective on gaining and maintaining expert performance. Their theoretical framework, deliberate practice, holds that a certain innate talent may permit a child to exhibit superior performance in a domain, but this fades by adulthood without adequate deliberate practice, such as in language acquisition. That is, without exception, all expert performance can only be brought about by well-designed, extended efforts to improve the current level of performance.

This framework provides a set of three constraints essential to attain expertise at a high level. First, the resource constraint explains that trainees need dedicated supporters to become specialists. Such supporters could be parents or guardians, especially if the trainee is a child. Ericsson *et al.* (1993) noted that international-level performers often begin participating in activities in their domains as early as the age of three. Therefore, parents are responsible not only for providing early access to the equipment and facilities related to the domain but also for providing their children opportunities for deliberate practice. Second, the effort constraint addresses limitations of the amount of deliberate practice at each phase over the course of development. In the early stages of acquisition, the effective duration of deliberate practice is best estimated to be about an hour. A rapid increase in the amount or the quality of practice in the initial stages provides no benefit. Rather, it results in an unexpected fatigue that affects subsequent training. Here, the increase in the quality of training refers to incorporating variations into tasks and providing feedback on how to modify the current performance. After the trainee slowly adopts to a constant level of practice, increases in quantity become possible. The gist of the claim of this framework is that the level of performance a trainee reaches depends on the amount of deliberate practice. Therefore, the effort constraint refers, in the short term, the idea that optimal deliberate practice should be designed with a balance between effort and recovery, and in the long term, that trainees should patiently continue their deliberate practice with increases quantity. Finally, the motivational constraint explains that trainees must remain motivated on the premise that deliberate practice is not inherently enjoyable. Especially, it emphasizes persuading the trainee that deliberate practice itself plays an instrumental role in improving performance. This pointing out of instrumental values is particularly important for children and novice learners until they can internalize methods for assessing improvement and can concurrently monitor the effects of practice by themselves.

Ericsson *et al.* (1993) distinguished deliberate practice from other activities, such as playful interaction, paid work, and the observation of others, in that it consists of a set of efficient practice activities to maximize improvement. Furthermore, they emphasized that, with the exception of gaining height, all learners can attain any sort of distinguishing characteristics of expert performance through deliberate practice. They deny that innate talent is immutable, based on their hypothesis that many characteristics that seem innate are, in fact, the result of effortful practice over a minimum of 10 years. This theory would benefit all kinds of learners in terms of providing a basic mindset toward learning.

2.2 The Generation Effect

2.2.1 An Initial Finding—Jacoby (1978)

The generation effect refers to the advanced retention of items in memory gained by trying to generate a missing part of an input during learning, rather than by reading a whole material passively. The effect was first reported by Jacoby (1978), who compared the retention performance between "discovery" learning and "reception" learning. In this study, students' performance recalling words was assessed on a paired-word association paradigm through six learning conditions using a between-participants design. The first condition involved simply reading a list of pairs of words (e.g., foot–shoe) once. In the second condition, participants were asked to report the target word based on a cue (e.g., foot–s__e) for each word pair. In the third condition, participants read each intact word pair twice without time intervals. In the fourth condition, the researchers asked participants to read the same intact word pairs twice with a lag of 20 items. In the fifth condition, participants had two opportunities to generate the target word from the cue word without intervals. Finally, in the sixth condition, participants were told to find the target word from a given cue, and the same item was presented again after 20 other intervening items were shown. Immediately after the item presentation, participants' retention of words was measured on an uninformed cued-recall test. The results revealed striking effects, generating items that were significantly better retained than simply read items. The probability of successful recall was even better for items that had been processed in a generative fashion once than for those that had been read twice. Although Jacoby discussed these results along with the spacing effect (Hintzman, 1973; Hintzman *et al.*, 1974)—probably because the conditions involving a spaced presentation also showed a superior recall performance than an immediate presentation regardless of the presence of generation—they can be taken as evidence of the generation effect as well.

2.2.2 An Initial Argument—Slamecka and Graf (1978)

Concurrently with the paper mentioned above, Slamecka and Graf (1978) presented the first set of solid evidence of the generation effect. Their research consisted of five experiments. A cued-recall paradigm was used for the "generate" condition in each experiment. Experiment 1 tested the hypothesized retention advantage of self-generated items over presented items. At the University of Toronto, 24 students studied 100 word pairs, each presented separately on an index card. For the participants in the "generate" condition, each card consisted of a set of stimulus words and the initial letter of a response word (e.g., rapid–f). After drawing a card from the pile, participants generated a word according to an operative rule (e.g., synonym) that was cued orally by the experimenter (e.g., "fast" can be generated in this). Five rules (e.g., associate, category, opposite, synonym, and rhyme) were separately and fixedly used for each of the 20 consecutive cards. For the read-only participants, an intact word pair (e.g., rapid–fast) was presented per card, and the relationship of the words (e.g., synonym) was provided. In both conditions, participants had to verbalize both the cue word and the response word. Before the study phase, they were informed that the recognition of the responses would be assessed at the end. Immediately after the studying phase, the recognition of all target words was assessed in a three-choice format (i.e., a correct choice and two distractor choices per item) together with a five-point confidence rating for each one. The results showed that self-generated items were more accurately recognized than read-only items at a significant level regardless of the operating rules. Analyses of the confidence rating also revealed that the generation group chose their answers more confidently than the read-only group. This recognition superiority was further tested in Experiment 2 by comparing the recognition performances with and without the notification of a subsequent recognition test at the beginning of the experiment. The researchers found that the difference of effect sizes from the informed-uninformed manipulation was very small, with an advantage for generative processing as a whole. Thus, the researchers concluded that the generation effect was so robust that a manipulation of whether a test was announced or not was “totally inconsequential” (p. 595).

A further investigation was carried out, in Experiment 3, on the way in which recognition probabilities were compared between response words and stimulus words. The results indicated an overall superior recognition rate of words in the “generate” condition compared to the read-only condition regardless of the location of words. However, researchers also discovered that the effect was much larger for response words than stimulus words. A post-hoc comparison further confirmed that the recognition superiority in the “generate” condition for stimulus words was within the error range, suggesting a distinctive nature, in which only generated items were beneficial for subsequent recognition. Experiments 4 and 5 involved comparisons of learning curves associated with the generate-read distinction. The participants in Experiment 4 learned the same set of word pairs five times, either in the "generate" condition or in the read-only condition, and they were given a free-recall test at the end of each learning trial, whereas the participants in Experiment 5 worked on a cued-recall test at each end of the learning trials. Both experiments showed the same tendency—that the generation effect was significant at earlier trials, especially for the first three; then, the learning curves started to converge and were finally level at the fifth trial. Through these five experiments, the researchers concluded that they could confirm "the existence of the generation effect" (p. 602) because the expected phenomena were readily replicated under various tests such as free versus cued recall, cued versus uncued recognition, and confidence ratings.

2.2.3 Data- vs. Conceptually Driven Processing—Jacoby (1983)

Jacoby (1983) further examined the retention superiority of self-generated items (Slamecka & Graf, 1978) from a different angle. He defined receptive learning to mean that a learner simply reads a word as "data-driven" processing because, in this case, the item in memory is considered to constitute a sheer copy of the external stimulus. In contrast, generative learning, in which a learner generates a word from a conceptual clue entails "conceptually driven" processing because, in this case, the word is considered to be a product of conceptual processing within the learner. The researcher found that each type of processing facilitates a particular subsequent processing in a different way.

In this series of experiments, 80 students at McMaster University, Canada, were exposed to a set of 50 high-frequency English antonym pairs on a computer screen in four types of presentation: no context, context, generate, and new. In the "no context" condition, a series of x's appeared first for 1 second and was replaced with a word from an antonym pair that stayed for 1 second (e.g., xxxx → cold). In the context condition, a word from an antonym pair was presented for 1 second and then replaced with the other word from the pair, which remained for 1 second (e.g., hot → cold). In the "generate" condition, a word from an antonym pair was displayed for 1 second and then replaced with a series of question marks (e.g., hot → ????). Finally, in the new condition, which served as a control item, neither member of an antonym pair was presented. Participants received these types of presentations in an intermixed manner. In each trial, the task consisted of uttering the final word as quickly as possible. Encoded words were assessed via a perceptual identification task and a recognition memory task. In each trial of the perceptual identification task, a word presented in the encoding phase was displayed on the screen for 0.5 seconds and then immediately masked by a series of ampersands (e.g., cold → &&&&). Participant were to report the word immediately after its presentation. In contrast to the perception test, the recognition memory test was paper-based and contained new words that had not appeared so far in addition to the previously presented words. Moreover, the letters "R" and "G" were put next to each word, and participants were prompted to circle either the letter "R" if they thought they had read the word or the letter "G" if they thought they had generated the word. The results showed that words processed in a data-driven way (i.e., those had been just read;) were best identified during the subsequent perceptual identification task. In contrast, words that relied on conceptually driven processing (i.e., namely self-generated words;) outperformed those associated with data-driven processing; (i.e., words that have been read;) in later recognition memory. Consequently, Jacoby argued that a greater degree of data-driven processing promotes subsequent perceptual identification, whereas a greater degree of conceptually driven processing discourages the facilitation of perceptual identification but enhances later recognition memory. He further hypothesized that interactive processes in reading may involve these different sensitivities.

2.2.4 Effects of Directed Forgetting—MacLeod and Daniels (2000)

In line with Jacoby's (1983) findings on the deferent sensitivity of learned words between data-driven and conceptually driven processing, MacLeod and Daniels (2000) investigated the generation effect by employing directed forgetting manipulation. Directed forgetting, in experimental psychology, refers to a typical retention impairment in explicit memory when an instruction is given to forget an item immediately after its exposure, in comparison to a relatively better retention of items when participants are told to memorize the word (MacLeod, 1998). Therefore, this experiment involved examining whether items generated and others that were only read would be influenced by the "forget" and "remember" instructions. Eighty-five undergraduates at the University of Toronto generated and read 20 English words in each of the "forget" and the "remember" conditions. In the "generate" condition, a participant produced a word (e.g., "umbrella") with a definition given, such as "this protects you from the rain– u____?" In the "read" condition, participants simply read a presented word. In each condition, half of the presentations included an instruction to forget the word, indicated by a series of Fs (e.g., "FFFFF") at the end of a trial. The other half had an instruction to remember the word, specified by a series of Rs (e.g., "RRRRR"). The retention of words was assessed in two ways: a perception-oriented test and a recall test. The perception test involved a sped-up reading of words, in which participants were asked to read out a series of words as quickly and accurately as possible. In the recall test, participants were instructed to recall as many of the words as possible, disregarding the "forget" and "remember" instructions in the study phase. These manipulations revealed contrastive characteristics between "generate" and "read" items. For generated words, "remember" words were more rapidly read than "forget" words at a significant level, but there were no significant differences in recall. Conversely, for words that were simply read, both "remember" and "forget" words were read equally fast, but the former were significantly well recalled compared to the latter. That is, the remember-forget instructional manipulation facilitated the priming of generated words but not read words, whereas this manipulation affected the explicit recall of read words but not generated words. As a result, the researchers asserted that two encoding manipulations, "generate" and "read," each affect different processes that underlie performance on an indirect test and a direct test. Therefore, these tests are differentially sensitive.

2.2.5 Self-Generated vs. Provided by Others—Crutcher and Healy (1989)

Crutcher and Healy (1989) explored the retention of items provided by others, as opposed to self-generated items, by using simple multiplication problems. In their experiment, 24 students at the University of Colorado were given 20 multiplication problems with answers that were unique two-digit numbers (e.g., 6 x 8 = 48) in four task conditions: read, calculate, generate, and verify. Each of these conditions involved two different independent variables: cognitive operations (other vs. self) and stimulus presence (present vs. absent). In the "read" condition, which corresponded to the pattern of "other and present" in the variable combination, participants simply read the problem and answered aloud. In the "calculate" condition (i.e., other and absent), participants read the problem aloud and used a calculator to generate the answer and orally report it.

In the "generate" condition (i.e., self and absent), participants read the problem and generated an answer without using calculators and reported it. Finally, in the verify condition (i.e., self and present), participants read the problem and answer aloud and verified whether the answer was correct or not without using calculators, additionally reporting it by saying "correct" or "incorrect." Participants processed five multiplication problems in each condition; thus, the experiment was a within-participant measure. After completing 20 problems, participants were asked to recall as many of the answers they had reported in the study phase as possible. The results showed that recall probabilities for the "generate" and "verify" conditions were significantly higher than those for the "read" and "calculate" conditions. Further examinations of long-term retention conducted two days and seven days after also confirmed the same retention characteristics as immediate recall. Because the rates of retention were at the same level in the "generate" and "verify" conditions, the researchers concluded that the typical generation effect must be considered as a phenomenon inherently derived from internal mental operations rather than the internal generation of stimuli.

2.2.6 The Read-Recite-Review Strategy—McDaniel, Howard, and Einstein (2009)

McDaniel, Howard, and Einstein (2009) proposed a study strategy for learning from educational texts, incorporating the generation effect into a practical use in education. The strategy, termed the read-recite-review technique, or the 3R in short, requires students to read texts once, as much of the text as they can remember aloud, and then read the same text again. To examine this approach's effectiveness, they contrasted the 3R with two other commonly employed techniques: rereading and note-taking. In their experiments, 72 students at Washington University studied four short passages containing about 250 words taken from an EFL textbook in one of the three conditions to which they were randomly assigned. All participants were notified that they would be tested about the materials later. The reread group was told to read the passages twice without any other additional activities. The note-taking group read the passages twice as the reread group did, and took notes during reading to help their memory for the subsequent tests. They were not allowed to use their notes on the tests. The 3R group read the passages once, recalled the passages as well as they could, and then read the passages again. Their memory of the text was assessed in three ways: a free-recall test, a multiple-choice test, and short-answer inference questions in which answers could be determined by synthesizing two or more ideas from the passages. An immediate posttest asked about the first and third passages. A delayed posttest performed a week later concerned the second and fourth passages. Analyses of the immediate and delayed free-recall performances showed a significantly greater recall probability for the 3R approach than the other study strategies. Meanwhile, differences in the learning conditions were not found on the multiple-choice test or the short-answer questions. However, when more complex passages were employed, the 3R significantly benefited in terms of participants' performance on a multiple-choice test, both immediate and delayed, immediate performance on short-answer questions, as well as free-recall performance. According to the researchers, these results suggested that generative learning is beneficial beyond simply improving retention. In conclusion, they proposed that students could be encouraged to utilize the 3R strategy for remembering information from class materials.

2.3 Prior Knowledge

2.3.1 An Early Empirical Report—Morris, Gruneberg, Sykes, and Merrick (1981)

It seems plausible that learners who know a great deal about English-related matters (e.g., American culture) can acquire the English language more successfully than those without an interest in such matters. In the field of second-language acquisition, Robinson (2007) listed prior knowledge as one of the components influential in the successful completion of a learning task. In the skill acquisition literature, discussions of prior knowledge can be traced back to two studies indicating that a chess grandmaster could memorize the positions of chess pieces at a glance and reconfigure them more quickly and accurately than novice players could (Chase & Simon, 1973; de Groot, 1965). Those findings were later replicated with Go players (Reitman, 1976).

Morris, Grunberg, Sykes, and Merrick (1981) investigated whether individuals who knew a great deal about association football (i.e., soccer) could learn more game results than those who knew little about this sport. In their experiment, 38 male students at University College Swansea answered a questionnaire about soccer (e.g., "Who is Manchester City's goalkeeper?" and "What was the last English team to win the UEFA Cup?") and then listened to a broadcast of game results on the radio. The broadcast lasted four minutes, including game results for four English and three Scottish divisions, 64 matches in total. Soon after learning the results, participants were asked to write down as many of the scores as they could remember on a piece of paper with all the matches on it. A week before the experiment, another 24 male students, serving as the control group, also answered the same questionnaire and were then asked to predict the scores in the matches to be played on the day of the experiment. The control group was employed to deal with a potential argument that soccer enthusiasts' memory for game results would be due to their better predictions. In the end, the experimental group recalled an average of 14.9 game scores, while the control group managed to remember 4.3 on average. Moreover, a significant positive correlation was found between soccer knowledge and the recall of game results for the experimental group but not for the control group. These results supported the researchers' assumption that existing knowledge would facilitate the acquisition of relevant new knowledge in a domain and that the performance was not related to better prediction ability associated with expert knowledge. The researchers argued that soccer fans could learn scores easily probably because their richer knowledge made it possible to draw immediate implications from the results such as the league positions of their clubs and others, which then could be understood as a result of these fuller cognitive activities. To a person uninterested in soccer, no such processes would occur.

2.3.2 A Theoretical Model—van Dijk and Kintsch (1983)

The superior domain-specific learning performance of individuals with a high level of knowledge in a particular domain has often been explained within a theoretical framework originally developed for text comprehension called "situation models" (van Dijk & Kintsch, 1983). A situation model constitutes a mental representation of the concepts a text describes. Moreover, such models can be formed by many dimensions, such as the integration of prior knowledge and the information provided in the text.

When individuals read a text, they create a "textbase," a propositional representation including the concepts represented in the text. Reader, then, supplement the textbase with relevant prior knowledge to create a personally comprehensible interpretation of the text. In this way, a situation model is constructed from both the textbase and prior knowledge associated with the textbase. Once a situation model is created, it can be used to aid the comprehension of other relevant texts. In addition, if the created situation model is found to relate to other existing models in long-term memory, the new model can be used to update the existing model. In such cases, learning occurs. Scholars have hypothesized that the integration of situation models becomes possible when at least one of five aspects in the current and the stored situation models match: space, causation, intentionality, protagonists and objects, and time (Zwaan & Radvansky, 1998). First, the spatial relations between an observer and the objects in the environment constitute crucial information. Among spatial relations, an above–below axis is particularly salient due to the constant presence of gravity. Furthermore, a front–back axis is prominent since it refers to the typical direction of movement. Second, causal relationships, often indicated by causal connectives such as “because” or “therefore,” are essential for text comprehension. Readers can infer these causal relations without those connectives by using their prior knowledge. For example, the sentence “Cathy poured water on the bonfire.” implies that the fire went out based on people’s general knowledge that water extinguishes fire. Third, intentionality can also be identified as a perceptual aspect because human behavior is generally goal directed. Therefore, it is assumed that readers can use general cognitive procedures to analyze human behavior and comprehend narratives. The fourth dimension of situation models is protagonists and objects. Many studies have suggested that readers comprehend narratives while keeping track of the persons or objects mentioned in the narrative (for details, see Zwaan & Radvansky, 1998). Finally, readers should know the relative temporal relations among the events of the story to achieve a proper understanding of the text, thus making time another dimension of situation models.

In this way, we can apply, to a certain extent, situation models into explanations for the superior learning performance of high-knowledge learners in a certain domain. Situation models imply that a greater amount or quality of knowledge leads such learners to have more links to existing representations and thus more opportunities to update their models, and learning can occur with greater ease. These models also predict, however, that those who have expert knowledge in a certain domain will show lackluster learning performances that can be comparable to low-knowledge learners when faced with non-domain-relevant information (van Overschelde & Healy, 2001).

2.3.3 Facilitative Effects on Learning Through Everyday Knowledge—Kole and Healy (2007)

Scholars have typically carried out earlier studies on the facilitative effects of prior knowledge in learning by contrasting the performance of experts or those with a high levels of knowledge of a particular domain, such as grandmasters of chess (Chase & Simon, 1973; de Groot, 1965), Go (Reitman, 1976), soccer enthusiasts (Morris *et al.*, 1981) or baseball fans (van Overschelde & Healy, 2001,) with that of novices or those with little or no knowledge of a particular domain. In contrast, Kole and Healy (2007) investigated whether learners could use their everyday knowledge as mediators in learning unfamiliar facts. Following the situation model, the researchers assumed that providing cues familiar to learners would facilitate the learning of non-domain-relevant information, given that experts' superior learning performances occur due to the number of links available between new information and existing knowledge.

In their experiment (Kole & Healy, 2007), 36 students at the University of Colorado learned a set of 144 fabricated facts. Each fact was presented as a short sentence including the person, a verb phrase, and an object noun (e.g., "Brian Burke likes to play lacrosse"). The name of the person in each sentence came from a list of professional beach volleyball players with whom the participants were unfamiliar. The control variable was a degree of prior knowledge that was manipulated between the participants at three levels. In the low-knowledge condition, participants learned fabricated facts about unfamiliar people. In the high-knowledge condition, the unfamiliar name in each sentence was replaced with one the participant was familiar with, such as names of friends or relatives, which had been elicited beforehand. Finally, in the mediated condition, participants first worked on paired-associate learning between an unfamiliar name in each sentence and a familiar name, and they were also exposed to fabricated facts. Furthermore, in this condition, participants received an extra instruction that they were to recall the associated familiar name while the researcher presented each fact. For all conditions, facts were presented for 3 seconds each, and a set of 144 facts were presented three times. After the learning phase, cued recall for all 144 facts was administered with each cue consisting of a name and a verb phrase that asked participants for the associated object noun (e.g., "Brian Burke likes to play _____."). The results for the high- and low-knowledge groups showed a typical prior knowledge effect, with significantly greater recall probability among the high-knowledge group than the low-knowledge group. However, the researchers also found that the mediated group significantly outperformed the low-knowledge group, even though both groups had studied unfamiliar facts. This finding implied that the mediated participants might have used the unfamiliar names as cues, which their existing representations made activated. The authors argued that these participants were able to integrate the new facts into their representations in long-term memory, and thus elaborate encoding of the new items became possible. The researchers further assessed that this finding provided support for the situation model's account of the prior knowledge effect because the only difference between the low and mediated group was the number of links available between new information and preexisting representations. Regarding practical implications for education, the researchers proposed that a mediation technique might be useful as a simple and quick way to improve learning, especially for those with little knowledge of a domain. This study implied that learners do not necessarily have to be experts in a domain to benefit from their prior knowledge.

2.4 Spacing

2.4.1 Terminology

Beneficial effects in learning from spaced or distributed practice have been extensively researched for more than 130 years (Ebbinghaus, 1885/1913; Thorndike, 1912). A quantitative meta-analysis conducted by Cepeda, Pashler, Vul, Wixted, and Rohrer (2006) identified 317 experiments in 184 articles in this research domain suggesting overall positive effects of spacing in learning with reasonably robust effects as well.

The term “spacing effect” refers to the practice of enhancing learning by leaving an interval between two study episodes for the same material. The interval that separates different study episodes is called an “interstudy interval,” which must be easily measurable, such as a unit of seconds (Cepeda *et al.*, 2006). Another interval called a “retention interval” is set between the end of the last study episode and subsequent tests. In a typical experiment design, the interstudy interval can vary, but the retention interval remains fixed. If the presentation of a given set of materials occurs multiple times with intervals in either an increasing or decreasing manner, researchers should consider a “lag effect,” which compares the effects of different levels of spacing. In learning, when the same material appears twice with an interstudy interval, it is spaced. However, when different material is inserted between the two presentations (e.g., material A appears, material A disappears, material B appears, material B disappears, and then material A appears again), researchers refer to a “distributed” practice. A typical counterpart condition to spaced practice is “massed” practice. When the same material is presented without any interventions or rest intervals, learning is said to be massed. When comparing the effects between spaced and massed practice sessions, the summed presentation time of a given item should be equal in the two conditions. The superior retention performance of spaced learning that many scholars have found results from comparisons with massed learning.

2.4.2 Optimal Interstudy Gap for Long-term Retention—Cepeda *et al.* (2008, 2009); Bahrick *et al.* (1993)

Research on the spacing effect has traditionally focused on comparing subsequent recall performance between spaced and massed practice schedules with a relatively short retention interval. For example, Bloom and Shuell (1981) compared American high school students' learning of 20 sets of French–English word pairs between a massed condition (i.e., giving all of them in a 30-min session) and a distributed condition (i.e., giving each 1/3 of them in a 10-min session for 3 consecutive days,) and discovered the that recall performance measured four days after the end of learning was significantly greater for the distributed learning group than massed learning one. However, another important research question in this field concerns what an optimal interstudy interval for long-term retention would be; the main interest here consists of comparing lag effects between varying gaps within the spacing framework (McDaniel, 2012).

In a study conducted by Cepeda, Coburn, Rohrer, Wixted, Mozer, and Pashler (2009), 161 students at the University of California learned a list of 23 unfamiliar facts twice with six interval conditions (i.e., 20 min, 1 day, 1 week, 1 month, 3 months, and 6 months) as between-participants variables. Each fact was presented as a question and an answer (e.g., "Who invented snow golf?–Rudyard Kipling"). In the first learning session, all 23 items were presented to participants, followed by three blocks of a test-with-feedback trial. In each of these trials, the participant first saw a question and then recalled the answer and wrote it down on an answer sheet. Immediately afterward, a correct answer was presented for 5 seconds. The second learning session, which revolved around relearning the material learned in the first learning session, involved two blocks of test-with-feedback trials of the same material, and it was carried out after the designated time passed for each experimental group (i.e., either 20 min, 1 day, 1 week, 1 month, 3 months, or 6 months after the first learning session). Following a six-month retention interval, the same items were tested for all participants without feedback on a recall test. Repeated measures of the analysis of the recall probability between gaps (i.e., from 20 minutes to 6 months) and times (i.e., second learning session and final recall) revealed that the 20-min gap (i.e., 0 day) produced significantly worse recall rates than all gaps longer than 1 day at the time of the final recall. Moreover, the 1-day gap produced significantly worse recall rates than the 1-month gap on the final recall. However, gaps longer than 1 month (i.e., 3 months and 6 months) also produced worse recall rates than the 1-month gap, although they were not statistically significant. Moreover, pairwise comparisons for each adjacent gap revealed a large benefit of increasing the gap from 20 minutes (i.e., 0 day) to 1 month. These results suggested that the 1-month gap was optimal for 6 months of retention.

Cepeda, Vul, Rohrer, Wixted, and Pashler (2008) further investigated optimal study gaps for retention intervals of up to 1 year (i.e., 7 days, 35 days, 70 days, and 350 days) on fact learning using 1,354 participants contacted online. Based on complete recall on the final test, the researchers identified the optimal gaps of 1, 11, 21, and 21 days for the retention intervals of 7, 35, 70, and 350 days, respectively. These findings by Cepeda *et al.* suggested that a month-long interstudy interval would enhance long-term retention most, from 6 months to 1 year.

Spacing effects with further long retention intervals were investigated by Barhrick, Barhrick, Barhrick, and Barhrick (1993). They conducted a 9-year longitudinal investigation with four subjects who learned and relearned 300 English–French or English–German language pairs at intervals of 14, 28, or 56 days. The retention rates were tested for 1, 2, 3, or 5 years after the completion of the learning process. According to the findings, increasing the interstudy gap to 56 days improved retention performance for all retention conditions (i.e., 1–5 years). However, some researchers speculated that an interval of 56 days was actually too short for the extremely long retention intervals. (Cepeda *et al.*, 2008).

2.4.3 Expanding and Equally Spaced Practice—Karpicke and Roediger (2007)

Expanding retrieval practice refers to a type of repeated practice with increasing interstudy intervals, as opposed to equally spaced practice. Scholars have frequently suggested that expanding retrieval practice promotes long-term retention more than equally spaced practice, especially for children, learners with lesser ability, and memory-impaired people, because an expanding pattern of retrieval encourages errorless learning (Cull, Shaughnessy, & Zechmeister, 1996). Researchers have investigated the effects of expanding retrieval relatively less than they have those of uniform spacing. In this regard, Karpicke and Roediger (2007) compared the effects of expanding retrieval practice with equally spaced practice for short- and long-term retention.

In their experiment (Karpicke & Roediger, 2007), 48 undergraduates at Washington University learned 36 infrequent English words selected from test preparation books for the Graduate Record Exam, divided into three practice conditions: massed, expanding, and equally spaced. Each item to be learned consisted of a pair constituting of a target word and its meaning (e.g., sobriquet–nickname, benison–blessing), which were tested three times after the initial exposure with varying spaces. In the massed condition, participants studied a word pair (e.g., sobriquet–nickname) and took three consecutive tests (e.g., sobriquet–____) without any interventions between tests. In the expanding condition, participants studied a target word pair and saw a filler word pair, and then the first test was presented. Between the first and second tests, five filler trials were inserted. Between the second and third tests, nine filler trials were inserted. In the equally spaced condition, five trials occurred between the initial study and subsequent tests. Therefore, the massed, expanding condition, and equally spaced conditions were denoted as 0-0-0, 1-5-9, and 5-5-5, respectively, in terms of the number of occurrence of intervening trials. After the training phase, half of the participants in each condition had a distractor task involving playing a video game for 10 minutes, and then they took the final recall test to assess the short-term retention of the material. The other participants were dismissed from the experiment of that day and took the final test 2 days after to assess their long-term retention. The results for short-term retention (i.e., 10 min) indicated that expanding retrieval practice yielded the best recall performance, followed by equally spaced retrieval and massed retrieval practice. However, for 2-day long-term retention, equally spaced practice showed a better recall probability than expanding retrieval practice, contradicting an earlier theoretical view that expanding retrieval practice promotes long-term retention. In their discussion, the researchers pointed out the fact that the recall accuracy of the initial retrieval attempt, measured on the first test trial after the initial exposure, was worse in equally spaced retrieval than in expanding retrieval, suggesting that the first retrieval in the equally spaced condition was more difficult than that in the expanding retrieval condition. Therefore, they contended that making an initial retrieval attempt difficult would lead to many errors on the test but eventually yield large gains in long-term retention. Based on this hypothesis, the researchers concluded that expanding intervals in repeated practice are not important. Rather, delaying an initial retrieval attempt constitutes a key factor to promote long-term retention.

2.4.4 Distributed Practice for Syntax Learning—Bird (2010)

While most research about the spacing effect with verbal materials has concerned with the retention of new words, Bird (2010) looked at the acquisition of English-language syntax using a spacing paradigm. In this study, 38 university students whose native language was Malay learned the simple past, present perfect, and past perfect tenses in a semester-long English-language course. In each study session, students engaged in a grammaticality judgment task for the target structures (e.g., "I saw that movie last week." or "*I have seen that movie last week") for 30 min; then, correct answers were presented using the overhead projector with brief explanations. Half of the students, Group 1, were taught distinctions between the simple past and present perfect with a 3-day interstudy interval and distinctions between the present perfect and past perfect with a 14-day interstudy interval. The other half, Group 2, received the reverse schedule: They were taught simple past–present perfect distinctions with a 14-day interstudy interval and present perfect–past perfect distinctions with a 3-day interstudy interval. Each study session occurred five times in a 14-week semester. Finally, students' grammatical knowledge was assessed on a surprise grammaticality judgment test on two occasions (i.e., 7 days and 60 days after the final study session). Group 1 received a posttest on simple past–present perfect distinctions after a 7-day retention interval and a posttest on present perfect–past perfect distinctions after a 60-day retention interval. Group 2 took the same tests in the reverse order. The results showed that, on the 7-day delayed test, scores associated with the 3-day interstudy interval were slightly better than those associated with the 14-day interstudy interval, but no significant differences emerged between them for all target structures. However, on the 60-day delayed test, scores associated with the 3-day interstudy interval were significantly lower than those associated with the 14-day interstudy interval. These results indicated that grammatical structures repeatedly taught with a shorter interval are more beneficial for short-term retention than those taught with a longer interval, but they are eventually forgotten in a relatively short term. In contrast, grammatical structures repeatedly thought with a long interstudy interval would not be fully useful in the short term, but they survive for a relatively long time at the same level. With these findings, the researcher suggested that the repeated practice of earlier learned material with a certain interstudy interval would enhance learning benefits and that these controls might be useful for language syllabus design.

2.4.5 Spacing Effect in Inductive Learning—Kornell and Bjork (2008)

In a common sense, learning can occur constantly outside the classroom. In many everyday situations, inductive learning, which involves the abstraction of principles, rules, and concepts from related examples, would happen more than verbal learning. Kornell and Bjork (2008) examined the effects of spaced practice involving inductive learning. They assumed that, unlike verbal learning, massed practice would enhance inductive learning more than spaced learning because massing makes it possible for learners to notice similarities of patterns in successive presentations, whereas spacing makes this process more difficult.

In this study, 120 students at the University of California viewed 72 paintings of scenery by 12 unfamiliar artists. Half of the participants were assigned to the massed presentation condition, and the rest were assigned to the spaced presentation condition. In the massed condition, the paintings were presented in such a way that six pictures by the same artist were presented successively, followed by six pictures by another artist. In the spaced condition, the paintings by the 12 artists were intermingled. Each painting was shown for 3 seconds on a computer screen with the name of the painter displayed under the picture. After the study phase, participants were shown 48 new paintings by the 12 artists and then asked to identify the name of the painter for each picture. According to the results, in contrast to the researchers' initial assumption, spacing performed better than massing. Participants in the spaced presentation identified the authors significantly more accurately than those who learned in the massed presentation.

The researchers further conducted a similar experiment by recruiting 72 new participants by setting the presentation condition (i.e., massing vs. spacing) as a within-participant variable. Participants viewed paintings by six artists in the massed condition and paintings by the other six artists in the spaced condition. Moreover, at this time, the researchers asked participants their intuition regarding which presentation pattern was more helpful for learning. This experiment demonstrated, again, superior subsequent recognition rates with spaced practice; namely, 85% of the participants did well in the spaced condition compared to in the massed condition. However, 83% of the participants answered that the massed condition was more effective than the spaced condition, despite the fact that massed practice was associated with inferior recognition performance. Based on these results, the researchers asserted that people rely on the illusory benefits of massed practice due to the intuitive ease in learning, but such belief may be misleading from an educational standpoint.

2.5 Overlearning

2.5.1 Two Notions of the Term

Two senses of the term "overlearning" exist. One of them refers to extra practice after reaching a set criterion at the initial stages of learning. A typical experimental manipulation in this sense of overlearning requires a learner to continue practicing beyond arriving at an errorless trial of the learning set, while a control learner stops the learning session after a perfect trial. Research has suggested that this overlearning promotes long-term retention of acquired knowledge and skills (Driskell *et al.*, 1992).

The other notion of overlearning pertains to a delayed practice of acquired knowledge and skills. This type of overlearning should be carried out sometime separately after the initial learning session. Therefore, it may include refresher practice on rusty knowledge by taking advantage of the benefits of the spacing effect in the long run. This notion of overlearning is assumed to be highly responsible for acquired skills being consolidated and automated (Dougherty & Johnston, 1996; Kim *et al.*, 2013).

2.5.2 Overlearning's Effects on Long-term Retention—Driskell *et al.* (1992); Rohrer *et al.* (2005)

In their meta-analysis paper, Driskell, Willis, and Copper (1992) examined 16 overlearning studies conducted from the 1920s to the 1980s. Eleven of those studies used cognitive tasks that mainly involved learning words, and five were based on physical tasks, such as maze tracing or assembling a military gun. Typically, participants in these studies engaged in training tasks a designated number of times, and several additional trials were conducted only for target participants. Retention intervals varied from 0 to 56 days, but most studies (nine of them) employed a shorter retention period of 2 weeks or less. According to the analyses of effect sizes conducted by Driskell *et al.*, the effects of overlearning were moderate in size for both cognitive tasks (d = .753) and physical tasks (d = .443), suggesting that overlearning can enhance subsequent retention. However, when comparing the magnitudes of the effects among the different retention intervals, the researchers discovered that the initial beneficial effects decreased by half 19 days after learning and almost disappeared 38 days after. Therefore, they mentioned that refresher training might be advisable at approximately 3-week intervals.

More recently, Rohrer, Taylor, Pashler, Wixed, and Cepeda (2005) replicated and extended the findings of Driskell *et al.* (1992) with longer retention intervals of up to 9 weeks. In their study, 130 undergraduates at the University of South Florida studied 10 city–country pairs (e.g., Axim–Ghana), with either 5 or 20 test trials. Knowledge retention was tested at 1, 3, and 9 weeks after learning. Therefore, the study involved a 2 (learning level) × 3 (retention interval) between-participants design. All participants first saw a list of 10 city–country pairs. Immediately afterward, they engaged in a test-with-feedback task for their memory with city names as a cue. Participants in the low learning group received the test-with-feedback procedure 5 times, whereas the high learning group received it 20 times. Here, the researchers distinguished "true low learners" as participants in the low learning group who produced no more than one perfect trial (N = 44,) and "true high learners" as participants in the high learning group who produced at least three perfect trials (N = 58). These participants were then randomly assessed for their memory after 1, 3, or 9 weeks after learning. The results were consistent with their previous findings indicating that the effects of overlearning lasted up to 3 weeks after learning but disappeared 9 weeks afterward.

Subsequently, Rohrer *et al.* extended the experiment above by manipulating the total number of items, rather than the number of trials. In this experiment, 88 students were assigned to the low or high learning conditions. The low and high learning groups learned 20 and 10 city–country pairs, respectively. The learning time was fixed for both groups. Therefore, the low learning group could spend virtually half the time per pair compared to the high learning group. The retention intervals were set at 1 and 4 weeks. The results showed, again, that the rate of recall of the high learning group, which was significantly better than the low learning group in week 1, sharply declined over time and converged to the low level of the low learning group in week 4. In conclusion, the researchers suggested that overlearning is efficient as previous studies had indicated, especially when comes to remembering information for a meaningfully long-term period. However, because the short-term effect of overlearning was found to be reliably strong, it might be appropriate for learners aiming at a short-term goal, such as students preparing for exams.

2.5.3 Overlearning in the Course of Mastery

Naturally, the quality levels of a certain skill vary at different stages of acquisition. In the early stages, performance is typically slow and full of errors, but speed increases, and the number of errors decreases through repeated analogous practice, and the learner finally gains the ability to perform well. According to an early study of skill acquisition in ergonomics (Crossman, 1959), a worker in a cigarette factory who needed 25 seconds to roll a cigarette immediately after beginning the job required only 9 seconds a year later. Seven years afterward, the same worker produced cigarettes at a rate of 7 seconds per roll, where a cigar-making machine at that time produced a roll every 6.5 seconds. Fitts (1964) postulated that the course of gaining expertise, especially motor skills, could be divided into three stages: the early, intermediate, and late phases. In the earliest phase of skill acquisition, learners develop a declarative encoding of the skill, very similar to the processes involved in rote learning. In the intermediate phase, learning predominantly shifts to instance-response association. Over long periods of practice, performance improves gradually and finally moves to the late phase characterized by stable, fast responses. This three-stage theory was further developed by Fitts and Posner (1967). The first stage, called the “cognitive phase,” involves understanding the task, and it is largely necessary to rely on cues or responses to perform the task. Instructions work well. Typically, learners in this phase rehearse a set of facts relevant to the task. In the second stage, called the “associative phase,” erroneous responses from the individual’s initial understanding are gradually altered. Furthermore, the connections among the various elements required for the successful completion of the task are reinforced, and the performance becomes gradually faster. The third stage is called the “autonomous phase.” In this final stage, the procedure becomes increasingly autonomous, and the individual can carry out the component processes adequately with little cognitive control. Referring to Fitts’ framework, Anderson (1982) also proposed three learning stages in his adaptive control of thought (ACT) theory. Initial knowledge, kept in a declarative form, transfers into a procedural form through transition stages that bridge these two forms of knowledge. The ACT framework, which mainly deals with cognitive skills, is extremely influential in cognitive science and frequently cited in the second-language acquisition literature, especially in relation to the formation of language automaticity (e.g., DeKeyser, 2001). A common claim among these theoretical frameworks is that skills develop through the power law of learning (Anderson, 2009; DeKeyser, 2001), suggesting that only by practicing can learners move to the next acquisition stage. Therefore, overlearning, in terms of the occasional repeated practice of existing knowledge, as opposed to massed practice within a single session for long-term retention, directly connects to any form of skill acquisition.

Recently, Kim, Ritter, and Koubek (2013) proposed optimal overlearning schedules focusing on the acquisition stage helped or hindered by this activity. They asserted that overlearning clearly helps acquisition during the first stage of skill learning because initial knowledge depends largely on declarative memory. However, declarative knowledge is vulnerable to decay. That is, initial knowledge always remains at risk of being forgotten. Therefore, the occasional activation of such declarative components is necessary. The researchers also argued that, among declarative components, only ones that can be readily retrieved would be transferred to a procedural form. Therefore, overlearning plays an important role in the task skill as the learner enters the second stage of acquisition, in which skill components reside in both declarative and procedural forms. In the second stage, overlearning is advisable as well. One of the main characteristics of this stage is marked by the transition of the form of knowledge from declarative to procedural. From a theoretical viewpoint, only after the skill knowledge components are fully proceduralized is a skill recognized as being at the final stage. Therefore, it can be hypothesized that overlearning is beneficial in encouraging knowledge components to exceed a transition threshold. In contrast, the researchers predicted that overlearning might not play an effective role at the third stage of acquisition because the task knowledge at this stage is kept in a procedural form characterized by its immunity to decay. However, overlearning could also play a certain role even in this final stage, as the scholars did not explore the forgetting of procedural knowledge in depth.

2.6 Task Difficulty

2.6.1 Introduction of Difficulty in Learning Tasks—Schmidt and Bjork (1992)

In their paper entitled “New Conceptualizations of Practice,” Schmidt and Bjork

(1992) mentioned that typical training procedures considered effective are, in fact, far from optimal, provided that the long-term retention and transferability of acquired knowledge and skills are emphasized in learning. They argued that correct responses during learning may not necessarily constitute reliable indicators of successful acquisition. They listed a number of earlier cases that possibly entailed the effects of task difficulty, such that random practice yielded better retention than blocked practice (Shea & Morgan, 1979), minimum feedback facilitated subsequent retention performance (Schooler & Anderson, 1990), and variability in practice, as opposed to uniform practice, reduced erroneous responses on subsequent new tasks (Mannes & Kintsch, 1987). A remarkable feature in these examples is that these studies questioned the commonly held belief that errorless trials during training increase acquisition, suggesting instead that, conditions that depress performance during training could maximize acquisition in the long run. In this study, Schmidt and Bjork implied that difficulty is a key component of a training task.

2.6.2 Desirable Difficulty—Bjork (1994)

In response to Schmidt and Bjork's (1992) argument about the need for difficulty in training tasks, Bjork (1994a; 1994b; 1999) argued that trainers should introduce desirable difficulties for the learner. Bjork's term "desirable difficulty" refers to a degree of difficulty that instructors should implement in learning tasks independent from the learner's perceptions about what constitutes a "good" task. He suggested that easy and errorless learning would likely boost acquisition because learners have a general idea that the current ease of identifying a correct answer predicts future success and that errors made during learning predict future failures. Learners typically do not view errors as opportunities for learning, although empirical evidence has suggested that learning with errors supports subsequent performance.

Researchers have proposed a number of distinctions regarding the human memory system, such as declarative versus procedural knowledge, explicit versus implicit memories, and short- versus long-term memory. In all these distinctions, an advantage of one can be a disadvantage of the other at the same time. However, an important point made here is that a latent disadvantage at a certain time can become an advantage at another time.

> Performance during training is an unreliable indicator of the extent to which the learning that is the goal of training has been achieved. Conditions that yield a high rate of correct responses during training can fail to support performance in the posttraining environment; conversely, conditions that appear to slow or impede performance during training can enhance the subsequent real-world performance that is the target of the training.
>
> (Bjork, 1994b, p. 299)

In relation to possible ways of introducing desirable difficulties into practice, Bjork (1994a) listed several practical manipulations, such as (a) varying the conditions of practice, (b) providing contextual interference, (c) distributing the practice of a given task, (d) reducing the levels of feedback to the learner, and (e) using tests as learning events. The desirable difficulty framework holds that learners' post-practice retention and transfer become maximized when an optimal level of difficulty is provided during training.

2.6.3 Contextual Interference Effects on Task Difficulty

Contextual interference is a form of task difficulty generated by controlling an item's presentation in a task. This paradigm originates from Battig (1972, p. 156) who proposed "intratask interference as a source of subsequent facilitation." An advanced point of this paradigm, compared with other skill acquisition theories, is that it mentions not only retention but also transfer performance on subsequent tasks.

Battig (1972) investigated the effects of contextual interference in terms of retention and transfer. In his experiment on retention, five groups, each consisting of 20 participants, learned 12 pairs of nonsense CVCs (e.g., KOW-CEL, BIK-LEX) with five different presentation orders. The only difference between each condition was the number of items blocked during the learning session, symbolically indicated as follows, where "S" stands for a stimulus term alone (e.g., KOW) and "P" denotes a stimulus–response pair (e.g., KOW-CEL).

Group 12: S1-S2-S3-S4-S5-S6-S7-S8-S9-S10-S11-S12-P1-P2-P3-P4-P5-P6-P7-P8-P9-P10-P11-P12

Group 6: S1-S2-S3-S4-S5-S6-P1-P2-P3-P4-P5-P6-S7-S8-S9-S10-S11-S12-P7-P8-P9-P10-P11-P12

Group 4: S1-S2-S3-S4-P1-P2-P3-P4-S5-S6-S7-S8-P5-P6-P7-P8-S9-S10-S11-S12-P9-P10-P11-P12

Group 2: S1-S2-P1-P2-S3-S4-P3-P4-S5-S6-P5-P6-S7-S8-P7-P8-S9-S10-P9-P10-S11-S12-P11-P12

Group 1: S1-P1-S2-P2-S3-P3-S4-P4-S5-P5-S6-P6-S7-P7-S8-P8-S9-P9-S10-P10-S11-P11-S12-P12

In the learning phase, all groups saw the 12 CVC pairs in the designated order and recalled response words with stimulus words serving as cues. This procedure continued until each participant made 60 correct responses in total. Immediately after this criterion trial, participants were given a 12 x 2 matrix with the two columns labeled "left-hand members" and "right-hand members," and told to write down as many pairs as they could remember. The five groups differed significantly in the number of errors they made while reaching the specified 60 correct responses. Group 12 (i.e., to whom all the 12 stimulus words were given, followed by all the 12 stimulus–response pairs) exhibited the most efficient learning (90.5 errors), followed by group 6 (102.2 errors), group 4 (111.0 errors), group 1 (123.0 errors), and group 2 (139.5 errors). However, an analysis of participants' retention rates revealed the opposite pattern. Group 1 recalled CVC pairs best (M = 7.85). Group 2 remembered equally well (M = 7.25). Group 12, however, had the worst retention rate (M = 5.80). Battig explained that the results indicated that acquisition difficulty, operationalized by block size, enhanced their subsequent retention.

The effects of contextual interference on subsequent transfer were also examined in paired-associate learning of nonsense CVCs. Two experimental groups participated in this experiment. One was to learn 8 CVC pairs including 4 CVCs used twice as stimulus and response terms for different pairs (e.g., SOC-TID, RAB-SOC). The other group had to learn 8 CVC pairs with no duplications (e.g., SIF-TOG, RAB-SOC). The two groups repeatedly studied their lists until reaching 40 total correct responses. After the criterion trial, both groups received a list of new CVC pairs and were told to learn them. An analysis of their relearning performance revealed that the group that learned the duplicate list in the initial learning phase made more errors in learning transfer pairs than the group that learned a non-duplicate list did. Battig (1972) reasoned that, because four CVCs were used twice, the duplicate list could readily be subdivided into four sets of two pairs, suggesting that the duplicated list was easier to learn than the other list.

The contextual interference effect has been replicated for both motor and cognitive skills by using random-versus-blocked practice schedules (Carlson & Yaure, 1990; Schneider *et al.*, 1995; Shea & Morgan, 1979). For instance, Shea and Morgan

(1979) found that a random presentation sequence of three motor tasks—knocking down three or five designated panels with a tennis ball in a designated order quickly by a cue of a stimulus light of a certain color—facilitated subsequent performances significantly better than a blocked presentation, as measured via posttests 10 minutes and 10 days after training. Moreover, performance on a transfer task that was made more complicated by adding two more steps from the initial acquisition procedure was significantly greater for the random training group than the blocked training group.

Carlson and Yaure (1990), who applied Shea and Morgan's (1979) findings to the acquisition of a cognitive skill, found that skills acquired under a random training schedule were superior in terms of subsequent transfer than a blocked learning schedule. In their study, participants learned four Boolean functions (i.e., AND[a, b], OR[a, b], NAND[a, b], NOR[a, b]) in three conditions: blocked per function, randomized, and mixed (i.e., blocked for two functions and randomized for the other two). The rate of learning in each condition was measured in terms of the mean reaction time for correct responses in the rule learning process. The transfer of knowledge was assessed on a separate problem-solving task that involved equation-changing calculations (e.g., Given that A = AND[1, 0]; B = OR[0, 1]; then X = NOR[A, B]?). The results showed that blocked practice conditions (i.e., low interference involved) boosted the rate of learning more than the random practice condition (i.e., high interference involved). However, in the transfer task, participants in the random practice condition produced correct responses significantly faster than those in the blocked practice condition.

In a context similar to those used by Carlson and Yaure (involving Boolean functions;), Schneider, Healy, Ericsson, and Bourne (1995) found that random practice schedules slowed down the rate of learning but yielded superior retention performance on an immediate posttest in comparison to blocked practice schedules. Furthermore, a remarkable contrast with long-term retention was found. Skills learned in blocked schedules decayed four weeks after the initial learning, whereas skills learned in random schedules remained relatively intact after a month.

2.6.4 The Training Difficulty Principle

While research has suggested that increasing interference during learning impedes speed of learning but enhances the durability and flexibility of the material learned, Schneider, Healy, and Bourne (2002) argued that any factor increasing difficulty during learning promotes subsequent retention and transfer. The findings of their two studies support this claim as they carried out a replication of the contextual interference effect (Schneider *et al.*, 1998) and its extension (Schneider *et al.*, 2002) on task difficulty in the field of foreign-language vocabulary acquisition.

Schneider, Healy, and Bourne (1998) replicated the contextual interference effect on paired-associate vocabulary learning tasks by presenting word pairs grouped by a conceptual category or arranged in a mixed order. In this study, 24 university students with little knowledge of French learned French–English word pairs delivered on a computer screen in either the grouped or the mixed condition. There were 25 French–English word pairs to learn, and they were presented to participants in blocks of five, each at a rate of 2 seconds. In each block, five intact French–English word pairs were first shown for studying. Then, participants saw the five French words one by one as cues and subsequently typed a counterpart English word for each one. For participants in the blocked condition, the five words in a block were conceptually related (e.g., *voiture*–car, *velo*–bicycle, *avion*–airplane, *camion*–truck, *bateau*–boat). For participants in the mixed condition, the five words in a block consisted of a word from each of five categories so that they were conceptually unrelated (e.g., *dos*–back, *avion*–airplane, *assiette*–plate, *jambon*–ham, *chemise*–shirt). In this way, participants were presented a set of French-English word pairs, asked to type English equivalents, and continued this cycle until all 25 word pairs were completed. Following this initial learning phase, all participants were asked to respond to all the French cue words until they completed two perfect trials consecutively. This step ensured that all participants had an equal level of knowledge of the target word pairs. Immediately after participants completed the criterion trial, the retention of the studied English words was tested by having participants type them on a computer based on the French cue words. A week later, participants returned, and the retention of the English words was tested again. They were then asked to study target word pairs again to assess the ease of relearning. At this juncture, half of the participants in each condition were placed in the other condition (i.e., blocked vs. mixed learning conditions) to assess their transfer performance.

The results of this study can be summarized in three points. First, participants in the blocked condition made fewer errors than those who in the mixed condition during the learning sessions and on an immediate posttest, but this pattern reversed in the relearning trials a week afterward, thus, a significant interaction between sessions (i.e., learning vs. relearning) was found. Second, participants' response times also followed this pattern. Participants in the blocked condition responded faster than those in the mixed condition during learning but the opposite was observed during relearning. Third, the researchers failed to obtain evidence for superior transfer effects, although they did not explicitly mention this finding in their paper. Rather, they interpreted the fact that participants performed when the retest matched the type of learning they had worked on in the initial phase as support for procedural reinstatement (Healy *et al.*, 1992), an idea analogous to Bransford's notion of transfer-appropriate processing. Consequently, only superior retention from contextual interference was emphasized in this research.

Schneider, Healy, and Bourne (2002) extended the study described above by employing translation direction and pretraining as independent variables in addition to their earlier manipulation (i.e., blocked vs. mixed). Following the revised hierarchical bilingual mental lexicon model (Kroll & Stewart, 1994), the researchers assumed that foreign-language word learning involving translation from a first language (L1) to a second language (L2) is more difficult than that involved the reverse direction and that pretrained items are easier to learn than non-pretrained items. The materials used, French–English word pairs, largely overlapped with those employed in their earlier study. Sixty-four English-speaking university students who had little knowledge about French participated in Experiment 1 to look at the effects of the translation direction and practice schedule (i.e., blocked vs. mixed). In Experiment 2, another 48 university students were tested to examine the effects of the translation direction and pretraining.

In Experiment 1, participants were divided into two groups. Half of them were to learn French–English word pairs with French words as cues and English words as responses. The other half were to study English–French word pairs with English words as cues and French words as responses. The participants in each group were further subdivided into the blocked presentation condition and the mixed-order presentation condition as in their previous study (Schneider *et al.*, 1998). Learning procedures were basically the same as that employed in the previous study, except the number of practice trials was fixed to three. Participants viewed 25 cue–target pairs on a computer screen in blocks of five at a rate of 2 seconds. Participants in the L2-to-L1 learning condition studied French–English word pairs (e.g., *voiture*–car) consecutively one at a time and then typed the English-language translation equivalent from a French cue word (e.g., *voiture*–?). Conversely, participants in the L1-to-L2 learning condition learned five English–French word pairs and typed a French word based on an English-language cue word (e.g., car–?). In each condition, half of the participants saw a set of five word pairs that were conceptually related, and the other half saw a mixed set. Training continued until the participants completed the learning process and typed of all the 25 word pairs three times. After the three training trials, participants took an immediate memory test in which they were required to type target words from cue words in the designated translation direction. After a week, participants took the memory test again and performed three relearning trials to assess the ease of relearning. At this time, half of the participants in each condition were required to respond in the opposite translation direction (e.g., a participant in the blocked English-to-French learning condition on day 1 might take the test in the French-to-English direction and relearning in a mixed presentation). Therefore, the delayed test and the delayed relearning on day 2 involved retention and transfer depending on whether the types of testing and relearning were the same or differed from that used a week before.

This experiment yielded three new findings of task difficulty for retention and transfer. First, translation direction as a variable of task difficulty functioned well regarding retention. A significant interaction between the translation direction (i.e., French cues or English cues) during the learning session and the time of learning (i.e., learning or relearning) was noted. Participants who recalled English words from French cues (i.e., easy learning condition) produced more correct responses than participants given English cues to recall French words (i.e., difficult learning condition) during learning, but this trend reversed during the relearning period a week after. Thus, a difficult learning condition operationalized by the translation direction during learning rendered superior retention than the easy learning condition. Second, the translation direction also affected subsequent transfer. At the delayed posttest, participants who received French cues (i.e., easy condition during a delayed test) but had practiced using

English cues in learning on day 1 (i.e., difficult learning condition) performed better than those who practiced from French cues on day 1 (i.e., easy learning condition). That is, learning in a difficult condition promoted subsequent transfer. In contrast, participants given English cues during the delayed posttest (i.e., difficult condition during delayed test) who had practiced with French cues on day 1 (i.e., easy learning condition) performed poorly compared to those who practiced using English cues on day 1 (i.e., difficult learning condition). Therefore, learning in an easy condition led to a lack of transfer measured with the opposite translation direction on the delayed posttest. Finally, it was found that other difficulty variables could interfere with a variable of difficulty. Practice schedule (i.e., blocked or mixed) exhibited only partial effects on performance during learning and the immediate test. Specifically, participants in the easy blocked presentation condition performed better than participants in the difficult mixed presentation condition during learning but this was found only for participants given French cues (i.e., easy in translation direction). An advantage also emerged for the blocked presentation on an immediate retention posttest, but this occured, again, only for participants given French cues. These results suggested that the effects of practice schedules, which seemingly had worked well in their previous study as a control variable for difficulty manipulation, were weakened when used with other components of difficulty.

Experiment 2 was carried out with pretraining manipulation instead of the practice schedule used in Experiment 1, while manipulations in the translation directions remained the same. In this experiment, all participants studied half of the French words, but not English words, of all word pairs prior to training. The researchers conducted this pretraining by displaying French words on a computer screen one at a time, then, participants typed those words immediately after each presentation.

Participants repeated this learning process for the designated French words three times. After pretraining, participants learned cue–response pairs in the same way as in Experiment 1: Half of them were trained in the French-to-English translation direction (i.e., easy learning condition) and the remaining half in the English-to-French direction (i.e., difficult learning condition). Furthermore, as in Experiment 1, they were given an immediate retention posttest. A week later, they took a delayed posttest and relearned the target word pairs, in which half of the participants in a given condition were placed in the opposite condition. Here, the researchers predicted that (a) a difficult learning condition would lead to limited performance both during learning and on the immediate posttest, but it would lead to superior performance during relearning and on the delayed posttest a week after, and (b) pretraining would lead to better performance during learning but not during relearning, especially for those recalling French words.

The primary findings of this experiment are twofold. First, the translation direction still played a significant role in both retention and transfer. Word pairs initially learned in a difficult condition in terms of the translation direction facilitated subsequent relearning. The retention loss rate also decreased after a week for the difficult English-cued learning condition. Second, pretraining showed moderate effects on retention. Participants performed significantly better during learning for pretrained word pairs than those not pretrained, especially when French responses were elcited from English cues. However, during the relearning session a week after, the participants' performance on pretrained pairs was no better than those not pretrained, rather, performance on unpretratined pairs was slightly better. Therefore, the researchers concluded that the effects of pretraining decreased as participants became familiar with those items through trials.

In sum, these two studies together confirmed the following facilitative effects on retention and transfer, which were led by increasing task difficulty: (a) a greater degree of difficulty in practice schedule-enhanced subsequent retention, (b) a greater degree of difficulty brought by manipulating the translation directions promoted both retention and subsequent transfer, and (c) a greater degree of difficulty brought by the absence of pretraining had marginal effects on retention. Based on these facilitative effects of task difficulty on retention and transfer, Schneider *et al.* argued that "any method used to increase difficulty during learning would lead to superior retention and transfer of the learning material" (Schneider *et al.*, 2002, p. 422), a source of the training difficulty principle in their later publication (Healy, Schneider, & Bourne, 2012). Meanwhile, the researchers also noted that these difficulty components would vary in terms of strength. Recently, Bjork and Kroll (2015) mentioned that bilingualism, operationalized in translation directions, would constitute one of the most reliable variables to produce difficulty in language-learning tasks.

2.6.5 Task Complexity and Difficulty

While Schneider *et al.* (2002) treated task difficulty as a single construct involving interference in learning tasks, Robinson (2001a, 2001b, 2003) classified task interference into separate aspects: those attributed to the inherent nature of a particular task (i.e., task complexity) and learner factors (i.e., task difficulty). In this framework, task complexity refers to a degree of demand for information processing that a particular task imposes on the learner. Task difficulty, in contrast, refers to the extent of the learner's capacity to direct cognitive resources to the task at hand. For example, generally, reading an advertisement flier is easier than reading a PC manual. However, even in such a case, performance could differ depending on the reader's affective state or motivation. Moreover, naturally speaking, IT engineers can read PC manuals more fluently than computer beginners. In this way, different types of interference can influence learners' apparent performance on a given task. Therefore, it is important to classify sources of task interference when designing a set of effective learning tasks.

Robinson (2003) mentioned that his term “task complexity” is a within-learner variable and that “task difficulty” is a between-learners variable for interference on a learning task. He further contended that learners can only develop their interlanguage systems by accessing factors of the task complexity dimension, which can be operationalized in cognition-oriented conditions such as +/– few elements, +/– here-and-now, +/– reasoning, +/– planning time, or +/– single task. Aspects of task difficulty, such as levels of motivation, aptitude, or proficiency, account for additional effects, positively or negatively, on learners’ attentional resource allocation to and learning of the target structures imposed by task complexity. Task performances are the results of interactions between task complexity and task difficulty. For example, a +/– here-and-now condition for pushing a use of the past-tense morphology might be more challenging for beginners than advanced learners, but a +/– planning time might impose an equal level of cognitive demand on a wide range of learners.

In Robinson’s framework, the factors of task complexity encompass two categories: resource-directing dimensions and resource-depleting dimensions. Resource-directing dimensions of task complexity consist of factors responsible for grammatical encoding, such as +/– few elements (e.g., pluralization), +/– here-and-now (e.g., past-tense marking), and +/– reasoning (e.g., if-then, because, therefore, or other logical connectors). In contrast, resource-depleting dimensions of task complexity (e.g., +/– planning time, +/– single task, and +/– prior knowledge) mainly deal with learners’ availability of attentional resources to allocate to linguistic demands but not linguistic codes themselves. Therefore, it is hypothesized that increasing task complexity along with resource-directing dimensions facilitates the development of learners' interlanguage systems, while increasing task complexity along with resource-depleting dimensions benefits the development of language automaticity.

2.6.5.1 Some Learning Effects on the Here-and-now Dimension of Task Complexity

As mentioned above, in Robinson's framework of task complexity, resource-directing variables are responsible for learners' language development, thus second-language acquisition researchers have investigated this topic as a primary area of research interest. In their meta-analysis of research on second language task complexity, Jackson and Suethanapornkul (2013) reported that the +/– here-and-now dimension was the most frequently used variable in pedagogic tasks, followed by the +/– few elements dimension and the +/– reasoning dimension. They also mentioned that the manners in which the +/– here-and-now condition was operationalized were relatively consistent throughout the existing studies. Typically, learners were encouraged to perform tasks with (i.e., + here-and-now) or without (i.e., – here-and-now) contextual support. The minus here-and-now condition also constitutes a "there-and-then" condition, as opposed to here-and-now.

In an early study of task complexity on pedagogical second-language learning tasks, Robinson (1995) compared narratives spoken in English by 12 non-native speakers between a context-supported condition (i.e., here-and-now) and a no context-supported condition (i.e., there-and-then). He assumed that a cognitively complex condition would elicit a dysfluent but more complex and accurate use of language because speaking without contextual support requires speakers to pay attention to their linguistic knowledge compared to when speaking while looking at material. Participants were shown a wordless cartoon strip and instructed to tell the story in English. In the here-and-now condition, they spoke while looking at the cartoon. In the there-and-then condition, the cartoon was taken away before participants began speaking. All participants were tested in both conditions in which different versions of the cartoons were provided. According to the results, he found that the +/– here-and-now manipulation did not affect the fluency of the participants' narratives, but they used a significantly greater variety of content words in the complex there-and-then condition than in the simple here-and-now condition. Moreover, their use of articles (e.g., "a" or "the") was more accurate in the there-and-then condition than the here-and-now condition.

Robinson, Cadierno, and Shirai (2009, Study 1) used data from the work of Robinson (1995) and further analyzed them in light of the aspect hypothesis (Andersen & Shirai, 1994), which predicts that morphological tense-aspect marking is developed in a specific order. For instance, achievement verbs, those describing movements with an instant endpoint (e.g., arrive), are cognitively easier for learners to produce in the past tense than state verbs (e.g., love), activity verbs (e.g., run), and accomplishment verbs (e.g., make). In contrast, progressive marking (i.e., -ing) is easiest for activity verbs compared to the other types of verbs. In relation to Robinson's framework of task complexity, the researchers hypothesized that the complex there-and-then narratives would contain more advanced tense-aspect markings than the simple here-and-now narratives, such as past-tense markings for state verbs or activity verbs, or progressive marking for achievement verbs. Analyses of frequencies of the non-prototypical use of tense-aspect markings between the simple and complex task conditions revealed that 55% of past-tense markings and 54% of progressive markings were made for prototypical, basic verbs in the here-and-now condition, whereas they were only 18% and 33%, respectively, in the there-and-then condition. The difference in a combined proportion of non-prototypical tense-aspect markings between the simple and the complex task approached a statistical significance, indicating that past and progressive morphologies were adopted for various types of verbs in the there-and-then condition compared to the here-and-now condition. Therefore, the researchers concluded that a cognitively demanding task promotes an advanced use of tense-aspect morphologies more than a less cognitively demanding task.

Revesz (2009) also employed the +/− here-and-now dimension of task complexity in her investigation of the interaction between task complexity and recasts. She hypothesized that recasts for erroneous utterances during a pedagogic second-language task would prove more effective in a cognitively complex task than a simple task. In this study, 90 Hungarian learners of English performed a role-playing task as witnesses at a crime scene as a treatment task. The dependent measures included the number of recasts made in the treatment task and gains in the accuracy of past-progressive forms between a pretest and a posttest. In the treatment task, participants were first shown a photo that was supposedly taken at the time a crime occurred. Their task consisted of describing the photo to the researcher who played the role of a police officer. Half of the participants described the crime scene without looking at the picture (i.e., there-and-then), while the remaining half spoke while looking at it (i.e., here-and-now). Both groups received the researcher's recasts for erroneous utterances. The results indicated that the + photo group received more recasts than the – photo group, suggesting that participants produced more errors in the simple here-and-now condition than the complex there-and-then condition. Furthermore, pre- and post gains for the target structure were significantly higher for the – photo group than the + photo group. These results suggested that cognitively complex tasks not only the enhance accuracy of utterances but also the effectiveness of recasts.

Chapter III

SECOND-LANGUAGE VOCABULARY ACQUISITION FRAMEWORKS

3.0 Introduction

This chapter reviews theories of vocabulary acquisition in the second-language acquisition literature, especially those focusing on the retention of newly acquired words. Not many theories clearly specify the relationship between vocabulary learning and the mechanisms behind it. Moreover, this chapter does not deal with the effects of input frequency, one of the general ideas or facts underpinning (with many replications) frequency-based vocabulary tests, such as vocabulary levels tests (Nation, 1990; Read, 1988), to deal more fully with issues directly relevant to the scope of the present study.

This chapter begins by introducing an early study on the acquisition of new words, specifically one concerning encoding processes. Bull and Wittrock (1973) used pictures and mental imagery to examine 87 fifth-graders' retention of the definitions of novel nouns. Students were assigned to one of three conditions: verbal definition only, verbal definition plus illustration provided by the researchers, and verbal definition plus self-discovered imagery, in which the students drew their own illustrations of the definitions. Recall on a multiple-choice test administered a week after learning revealed that the self-discovered imagery group performed significantly better than the definition-only group. Overall, the results supported their prediction that the learning and retention of definitions would be facilitated most when students created their own images for the definitions, followed by those giving images developed by someone else and those who copied the definitions (Bull & Wittrock, 1973, p. 292). Although these differences were not statistically significant, the better retention rate of the self-discovered imagery group could be interpreted as resulting from the act of receiving different levels of involvement load in their nonverbal codes. Thus, these findings provide us an insight into the importance of cognitive load on nonverbal material codes in vocabulary acquisition to some extent.

The following sections review three theories dealing with the acquisition and retention of new words that have a central focus that relates to cognitive processes during learning. Specifically, this section discusses (a) the dual coding approach of human verbal memory, widely acknowledged in the psychology literature; (b) the involvement load hypothesis, which specializes in vocabulary learning in the field of second-language acquisition; and (c) technique feature analysis, a set of hypothetical measures for successful vocabulary learning derived after the involvement load hypothesis.

3.1 Dual Coding Theory

3.1.1 The Theory—Paivio (1971, 1986)

The dual coding theory (Paivio, 1971, 1986) holds that word knowledge is stored in two separate but partly interconnected systems: a verbal system and a nonverbal system. The contents of the two systems vary in accordance with sense modalities. For example, in the visual modality, the verbal system might hold a mental representation of an orthographic form such as "CUP," while the nonverbal system might deal with a mental representation of an object or a scene, such as a coffee cup with a saucer. In the auditory modality, the verbal system holds the sound of the word, /kup/, while an environmental sound such as the clinking of cups on saucers might be coded in the nonverbal system (Sadoski & Paivio, 2001). Therefore, the theory maintains that the nonverbal system possesses meanings of words based on worldly knowledge about objects or events, and the verbal system consists of labels referring to those meanings.

The dual coding theory assumes that information encoded both verbally and nonverbally should result in better retention than information stored in only one code. Information stored in two codes could lead to more elaborate processing than single-coded information, thus promoting better comprehension of inputs and generating stronger memory traces. Therefore, people retain concrete words more readily than abstract words. Furthermore, the theory assumes that the connection between the verbal and nonverbal systems is not necessarily a one-to-one but a one-to-many linkage. A major difference between concrete and abstract words, in this theory, is that a label of each concrete word has more direct connections with the referent imagery than abstract words have, thus, they are imageable. For example, a concrete word “tree” evokes referent images of trees of various kinds from the learners’ experiences, such as trunk, roots, branches, or leaves (Paivio, 1986, p. 121–123). Paivio (1971) argued that recall increases with concreteness and that dual coding enhances retention: “The increased availability of both codes increases the probability of item recall . . . [O]ne could be forgotten during the retention interval, but verbal recall [may] still be possible provided that the other is retained” (Paivio, 1971, p. 207).

3.1.2 Recall of Concrete and Abstract Sentences—Sadoski, Goetz, and Fritz (1993)

Sadoski, Goetz, and Fritz (1993) investigated the immediate and delayed recall probabilities of factual sentences about historical figures consisting of concrete and abstract sentences. In this study, 79 students at Texas A&M University read 40 short sentences about 10 historical figures of varying familiarity. Participants were randomly assigned to one of four conditions: concrete–concrete, concrete–abstract, abstract–concrete, and abstract–abstract. In the concrete–concrete condition, four factual sentences for each of the 10 figures were relatively concrete. In the concrete–abstract condition, the first two of the four sentences were concrete, and the latter two were abstract. Similarly, in the abstract–concrete condition, the first two were abstract and the latter two were concrete, and in the abstract–concrete condition, four sentences for each of the 10 figures were all abstract. The concreteness of each of the 40 sentences had been confirmed in advance with ratings of the other 22 students on a 7-point Likert scale. Participants then received a booklet including the 40 sentences consisting of 94 idea units. Immediately after reading the booklet, participants were given another booklet with the names of the 10 historical figures as cues and were asked to write down everything they could remember from what they had just read as accurately as possible. Five days later, the participants again received the same booklet to provide written recall. The results showed that the abstract–abstract group recalled significantly less than the other experimental group on both the immediate and delayed posttests. Furthermore, a significant correlation was found between the other 22 participants' sentence ratings on concreteness and the probability of recall on the immediate (r = .73) and delayed posttests (r = .65). Thus, the researchers concluded that concreteness is highly predictive of recall.

3.1.3 Foreign-Language Learning from Media Exposure—Kuppens (2010)

Evidence for the dual coding theory has been reported not only from laboratory-based studies but also based on a survey on children's long-term use of media in natural settings. Kuppens (2010) investigated relationships between Dutch-speaking children's ability in English and their exposure to English-language media. In this study, 361 pupils, aged 11 on average, who had not yet received formal English instructions, answered a survey asking the use of a range of English-language media, such as subtitled television programs and movies, music, and computer games. After filling out the survey, the pupils' English skills were tested by having them translate eight short sentences from Dutch to English (e.g., "Ik heb je gemist. → I missed you"; "Je hebt gelijk. → You're right"), as well as eight other short sentences from English to Dutch (e.g., "I feel great, Nice to meet you"). Because participants were preteens without formal English-language education, the researchers conducted these tests orally. An analysis of the survey and the English tests indicated that pupils who watched English-subtitled television programs or movies performed significantly better on the translation task than those who had had fewer opportunities to engage with English-related media. Specifically, pupils who watched English-subtitled television or films six times or more a week translated Dutch sentences into English significantly better than those who watched them only three to five times a week. In turn, the latter were significantly better on Dutch-English translation than participants who watched English-subtitled television or films less than three times a week. This pattern also emerged in their performance on English-Dutch translation task. Furthermore, pupils who played English computer games three times or more a week translated the sentences from English to Dutch significantly better than those who played English computer games less than three times a week. In fact, this study did not mention the dual coding theory, but the findings imply that children's incidental foreign language acquisition could be facilitated by watching images and languages at the same time.

3.1.4 Dual Coding with Olfactory Information—Lwin, Morrin, and Krishna (2010)

Scholars have extended the dual coding theory beyond its traditional focus on verbal and visual information to olfactory information. Lwin, Morrin, and Krishna (2010) manipulated the presence or absence of olfactory information in addition to visual stimuli, both at encoding and retrieval to examine their effects on verbal recall.

In this study, participants learned information about a cosmetic product that had been delivered in an envelope as if they had received a piece of direct mail from a hypothetical company. The researchers focused on two issues: (a) the effect of olfactory information at encoding and (b) the effect of scent-based cues at retrieval, each examined in two separate experiments. In the first part of the study examining the effects of olfactory information at encoding, 100 females received an envelope containing an advertisement about a moisturizer, and they were asked to inspect it in four between-participants conditions. In the "no image–no scent" condition, the advertisement contained only the brand name, several copy points about the product, and a picture of the product. In the "image–no scent" condition, two pictures of flowers were added to the advertisement. In the "no image–scent" condition, the advertisement contained only the brand name, the copy points, and the picture of the product, but a rose-sandalwood scent was infused into the paper. In the "image–scent" condition, the advertisement contained both the floral images and the scent. The researchers took the direct mail away immediately after participants had examined it briefly. After a 5-min retention interval, participants were asked to recall everything they knew about the product in two ways. Unaided recall, performed first, required them to recall without any cues. Aided recall, performed next, provided the scent as a cue to remember anything else about the product. Participants were tested in this way again after two weeks. This experiment revealed significant facilitative effects of scent on long-term verbal memory. In short-term spans (i.e., 5 minutes), olfactory information facilitated verbal retention at a level comparable to the effect of visual information. However, in the long term (i.e., 2 weeks), regardless of the presence or absence of pictorial information, participants who encoded information with scent retained verbal information significantly more than those who did not receive a scented envelope.

The facilitative effects of olfactory information on verbal recall were further confirmed in their next experiment, which examined the effects of scent as a cue at the time of retrieval with a larger sample size. In this experiment, 140 undergraduates were given the fully informative advertisement at the time of encoding; the same material provided in the "image–scent" condition in the first experiment. The four-way manipulation of the image (i.e., present or absent) and scent (i.e., present and absent) was made only at the time of retrieval a week after encoding as the between-subjects variable. The results were consistent with the facilitative effect of scent information found in the first experiment. Participants given scent as a cue in one-week delayed retrieval remembered significantly more information about what they had learned a week before than those who recalled without scent as a cue. Therefore, the researchers concluded that scent-based retrieval potentiates facilitative effects on recall, especially when used together with pictorial information.

3.2 The Involvement Load Hypothesis

3.2.1 The Theory—Laufer and Hulstijn (2001)

The involvement load hypothesis (Laufer & Hulstijn, 2001) states that the learners' levels of involvement while learning new words determine the subsequent retention of newly learned words. The term "involvement load" refers to a sum of three motivational-cognitive indices that a particular task imposes, including need, search, and evaluation. This theory overlaps in many parts with the depth-of-processing framework (Craik & Lockhart, 1972), which claims that elaborate processing enhances retention.

First, need is the motivational component of the involvement load and refers to the degree to which learners require new words to complete a given task. Need has three levels. Need is hypothesized to be strong when imposed by the learner, such as when the learner wants to know the concepts related to the word. Need also could be moderate when imposed by a task or a teacher, as when a teacher asks them to use the word. Need is empty when the new word does not relate to the completion of the task.

Second, search is the cognitive component of the involvement load and indicates the presence or absence of the learner's attempt to seek the word during learning. Search is a dichotomous index, and is present when learners check the meaning of unknown words to complete a task (e.g., when learners consult a dictionary for the meaning). Search is absent when there is no need for such an effort. For instance, search does not occur when the meaning of an unfamiliar word is given by a glossary in the margins.

Evaluation constitutes another cognitive component of the involvement load and refers to the extent to how precisely the unknown word is compared with other plausible alternatives to make a decision to use it. Evaluation is a three-scale index. Evaluation is strong when the task requires understanding how other words could be combined in production (e.g., *military stroll vs. military march), as in composition writing. Evaluation could be moderate when such a decision is made by receiving an answer. Evaluation is empty when there is no need to make such decisions.

The magnitude of the involvement load on a given task is determined as a combination of the degrees of these three components. The theory assumes that words processed with a higher involvement load are retained better than those processed with a lower involvement load. Because the values of each involvement component and the corresponding conditions are explicitly defined, this framework is empirically testable. Laufer and Hulstijn (2001, p. 18) found the components to be activated on specific tasks as follows.

1. Reading and comprehension questions, included in the text but irrelevant to task.

   (Need: 0, Search: 0, Evaluation: 0)

2. Reading and comprehension questions, included in the text and relevant to the task.

   (Need: 1, Search: 0, Evaluation: 0)

3. Reading and comprehension questions, not included but relevant to the task.

   (Need: 1, Search: 1, Evaluation: 0)

4. Reading and comprehension questions and filling gaps, relevant to reading comprehension, or listed in a glossary at the end of the text.

   (Need: 1, Search: 0, Evaluation: 1)

5. Writing original sentences, listed in a glossary.

   (Need: 1, Search: 0, Evaluation: 2)

6. Writing a composition on concepts selected by the teacher or when the L2 learner-writer must look up the L2 form.

   (Need: 1, Search: 1, Evaluation: 2)

7. Writing a composition on concepts selected by the L2 learner-writer.

   (Need: 2, Search: 1, Evaluation: 2)

3.2.2 Effects of Involvement Load on L2 Beginners—Keating (2008)

Keating (2008) investigated the hypothesized benefits of the task-induced involvement load for the retention of new words by using Spanish-like nonwords for beginners of Spanish. In this study, 79 English-speaking students enrolled in elementary-level Spanish classes at the University of Illinois learned eight Spanish-like nonwords as targets, divided into three conditions. The conditions differed in terms of the levels of the task-induced involvement load introduced by Laufer and Hulstijn (2001). The first condition involved reading comprehension with a glossary in the margins. Participants in this condition were provided with a text and a set of five true–false comprehension questions about the passages. In terms of the strength of the involvement load, need was moderate, because knowledge of the target words was required both to understand the passages and to answer the comprehension questions. Search and evaluation were both absent. Therefore, the involvement load index was 1 (i.e., need: 1, search: 0, evaluation: 0). The second condition entailed reading comprehension and filling gaps. In this condition, the target words in the text were replaced with blank spaces. The target words appeared on a separate page, intermixed with four extra words that were not in the original text. Participants in this condition read the text, filled in the blanks with the words from the list, and then answered comprehension questions. This task induced moderate need and moderate evaluation but nothing in search. Therefore, the involvement load index was 2 (i.e., need: 1, search: 0, evaluation: 1). The third condition required writing original sentences using the target words. In this condition, the researchers did not provide a text. Instead, participants were asked to review a word list, the same list provided in the second condition, for 10 minutes, and then write original sentences by using the target words for another 10 minutes.This task induced moderate need and strong evaluation. Evaluation was strong because participants needed to produce new words in combination with previously known words in composition writing. Therefore, the involvement load index was 3 (i.e., need: 1, search: 0, evaluation: 2). Participants' knowledge of the new words was tested immediately after the treatment task and two weeks later, both receptively and productively. On a passive recall posttest, participants translated the new words (i.e., Spanish-like nonwords) into English. In an active recall posttest, participants translated English sentences into Spanish by using the target nonwords.

The results were consistent with the predictions of the involvement load hypothesis. Retention was highest in the sentence-writing condition (i.e., need: 1, search: 0, evaluation: 2) and lowest for those doing the reading comprehension task with a glossary in the margins (i.e. need: 1, search: 0, evaluation: 0), the reading comprehension plus filling in the gaps condition (i.e., need: 1, search: 0, evaluation: 1) was in between the other two in terms of immediate passive recall and immediate active recall, as well as delayed passive recall. Moreover, the differences in retention between the reading plus fill-in condition and the reading a glossary condition were all significant regardless of test type (i.e., passive or active) and time (i.e., immediate or delayed), whereas the differences were marginal between the reading plus fill-in condition and the writing condition. Therefore, the researchers concluded that the evaluation component of task-induced involvement plays a crucial role in learning new words.

3.2.3 Effects of Involvement Load Between Different Proficiency Levels—Kim (2008)

Kim (2008) compared the effects of task-induced involvement between two different populations in English proficiency. Sixty-four adult learners of English from a variety of countries participated in the study. Among them, 34 were undergraduates with TOEFL scores above 520 and thus were acknowledged as highly proficient learners of English in this study. The remaining 30 learners, classified as less proficient learners, had TOEFL scores ranging from 470 to 520. The three treatment tasks with the varying task-induced involvement loads employed in this study were very similar to those used in the work of Keating (2008), namely, (a) reading comprehension with a glossary in the margins, (b) reading comprehension with a glossary plus gap-filling, and (c) writing a composition using target words. The involvement load indices were 1, 2, and 3, respectively. The target vocabulary words consisted of 10 English words that the participants' teachers assumed their students did not know (e.g., vexed, caveat, or divulge). Unlike Keating's (2008) approach, in this study, knowledge of new words was measured by self-reporting about them on a 5-point scale format: 1 (not familiar), 2 (familiar but the meaning is not known), 3 (correct synonym or translation is given), 4 (the word is used with semantic appropriateness in a sentence), and 5 (the word is used with grammatical and semantic appropriateness in a sentence). According to the researcher, the use of this rather unclear measure was inevitable because the participants' first languages varied. Levels of knowledge of the new words were obtained immediately after the treatment and two weeks after, then, they were compared in task type as a between-subjects factor and in proficiency as a between-groups factor.

A one-way analysis of retention for task type showed the hypothesized effects of the task-induced involvement. The rate of retention of the composition group was greater than the gap-filling group, followed by the reading group, regardless of their proficiency levels, both on the immediate posttest and the delayed posttest. Furthermore, the differences between each of the three task types were independently significant on the delayed posttest, suggesting strong long-term effects of the involvement load. However, a two-way comparison between task type and proficiency revealed no significant differences between the two groups, indicating that the equal level of the task-induced involvement load was equally beneficial to learners with various proficiency levels. Therefore, Kim concluded that the involvement load approach could be applied to language learners regardless of their proficiency levels.

3.2.4 Effects of Involvement Load with Listening Materials—Jing and Jianbin (2009)

Jing and Jianbin (2009) explored the effects of the task-induced involvement load by using auditory material. In this study, 86 English-language students learned 15 English words in one of the following three types of tasks. Task A involved listening comprehension with irrelevant glossaries in the margins. In this task, participants could answer the comprehension questions without referring to glossaries, and the task did not induce any involvement load among learners. Therefore, the involvement load index was 0. Task B included listening comprehension with glossaries relevant to the listening passage. On this task, participants needed to complete the listening comprehension questions in a multiple-choice format. Choosing correct answers required reference to the glossaries, thus, need was moderate (i.e., need: 1). Therefore, the involvement load index was 1. Finally, Task C entailed listening comprehension with glossaries relevant to the listening passages plus writing a composition. This task consisted of two parts. First, participants answered the comprehension questions in the same way as in task B. The second part of the task required participants to write a short article by using all the target words. A productive use of new words with selection induced strong evaluation (i.e., evaluation: 2). Therefore, together with moderate need (i.e. need: 1), the involvement load index of this task was 3. Participants' knowledge of the new words was assessed only passively by providing Chinese equivalents (their L1) for each word. Posttests were administered immediately after the treatment and a week after.

The results confirmed the prediction of the hypothesis. Retention of the target words learned through writing was highest, followed by those learned through text-relevant questions. Learning with text-irrelevant questions resulted in the lowest performance on both the immediate and delayed posttests. As a result, this study showed that task-induced involvement load is also effective in learning through listening.

3.2.5 Issues of Time on Task

The involvement load hypothesis holds that tasks with a greater involvement load are more learnable than those with lower involvement load. Naturally, tasks with a high involvement load (e.g., writing a composition) require more time to complete than less loaded tasks (e.g., reading). Therefore, some have criticized the efficiency of this approach. Webb (2005) found that reading was more effective than writing when comparing these tasks in terms of the average time spent. Folse (2006) demonstrated that completing of multiple-choice exercises three times after reading outperformed a single attempt at writing in terms of subsequent retention. Thus, he concluded that writing sentences with new vocabulary words is not only time-consuming but also ineffective. These controversies can be reduced into one basic argument, namely, words learned on a task with a high involvement load are well retained because learners engaged in such high-loaded tasks are typically be exposed to new words longer than for simple tasks.

Nevertheless, Hulstijn and Laufer (2001) maintained that the time needed to complete a given task is an inherent property of the task. The involvement load hypothesis assumes incidental vocabulary learning. “Incidental” is the key point here. That is, for example, the time needed to consult a dictionary for the meaning of an unknown word is a part of the involvement load the task imposed. Hulstijn and Laufer (2001) considered time on task to be a variable influenced by the degree of the involvement load.

3.3 Technique Feature Analysis

3.3.1 The Theory—Nation and Webb (2011)

Technique feature analysis (Nation & Webb, 2011) is a theoretical framework for second-language vocabulary acquisition that extends the involvement load hypothesis by providing a detailed set of criteria. Nation and Webb (2011) argued that other motivational, cognitive, and environmental factors were, in fact, necessary in vocabulary learning in addition to those focused on in the involvement load hypothesis. For example, Nation (2001) suggested that setting the goal of an activity, enhancing noticing with negotiation, providing retrieval opportunities, and encouraging a generative use of words were among the key properties for successful vocabulary acquisition. Therefore, this framework consists of the following five components as major criteria: motivation, noticing, retrieval, generation, and retention. Furthermore, for these criteria to be empirically operationalized, each criterion unit has a set of several specific sub-criteria and point values, resulting in an 18-criterion framework, as shown below (see also Nation & Webb, 2011, p. 8–10, for detailed features of the criteria and reasons for their inclusion).

Motivation

- Is there a clear vocabulary learning goal? [0 / 1]
- Does the activity motivate learning? [0 / 1]
- Does the learner select the words? [0 / 1]

Noticing

- Does the activity focus attention on the target words? [0 / 1]
- Does the activity raise awareness of new vocabulary learning? [0 / 1]

Does the activity involve negotiation? [0 / 1]

Retrieval

Does the activity involve the retrieval of the word? [0 / 1]

Is it productive retrieval? [0 / 1]

Is it recall? [0 / 1]

Are there multiple retrievals of each word? [0 / 1]

Is there spacing between the retrievals? [0 / 1]

Generation

Does the activity involve generative use? [0 / 1]

Is it productive? [0 / 1]

Is there a marked change that involves the use of other words? [0 / 1]

Retention

Does the activity ensure successful linking of form and meaning? [0 / 1]

Does the activity involve instantiation? [0 / 1]

Does the activity involve imaging? [0 / 1]

Does the activity avoid interference? [0 / 1]

In proposing technique feature analysis, Nation and Webb (2011) did not provide any direct evidence from empirical examination for their framework. Instead, they calculated scores in both the involvement load hypothesis and technique feature analysis and contrasted them to show that tasks with the same involvement load could be differently evaluated in their framework. For example, in the case of a true–false question task for a given text, the involvement load index is 3, but the score of the same task in technique feature analysis is 6. In the case of a word card activity, such that a learner writes an unknown second-language word on one side of a card and its translation on the other side and goes through the cards at increasingly spaced intervals, the involvement load is 3, but this figure becomes 11 in their framework.

Nation and Webb (2011) stated that this analysis was one possible way of examining the effectiveness of vocabulary teaching. Meanwhile, they also noted that these scales of analysis were clearly in an early stage of development during their study.

3.3.2 Empirical Comparison with the Involvement Load Hypothesis—Hu and Nassaji (2016)

Hu and Nassaji (2016) compared the predictive power of learning between the involvement load hypothesis (Laufer & Hulstijn, 2001) and technique feature analysis (Nation & Webb, 2011). As technique feature analysis is a relatively new concept, empirical research on this topic remains scarce. This was probably one of the first investigations that empirically looked at the framework in question. In this study, 96 Taiwanese college students learned 14 English words previously unknown to them at a pilot stage in one of four task conditions. Two of the four tasks predicted an equal level of learning in both frameworks. The other two tasks predicted contradictory outcomes. The design of the tasks is summarized below.

Task 1: Reading and multiple-choice items	ILH: 3	TFA: 6
Task 2: Reading and choosing definitions	ILH: 3	TFA: 6
Task 3: Reading with filling gaps	ILH: 2	TFA: 7
Task 4: Reading and sentence rewording	ILH: 3	TFA: 6

Tasks 1 and 2 were hypothesized to be equally effective in both frameworks, thus, new words should be retained equally. In contrast, task 4 should be better than task

3 when assuming the involvement load hypothesis, whereas the opposite should be predicted based on technique feature analysis. The knowledge of new words was assessed immediately after the treatment by means of English–Chinese translations.

Cross-task comparisons of retention showed that participants engaged in task 3 remembered significantly more words than those working on task 4. Moreover, the retention of words was better for task 3 than task 1. These results indicated the advantages of technique feature analysis over the involvement load hypothesis. Therefore, the researchers concluded that technique feature analysis is a stronger predictor of word learning than the involvement load approach. Pedagogical implications could also be drawn. As suggested by Nation and Webb, instructors could make use of this framework as a useful tool for effective teaching.

Chapter IV

THE PRESENT STUDY—STUDY 1

4.0 Introduction

The present study investigates the retention and transfer of new second-language vocabulary words, along with the depth-of-processing framework, with a particular focus on the initial acquisition of unknown words. At the beginning of the chapter, we would like to summarize the previous review chapters and state the motivation for the present research.

The first chapter provided an overview of the theoretical mechanisms involved in the human memory system. This introductory chapter articulated the basic ideas on the processing view of the human memory system, which holds that processes are highly responsible for the retention of inputs, in contrast to the traditional storage views. The depth-of-processing framework (Craik & Lockhart, 1972), one of the most influential theories in the processing view, for example, argues that processes at a deep level at the time of encoding promote the retention of inputs because such processes involve semantic processing, as opposed to shallower processing in which only the physical features of words, such as of those where only orthographic or phonetic information is processed. Another influential theory in this field, transfer-appropriate processing (Bransford *et al.*, 1979; Morris *et al.*, 1977), suggests that conditional similarities in encoding and testing affect recall performance, suggesting that shallow processing could also be effective when shallow features (e.g., orthographic or phonological aspects) of words are tested at recall. Nevertheless, because understanding the meanings of words is a crucial part of language learning (vanPatten, 1996), and because general language teaching does not presuppose any particular test, the ideas of transfer-appropriate processing would serve as complementary notions for the depth-of-processing paradigm, at least when looking at these two processing theories from a pedagogical viewpoint.

In response to the aforementioned importance of elaborate processing for durable memory, the second chapter discussed some theories of skill acquisition that inherited, at least partly, the idea of retention advantages led by deep processing. These theories include the generation effect (Slamecka & Graf, 1978), distributed practice (Cepeda *et al.*, 2006), task difficulty imposed by contextual interference (Battig, 1972; Carlson & Yaure, 1990), and a more generalized claim of the task difficulty principle (Schneider *et al.*, 2002), which suggests that difficult tasks promote subsequent retention and transfer. Among these research paradigms, the idea of the introduction of "difficulty" into learning tasks can be viewed as a framework that deals with the depth-of-processing framework in the most inclusive.

In this line of research, we found the task difficulty principle (Schneider *et al.*, 2002), one of the strongest claims associated with benefits from difficult tasks, to be a theoretically stimulating idea that needs further investigation. The reasons for this are twofold. First, the task difficulty principle casts doubt on the general intuition that what is easy to learn is retained longer than difficult concepts. More theoretically, it challenges the cognitive load theory (Chandler & Sweller, 1991), which maintains that heavy cognitive operations would deprive the cognitive resources responsible for memory, thus limiting the elaborate encoding of inputs. Second, task difficulty is an ongoing research agenda that is now yielding a number of second-language acquisition studies, including a viable factor that contributes to the optimization of training (e.g., the special issue of *The Modern Language Journal*, Fall 2019, featuring studies on optimizing second language practice).

In Chapter 3, we looked at three theoretical frameworks for effective second-language vocabulary acquisition characterized by elaborative processing at encoding.

First, the dual coding theory (Paivio, 1971, 1986) argues that materials are encoded in greater detail when input processing involves two different modalities—for example, orthographically as well as phonologically—than when the process is unimodal. The theory explains that information in different types of modality does not compete with other forms of information in the encoding process, allowing sophisticated processing in both quality and quantity, thus leading to superior retention of the information as a whole.

Limiting the scope to the field of second-language vocabulary acquisition, we looked at two more frameworks regarding the retention efficiency of new words. Both frameworks stand on the basic notion of the depth-of-processing view: A degree of elaboration in encoding determines subsequent retention. Whereas the levels of efforts made during the learning process are generally ambiguous aspects, the involvement load hypothesis (Laufer & Hulstijn, 2001) provides specific variables to operationalize degrees of what they termed "involvement" in words to learn. Need, search, and evaluation are the three variables used to determine the levels of motivation for task completion, attention to word meaning, and attention to the distinctiveness of the word, respectively. The hypothesis predicts that, the more overall involvement the learner has, the more the knowledge of the word becomes sustainable. The other framework, the technique feature analysis (Nation & Webb, 2011), is also a criterion-based framework consisting of five major variables mentioned in the work of Nation (2001). Motivation is a variable regarding how well the task drives the learner's motivation to achieve the goal of the task. Noticing is a variable for how well the activity enhances the learner's attention on the target words. The retrieval variable concerns whether the task provides retrieval opportunities for the target words. Generation is a variable for the task that provides a generative use of words. Finally, retention constitutes a variable for whether the task includes cognitive support to remember words such as instantiation or imaging.

To sum up, the gist in the previous review chapters can be summarized as follows. In the human memory system, the quality of the encoding process matters to the retention of new information (Chapter 1). This processing quality might be achieved by manipulating the difficulty levels of the task (Chapter 2). However, current frameworks on second-language vocabulary acquisition have not yet touched upon task difficulty (Chapter 3). Therefore, the present study investigates the effects of task difficulty in the context of second-language vocabulary acquisition, with a particular focus on the retention and transfer of new words as learning benefits, as the task difficulty principle insists.

The present study consists of two main parts. The first part of the study involves a replication of the work of Schneider *et al.* (2002) by employing a different population and a different set of target languages from those employed in the original study. The second part of the study applies the notion of the task difficulty principle to a pedagogical task that can be practically useful, especially for typical language-learning classes. Specifically, the retention and transfer of new words through reading tasks entailing different levels of difficulty will be investigated.

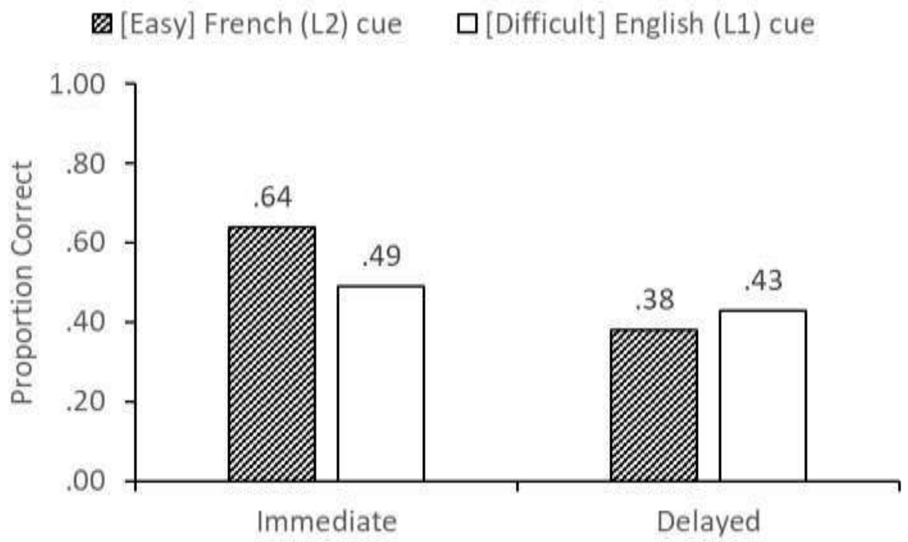
4.1 Research Questions and Hypotheses

As described in Chapter 2, Schneider *et al.* (2002) conducted a vocabulary-learning experiment in which native English speakers with little knowledge of French studied English–French or French–English translation pairs, setting the translation direction as a main dependent variable for task difficulty. They assumed that recalling French words from English cues was more difficult for their English-speaking participants than recalling English words from French word cues. The immediate posttest for retention administered at the end of the day1 training session showed that English words (i.e., an easy French-to-English training condition) were remembered more easily than French words (i.e., a difficult English-to-French training condition), suggesting a short-term advantage of learning in an easy condition. However, when the participants returned a week after and took the test again, the English words were less well remembered, while the French words were retained. Moreover, when tested in the reverse translation direction from a week earlier, participants trained in the difficult English-to-French translation condition on day1 recalled more English words in comparison to the amount of recall of French words that those in the easy French-to-English translation condition needed to recall. Therefore, Schneider *et al.* (2002) concluded that difficulty benefs retention and transfer in the long run.

Fundamentally, Schneider *et al.*'s (2002) study can be characterized by their finding of an inverse proportion of correct responses between the immediate posttest and the delayed posttest as a function of two different levels of difficulty at learning, as illustrated in Figure 4.1.1. Participants given easier French cues during learning performed better on the immediate post-recall test than those given more difficult English cues. However, on the delayed posttest, after a one-week retention interval, those given difficult English cues at learning outperformed those with easy French cues, suggested by the title of their paper—"What is learned under difficult conditions is hard to forget."

Figure 4.1.1

Schneider et al. (2002) Experiment 1: Proportion Correct as a Function of Time of Test and Cue in Learning in the Mixed Condition (based on Table 5, p. 429).



Now, to determine our research questions in terms of a replication study, we will look at their results in Experiment 1 more closely to grasp exactly what they found. Referring to Table 5 in the study by Schneider *et al.* (2002, p. 429), we extracted, in Table 4.1.1, proportions of correct responses at each time of their Experiment 1 posttests in the mixed presentation condition and added labels of the corresponding task types. Table 4.1.1 indicates, as Schneider *et al.* emphasized, that, even though difficult learning sessions would be less effective than easy learning sessions in terms of immediate post-learning recall (easy = .64 vs. difficult = .49), they are effective in the long run (easy = [.63 + .13] / 2 = .38 vs. difficult = [.62 + .24] / 2 = .43). However, the table also reveals the distinct characteristics of each task type. While the retention rates associated with easy learning do not change after a week (immediate = .64 vs. delayed = .63), those associated with difficult learning clearly decrease (immediate = .49 vs.

delayed = .24). This finding contradicts their claim that "what is learned under difficult conditions is hard to forget." Conversely, performance on the transfer tasks is clearly greater for those associated with difficult learning objectives than easy learning ones (easy = .13 vs. difficult = .62). Moreover, accuracy on the transfer posttest associated with difficult learning far surpassed the level of their immediate posttest (immediate = .49 vs. delayed = .62). Therefore, the characteristics of the long-term advantage of difficult learning can be understood to imply that difficult learning boosts delayed transfer performance, thus compensating for the lack of long-term retention. Their results indicated that manipulations of difficulty during learning may not affect both retention and transfer uniformly. Now, a view emerges that easy learning maintains the retention of knowledge but not its transferability, however, difficult learning facilitates the transfer of knowledge but not its retention. They did not mention this view in their original paper.

Table 4.1.1

Schneider et al. (2002) Experiment 1: Proportion Correct for Immediate and Delayed Tests as a Function of Time of Test and Cue in Learning in the Mixed Condition (based on Table 5, p. 429).

Learning	Immediate	Delayed		Task attribution
French (L2) cue	.64	French (L2) cue	.63	Easy learning & retention
[Easy]		English (L1) cue	.13	Easy learning & transfer
English (L1) cue	.49	French (L2) cue	.62	Difficult learning & transfer
[Difficult]		English (L1) cue	.24	Difficult learning & retention

With an emphasis on the replication of the abovementioned findings by

Schneider *et al.* (2002), the research questions and hypotheses of the present research were formulated as follows.

Research Question 1:

> Is learning more efficient on an easy learning task than a difficult one? Specifically, on a paired-word learning task involving translation pairs between participants' first language (L1) and a second language (L2), do participants who learn word pairs in an easy L2-to-L1 direction produce more correct responses at each trial during training, as measured by responses to the translation equivalents of each word pair, than those who learn the word pairs in a difficult L1-to-L2 direction?

Hypothesis 1:

> Participants who learn the word pairs in the easy L2-to-L1 direction will demonstrate superior learning gains compared to those who learn the word pairs in the difficult L1-to-L2 direction. Therefore, learning with less difficulty will be more efficient than learning with more difficulty.

Research Question 2:

> Does a difficult learning condition promote greater delayed recall of learned words than an easy learning condition, in terms of retention and transfer as a whole? Specifically, a week after learning, when testing half of participants in the same translation direction (retention) and the other half in the opposite translation direction (transfer), do those who learned word pairs in a difficult L1-to-L2 direction recall more word pairs than

those who learned the word pairs in an easy L2-to-L1 direction?

Hypothesis 2:

Participants who learn word pairs in the difficult L1-to-L2 direction will produce more correct responses a week after learning than those who learn the word pairs in the easy L2-to-L1 direction. Therefore, learning with a certain degree of difficulty will be efficient in the long run.

Research Question 3:

Do the effects of learning on retention and transfer show distinct patterns at two different levels of training difficulty? Specifically, when tested a week after, does an easy learning condition lead to the retention of knowledge but not transfer? In contrast, does a difficult learning condition lead to transfer but not retention?

Hypothesis 3:

Regarding the retention of knowledge, when tested a week after learning, the participants learning word pairs in the easy L2-to-L1 direction will produce more correct responses in the same translation direction than those learning the word pairs in the difficult L1-to-L2 direction. Regarding the transferability of knowledge, when tested a week after learning, the participants who learn word pairs in the difficult L1-to-L2 direction will produce more correct responses in the opposite direction than those who learn the word pairs in the easy L2-to-L1 direction.

4.2 Participants

A total of 54 Japanese-speaking undergraduates at Aoyama Gakuin University participated in the present experiment. Two students did not return in the course of the experiment sessions, so the final number of participants was 52. Of those who completed the experiment sessions, 14 students were majoring in French literature (27%) and 38 in English literature (73%); 10 were male (19%), and 42 were female (81%), aged from 18 to 25. In recruiting participants, the researcher visited classes with the permission of the teachers in charge of those classes and distributed a flyer for the experiment (see Appendix A). The researcher then arranged the schedule for the experiments for each of those who had sent an email to the researcher of their own free will. The participants received 1,000 JPY per session on weekdays or 1,500 JPY on weekends or during vacation periods.

The planned number of participants (i.e., 52 samples) was established on the basis of the statistical indices provided by Cohen (1992). According to Cohen, each group should include at least 26 samples for a reliable difference of two independent means, given that the alpha level (i.e., criterion for significance) is set at .05, the statistical power is .80, and the estimated effect size is medium (ES = .50).

4.3 Experimental Software

4.3.1 Selection of Packages

At the planning stage of the study, several stimulus presentation packages that could be suitable for the present study were available. In the following sections, we will briefly describe the distinctive features of each package considered use in the present study.

4.3.1.1 SuperLab

SuperLab runs both on Windows and Mac-OSX, and the program presents texts, pictures, and sounds as experimental stimuli. Users can also employ audio files as stimuli if DirectX is additionally installed. This software provides graphical user interface (GUI) so we did not need to have knowledge of program coding to create a set of experiments.

Figure 4.3.1.1

A Screenshot of the SuperLab Main Window.

When starting SuperLab, a main window with three sub-panes appears. All events needed for an experiment are created on the main window (Figure 4.3.1.1). The three sub-windows, entitled Blocks, Trials, and Events constitute the hierarchical levels of a presentation set. Block is the highest level of an experiment and should include at least one trial. Trial is the second-highest level and should include at least one event.

Finally, Event, the minimal level and an actual stimulus (e.g., texts, sounds, pictures, or movies), is implemented at this level. Details such as the duration of a stimulus, screen locations, or feedback settings can be specified on this option menu.

After creating components in the three categories, we need to associate them so that SuperLab recognizes which blocks include which trials and which trials contain which events. Basically, components listed in Blocks, Trials, and Events are executed from top to bottom, but a randomized presentation is also possible by the setting at the parent level (e.g., events can be randomized by setting the option of the superordinate trial to which the events belong).

SuperLab is a product of Cedrus Corp., USA. The latest version is SuperLab 5 (as of January 2020). Customers can purchase it on CD or through the Internet (download version). The CD version costs 98,280 JPY (tax-included) for a single license, and the download version costs 92,880 JPY (tax-included) at Tegara Corp., a Japanese vendor, as of January 2017. Cedrus Corp also offers a student price for a single license at 199 USD, but it is only valid for two years.

A 20 minute YouTube video “A First Experiment in Superlab” by Martin Hilpert (available at www.youtube.com/watch?v=HiYidKQpRQ0, as of October 2019) demonstrates how to create a simple word-recognition task from scratch. This is informative, especially for beginners.

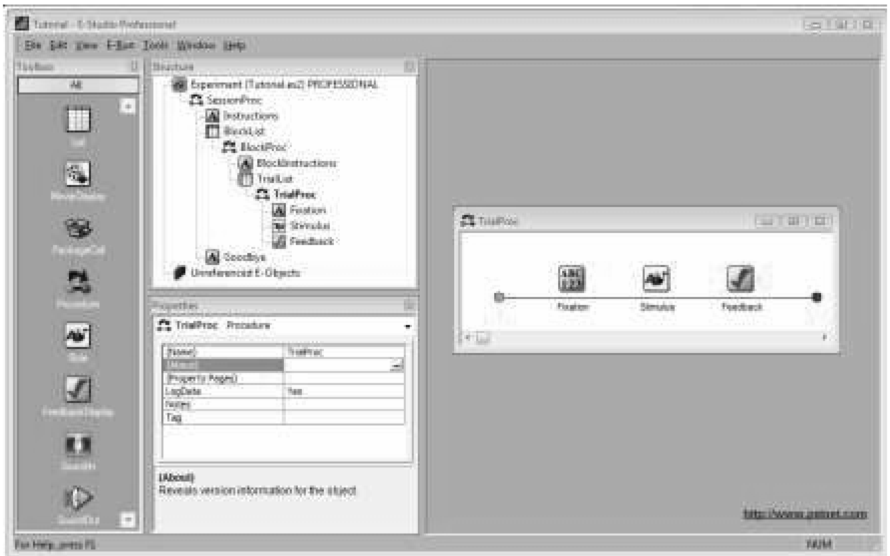
4.3.1.2 E-Prime

E-Prime is a software package used to conduct stimulus presentation experiments, including an experiment-designing tool, a coding tool, and a data-merging tool. E-Prime runs on both Windows and Mac-OSX computers. Unlike SuperLab, this program can play movies without any additional installation. The latest version supports touch-screen devices.

E-Prime consists of six component software programs: E-Studio, E-Basic, E-Run, E-Merge, E-DataAid, and E-Recovery. Most of an experiment's design occurs on E-Studio, a GUI-based experiment editor. Similar to SuperLab, an experiment consists of three hierarchical levels: Brocks, Trials, and Events. An actual hierarchy of components is shown in the tree in the window on the upper left. We can create an experiment by drag-and-dropping the ready-made controls (e.g., TextDisplay, ImageDisplay, SoundOut) shown on the left onto the timeline in the Procedure window (Figure 4.3.1.2). Typical experiments can be created only on E-Studio, but by coding lines in E-Basic, a script language for E-Prime, more detailed control is possible.

Figure 4.3.1.2

A Screenshot of the E-Prime Main Window.



E-Run is a compiler. It converts an experiment designed on E-Studio to the E-Basic language. E-Merge is a data-merging software. On E-Merge, we can select data files taken in some experiment sessions and merge them into a single file to facilitate further analysis. E-DataAid is a data viewer and editor similar to a spreadsheet program. Finally, E-Recovery is a data salvage program that may be useful when a forced termination occurs in the experiment.

E-prime is a product of Psychology Software Tools Inc., USA. The latest version is E-Prime3, launched in December 2016. E-Prime3 comes with USB memory. On the developer's site, it costs 995 USD for a single license. No academic prices are offered. Customers can buy E-Prime3 from two Japanese vendors, Tegara Corp. and IBS JAPAN Co. Ltd. Prices are available upon request.

Two instruction videos provided by the developer are available online at IBS JAPAN (available at www.ibsjapan.co.jp/tech/details/eprime-solution/movie.html, as of October 2019). The video entitled "E-Prime 2.0 Webinar," has Japanese subtitles. Another "E-Prime 3.0 Webinar" is only in English.

4.3.1.3 PsychoPy

PsychoPy is free software for psychology experiments for Windows, Mac, and Linux. The program can handle various stimuli, such as texts, sounds, pictures, and movies. A unique feature of this software is that it can display polygons or dots through the built-in settings so researchers do not need to prepare bitmap files.

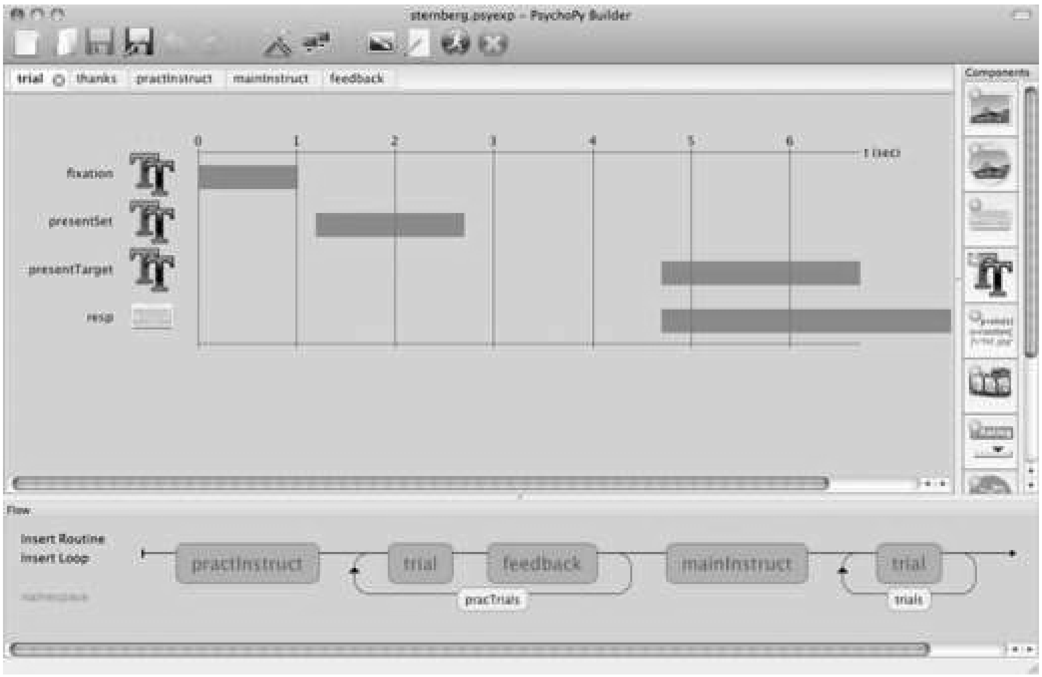
PsychoPy consists of two functional tools: Builder and Coder. Builder is a GUI-based experiment-designing tool. We can create an experiment by drag-and-dropping the ready-made functions onto the timeline of a trial as on E-Prime. There are three parts in the Builder window: the Trial, Flow, and Component panes.

The Trial pane, the largest one, shows actual events to present in a trial in the manner of a timeline. A timing to present the trial is specified in the Flow pane shown at the bottom of the window (Figure 4.3.1.3). In PsychoPy, a multiple-trial scenario is defined by setting a loop in the Flow pane. For example, we can add a loop to a trial containing a list of words and nonwords and randomize them in the actual presentation by setting the options of the loop.

A disadvantage of PsychoPy is that we need to code lines, by using Coder, for relatively typical controls. For example, coding in Python is required to separate the ongoing process into a correct event and an incorrect event according to a certain response.

Figure 4.3.1.3

A Screenshot of the PsychoPy Main Window.



An advantage is that the program is free. In addition, several instruction articles in Japanese are available on the Internet, whereas finding such documents on SuperLab or E-Prime is typically difficult.

PsychoPy is an open-source software program, so many developers revise and release the version. Users can download an installation set of PsychoPy at www.psychopy.org (as of October 2019). At the time of writing, the latest version was not stable. Generally, the second or the third newest version has been well verified by many users, making them safe.

Some instructional sites are available in Japanese, such as one that provided by the Perception & Cognition Laboratory at Kwansei Gakuin University (available at ogwlab.org/?page_id=460, as of October 2019), or one provided by Ehime University (available at www.s12600.net/psy/python/ppb/, as of October 2019).

4.3.1.4 OpenSesame

OpenSesame is a free program for the development of behavioral experiments for psychology, employing a point-and-click graphical interface. As we searched on the Internet, we found that this program had not gained much popularity yet. Similar to the programs reviewed so far, OpenSesami can handle most types of media. The icons listed in the toolbar on the left show ready-made functions. Users create a variety of experiments by placing these functions on the Overview pane next to the vertical toolbar. The Overview pane displays an overall flow of the experiment in the style of a tree (Figure 4.3.1.4).

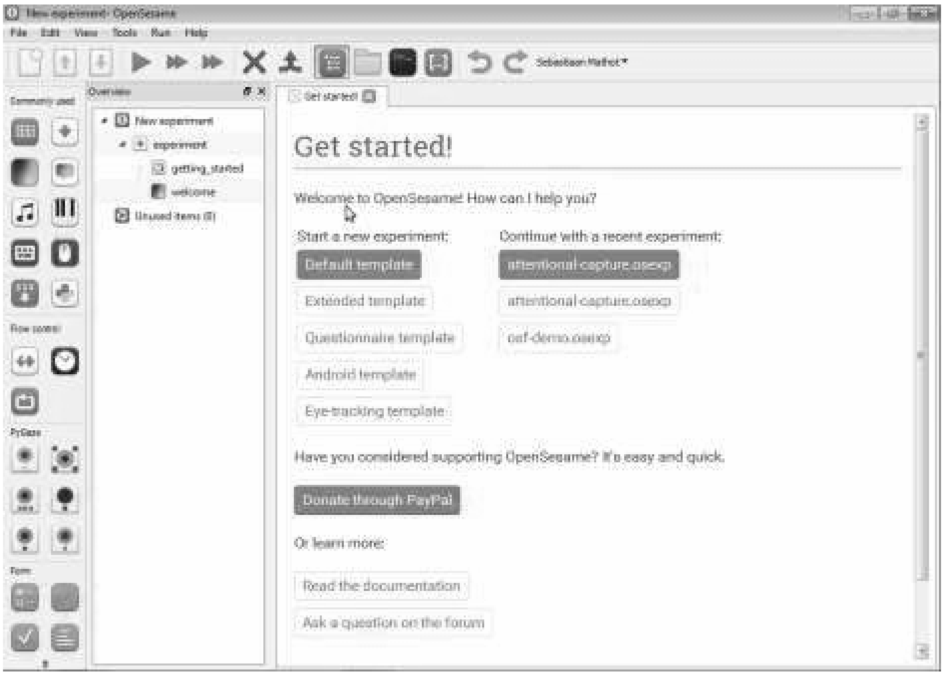
A unique function of this software is that it has a built-in sketchpad. Users can draw pictures on the sketchpad and set them up as visual stimuli. Another useful feature is that users can choose some templates for typical experiments when starting the program.

OpenSesami is available for download at osdoc.cogsci.nl/ (as of October 2019).

The latest stable version is OpenSesami 3.2.8 as of the time of writing. A runtime file for Android (smartphone) is also available. Moreover, the OpenSesami homepage provides a beginner tutorial in creating a gaze-cueing experiment, and test-takers respond which side (right or left) the picture face on the screen gazes toward (available at osdoc.cogsci.nl/3.2/tutorials/beginner, as of October 2019). A step-by-step video tutorial on how to create a visual search experiment is also available on this site. In this experiment, participants indicate whether the target object is present or not among various objects shown at a time by pressing keys.

Figure 4.3.1.4

A Screenshot of the OpenSesami Main Window.



4.3.2 The Experimental Package to Be Used

As a result of a comprehensive comparison of various dimensions, such as their reliability, usability, familiarity, and cost, we chose SuperLab. Although it is a licensed software, and thus involves a certain financial cost, SuperLab is one of the most widely used packages that provides a platform in which to create reliable experiments. Another reason is that SuperLab offers the simplest way to compose experiment tasks among the four candidate packages. At the time of purchasing a SuperLab license, SuperLab5 was the latest stable version. SuperLab6 was also available as a beta version that license holders could use for free. The tasks of the present study were mainly constructed on the beta version of SuperLab6.

4.4 Materials

Schneider *et al.* (2002) used 25 English–French word pairs for native English speakers who had little knowledge of French. Those words pairs basically consisted of five different categories, although their results showed that the facilitative effects on the categorical interference were largely submerged by significant effects of the translation direction. Therefore, in designing the present translation pairs in a conceptually similar way to those of Schneider *et al.*'s method, we paid attention to the following points: (a) Since Japanese speakers were going to participate, one of the languages used in the translation pairs should be Japanese; (b) the other language of the pairs should be one that is extremely unfamiliar to average Japanese people. (c) Because the original study used French as a second language with English speakers, and the words were considered to be pronounceable to some degree, the second language in the present study should also be one that Japanese students find somewhat pronounceable. Theoretically, pronounceability is an important issue in word encoding as well since the phonological loop mechanism is responsible for the short-term holding of verbal inputs (Gathercole & Baddeley, 1993). Finally, (d) since keyboard typing is necessary for responses on the planned learning task, the second language should be one that could easily be typed on a standard keyboard.

Taking all these points into consideration, we chose the Indonesian language as the target second language of the translation pairs. Indonesian is generally not taught in the standard curriculum in Japan. Furthermore, this language consists only of familiar alphabetic characters (A-Z), which most Japanese students who have learned English in school can pronounce, but their meanings are not likely to be guessed. Furthermore, this characteristic permits participants to type the words on a standard keyboard.

Target translation pairs were chosen from simple daily words from five different categories that everyone knows. Table 4.4 shows the target items in the present study, in parallel with those used in Schneider *et al.*'s (2002) Experiment 1.

Table 4.4

Word Pairs Used in the Present Study and Schneider et al. (2002).

The present study Indonesian–Japanese (Eng.)	Schneider *et al.* (2002) French–English
Geography	Body parts
gunung – 山 (mountain)	dos – back
laut – 海 (sea)	bouche – mouth
hutan – 森 (forest)	figure – face
pulau – 島 (island)	doigt – finger
sungai – 川 (river)	yeux – eyes
Interior objects	Vehicles
meja – 机 (table)	voiture – car
pintu – ドア (door)	velo - bicycle
jendela – 窓 (window)	avion – airplane
kursi – いす (chair)	camion – truck
kamar – 部屋 (room)	bateau – boat
Animals	Tableware
burung – 鳥 (bird)	fourchette – fork

anjing – 犬 (dog)	cuiller – spoon
kuda – 馬 (horse)	couteau – knife
kucing – 猫 (cat)	assiette - plate
ikan – 魚 (fish)	verre – glass
Colors	Food
merah – 赤 (red)	pain – bread
hijau – 緑 (green)	beurre – butter
biru – 青 (blue)	lait – milk
hitam – 黒 (black)	oeufs – eggs
putih – 白 (white)	jambon – ham
People	Clothes
guru – 教師 (teacher)	manteaux – coat
siswa – 学生 (student)	chemise – shirt
istri – 妻 (wife)	chaussures – shoes
supir – 運転手 (driver)	cravat – tie
raja – 王 (king)	jupe – skirt

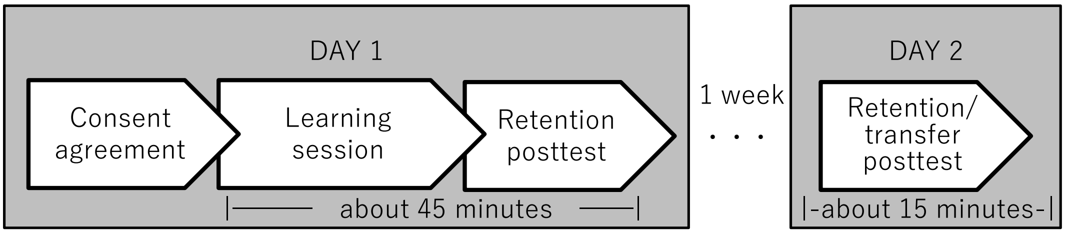
4.5 Procedure

In a quiet and well air-conditioned classroom, the present experiment was conducted by strictly following the method of the original researcher, Schneider *et al.* (2002), involving two days with a one-week retention interval. Day 1 included signing a consent form (see Appendix A), a learning session, and an immediate retention posttest. On day 2, participants returned and were given either a delayed retention posttest or a transfer posttest. The average time required to complete all the day 1 tasks was approximately 45 minutes. Day 2 tasks required about 15 minutes on average. The timeline of Study 1 is illustrated in Figure 4.5.1.

On day 1, the participants learned the 25 Indonesian–Japanese word pairs and took an immediate posttest to measure retention. They were assigned alternatively to either an easy or a difficult learning condition, which was manipulated in terms of translation directions according to the order of their arrival.

Figure 4.5.1

The Timeline of Study 1.



The task in the easy learning condition was to memorize word pairs presented in an Indonesian–Japanese format (e.g., gunung–山) and then to type the Japanese counterpart word of each pair (e.g., 山) from an Indonesian cue word (e.g., gunung). The difficult learning condition involved the opposite process; the students memorized the same word combinations presented in a Japanese–Indonesian format (e.g., 山–gunung), and then typed the Indonesian counterpart of each pair (e.g., gunung) from a Japanese cue word (e.g., 山).

In both learning conditions, the 25 target word pairs were presented by dividing them into five blocks, with five pairs in each block. In each block, five word pairs that were randomly chosen from each category were displayed for 2 seconds one at a time. After the five pairs were sequentially presented, an input box appeared with a cue stimulus word displayed above it. Participants typed in the corresponding word to this cue in the language designated by the task at hand. There was no time limitation during typing. Each stimulus word was presented in the same order in which the word pairs were presented. Two or more word pairs in the same category were not included in a certain block; nor did the same word pair appear multiple times in different blocks. All the words, including word pairs and stimulus words, were shown in the center of a

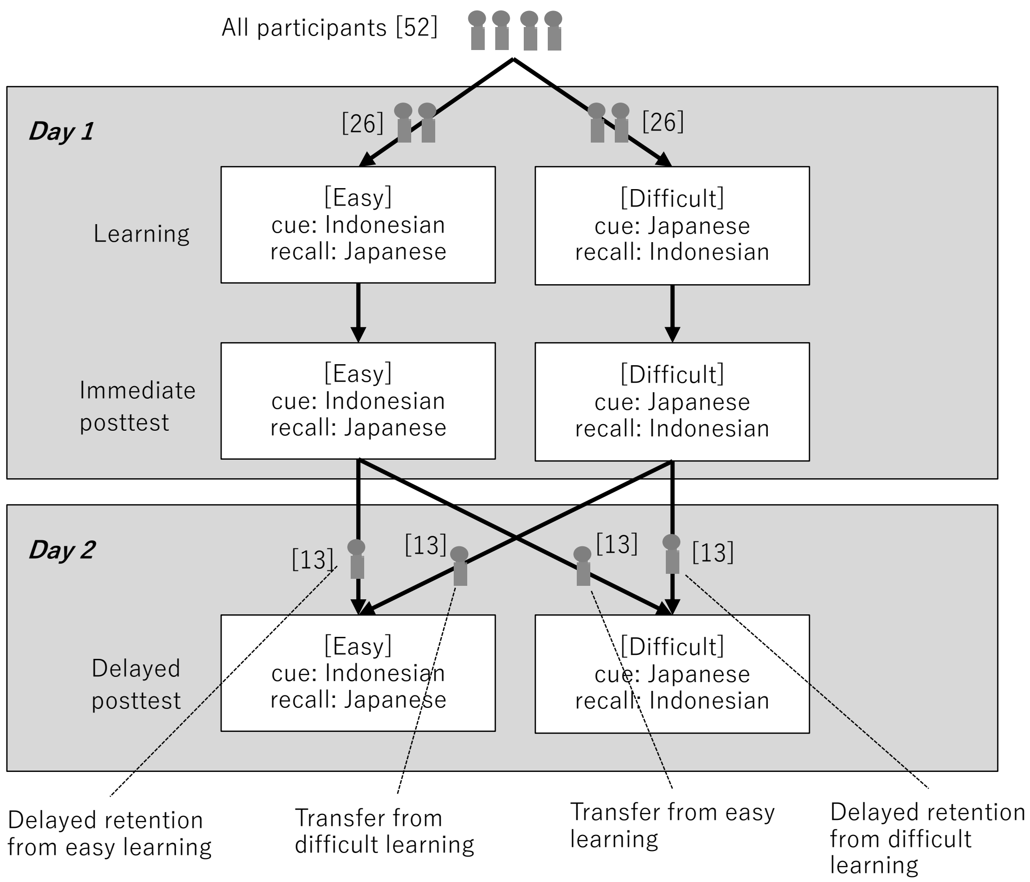
14.1-inch white screen in black 36-point Georgia font for alphabetic characters and in the black 36-point MS Gothic font for Japanese characters. A flowchart of a given learning block is illustrated in Appendix B for both the easy and difficult learning conditions. In both learning conditions, after completing all five blocks in which the participant had been exposed to all target items, they repeated this learning procedure two more times in different randomized presentation orderings. Thus, the participants had opportunities to learn the 25 word pairs three times in total, as Schneider *et al.*'s (2002) participants did.

Immediately after the learning phase, participants took a posttest to measure their memory of the learned word pairs. The format of the posttest was the same as that used for the cued inputs in the learning session. Twenty-five cued input screens for each word pair appeared successively, and participants had to type the corresponding word of each pair in either Japanese or Indonesian depending on the conditions they were assigned to in a self-paced manner. Screens of all test items on the immediate posttest in the easy learning condition (i.e., L2 cue–L1 response) are provided in Appendix C. Moreover, those in the difficult learning condition (i.e., L1 cue–L2 response) appear in Appendix D. For both conditions, each screen for a given item was shown only once in a randomized order. As noted in Appendix C, in a posttest where Japanese responses were required, a response in Hiragana format was also acceptable since it did not make an impact on the meaning.

One week after the first experiment session, on day 2, participants returned for a delayed posttest. This time, half of the participants in each learning condition took a test to measure retention, but the remaining half took a test to assess transfer. The delayed retention test required students to simply repeat the cued typing of the target response words in the same translation direction as they studied them on day 1 but in a new randomized order. The delayed transfer test required students to type in the translation equivalents in the other language from the ones they had typed in on day 1. Therefore, the participants in the easy learning condition on day1 who had typed Japanese words from Indonesian cue words had to type, this time, Indonesian words from Japanese cues (see Appendix D for the test format). Similarly, the participants in the difficult learning condition who had typed Indonesian words from Japanese cues should type, this time, Japanese words from Indonesian cues (see Appendix C for the test format).

Figure 4.5.2

A Flowchart of the Present Experiment.



In this way, the research design allowed us to obtain word recall data during learning and on an immediate posttest, performed at two different levels of difficulty, on day 1. Moreover, four types of data were collected on day 2, which were on (1) the delayed retention associated with an easy learning process, (2) the transferability of knowledge associated with the easy learning process, (3) the delayed retention associated with a difficult learning process, and (4) the transferability of knowledge associated with the difficult learning process. Figure 4.5.2 consists of a detailed flowchart of the experiment with the number of participants performing each task.

4.6 Results

4.6.1 Research Question 1

The first research question asked whether a learning task with lower degree of difficulty (i.e., easy L2 cue–L1 recall) promotes learning more than a difficult learning task (i.e., difficult L1 cue–L2 recall). To answer the question, we started analyses by looking at distributions of collected data. In the learning phase, 52 participants were divided into halves, the easy learning condition (N = 26) and the difficult learning condition (N = 26), and each group studied target word pairs three times.

Table 4.6.1.1 shows descriptive statistics for correct responses of each trial in the easy and difficult learning conditions. As seen in the mean values in the proportions of correct responses, participants in the easy learning condition, who responded in their L1 from L2 cues, made more accurate responses than participants in the difficult learning condition whose translation direction the opposite way. The median values were also higher for the easy learning group than the difficult learning group. These values increased as a function of time for both group.: The second trial was more accurate than the first trial, and the third trial was more accurate than the second trial. The effects of the times of trials were also found in the standard deviation (*SD*) for the easy learning group. The *SD*s converged as participants repeated practice, suggesting that stable intake took place. However, this was not the case for the difficult learning group, as evidenced by the increased *SD*s to the final trial.

Table 4.6.1.1

Descriptive Statistics on Response Accuracy in Learning.

	Easy Condition (L2 cue–L1 recall)			Difficult Condition(L1 cue–L2 recall)		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
Mean	.612	.799	.894	.229	.471	.600
Median	.640	.840	.920	.200	.520	.640
SD	.173	.102	.098	.124	.180	.253
Min	.320	.560	.640	.080	.200	.160
Max	.920	1.000	1.000	.640	.800	1.000
Skewness	−.301	−.558	−1.339	1.926	−.006	−.252
Kurtosis	−.947	−.119	.867	4.687	−1.279	−1.130

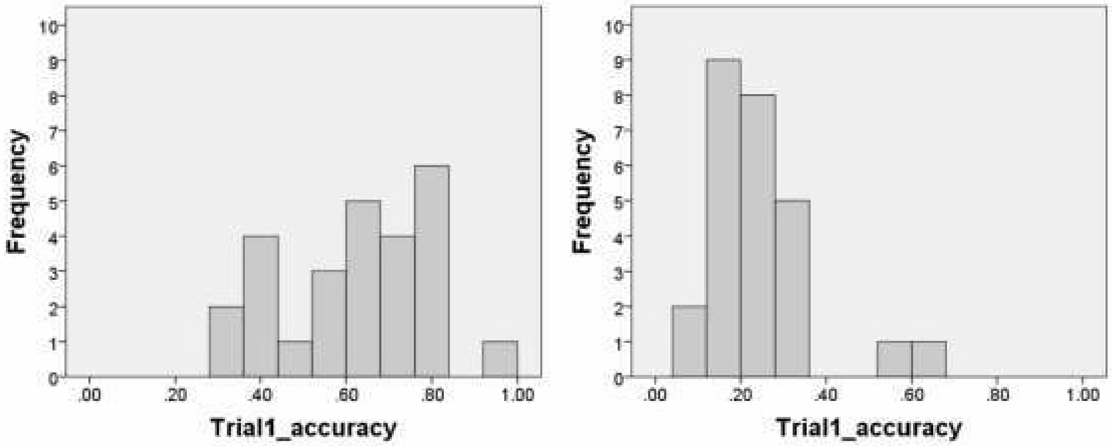
For both learning conditions, we checked the distributions of data of each trial (Figure 4.6.1) and calculated values for normality. Kolmogorov−Smirnov tests indicated that data on Trial 1 in the easy condition maintained a normality (p = .200), whereas trials 2 and 3 were not normally distributed (both p < .001). The tests also showed that, in contrast, data on trials 2 and 3 in the difficult condition maintained a normality (p = .162, p = .051, respectively), but Trial 1 did not (p = .011). Therefore, further analyses of the significance of differences involved non-parametric tests.

Figure 4.6.1

Histograms on Response Accuracy during Learning Trials 1, 2, and 3.

Learning Trial 1

Easy Learning Condition **Difficult Learning Condition**



Learning Trial 2

Easy Learning Condition **Difficult Learning Condition**

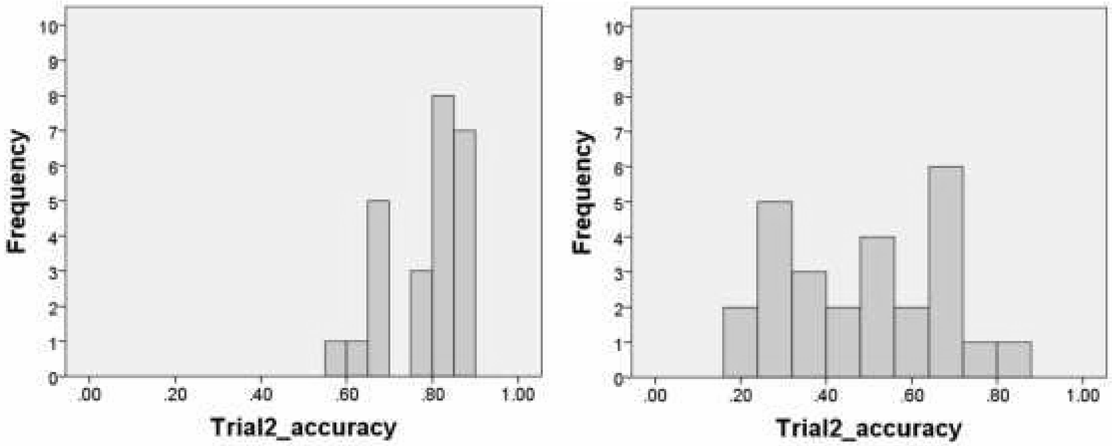
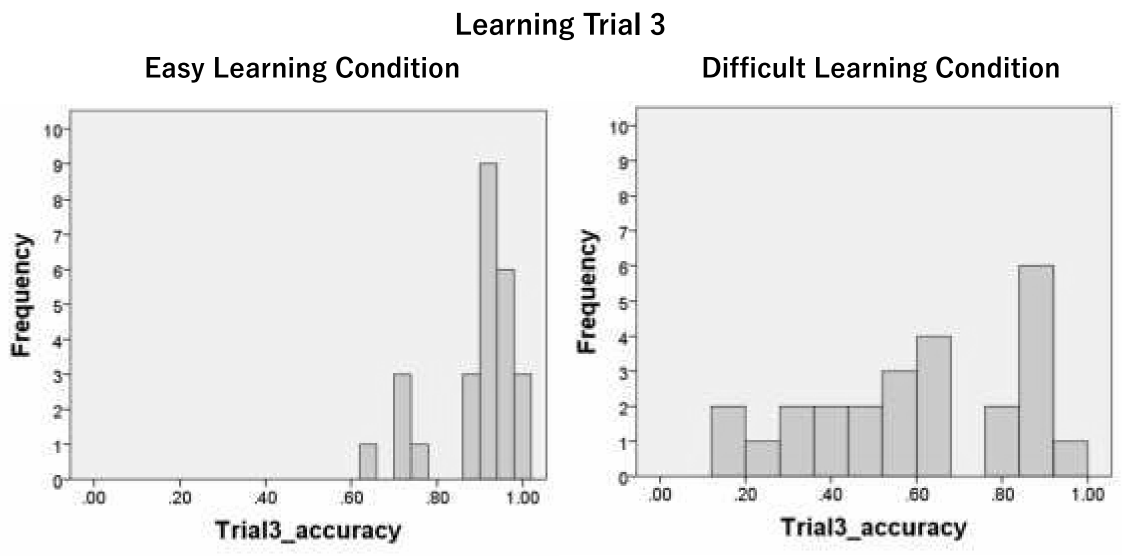


Figure 4.6.1 (Continued)



To check the significance of the gains over trials within the same condition, two non-parametric Friedman tests of differences among the proportions of correct responses for each task were conducted separately, and they rendered a chi-square value of 37.604 for the easy learning condition and 43.820 for the difficult learning condition, which were both significant (both at $p < .001$). Post-hoc comparisons between the accuracies in trials 1 and 2, trials 1 and 3, and trials 2 and 3 were investigated for each condition by performing six Wilcoxon sign-rank tests separately. The results showed that, in the easy learning condition, the response accuracy was significantly higher for Trial 2 than Trial 1, $Z = -4.057$, $p < .001$, and for Trial 3 than Trial 2, $Z = -3.050$, $p = .002$. Likewise, in the difficult learning condition, the accuracy was significantly higher for Trial 2 than Trial 1, -4.379, $p < .001$, and for Trial 3 than Trial 2, $Z = -2.990$, $p = .003$. Not surprisingly though, these results confirmed significant improvements during learning as a function of time for both experimental conditions.

The evaluation for the significance of differences in response accuracy between the easy and difficult learning conditions was conducted on a non-parametric Mann−Whitney test. According to the results, at each of the trials, easy learning yielded a significantly higher proportion of correct responses than difficult learning (U = 22.500, p < .001 for Trial 1; U = 33.000, p < .001 for Trial 2; U = 82.000, p < .001 for Trial 3). To summarize, the first hypothesis ("An easy learning condition will yield more learning gains than a difficult learning condition,") was confirmed on the grounds of the remarkable contrast of correct responses between the easy and difficult learning conditions (Table 4.6.1.2).

Table 4.6.1.2

Mean Proportions of Correct Responses on Each Learning Trial.

	The Number of Trials		
Learning Condition	1	2	3
Easy (L2 cue–L1 recall)	.61	.80	.89
Difficult (L1 cue–L2 recall)	.23	.47	.60

4.6.2 Research Question 2

The second research question concerned the prolonged recall advantages associated with difficult learning. Namely, we asked whether knowledge associated with difficult learning is more frequently recalled, as the sum of retention and transfer, than that associated with easy learning when tested a week after learning. We investigated this question by comparing participants' response accuracies on the 1-week delayed posttest with those on the immediate posttest as a baseline performance for each learning condition, as in the work of Schneider *et al.* (2002).

Table 4.6.2.1 displays descriptive statistics for correct recall for learned words on the immediate posttest and the delayed posttest after a 1-week retention interval. Not surprisingly, in both conditions, the mean immediate recall performances (.723 in the easy condition; .482 in the difficult condition) were better than the mean delayed recall performances (.172 in the easy condition; .184 in the difficult condition), suggesting a certain amount of forgetting as a function of time passing. Comparing the learning conditions, participants in the easy learning condition recalled more words immediately after the learning process than participants in the difficult learning condition, indicating inherited short-term advantages of gains from easy learning. However, when tested a week after learning, the participants in the difficult learning condition outperformed those in the easy learning condition, as reported by Schneider *et al.* (2002).

Table 4.6.2.1

Descriptive Statistics on Response Accuracy on the Immediate Posttest and the Delayed Posttest.

	Easy Learning Condition (L2 cue–L1 recall)		Difficult Leaning Condition (L1 cue–L2 recall)	
	Immediate	Delayed	Immediate	Delayed
Mean	.723	.172	.482	.184
Median	.780	.080	.440	.160
SD	.198	.184	.235	.163
Min	.200	.000	.120	.000
Max	.960	.520	.880	.440
Skewness	–1.124	.715	.252	.244
Kurtosis	1.063	–1.015	–.772	–1.596

To calculate the degrees of significance of these differences, we checked the distribution of data on the immediate posttest and the delayed posttest for both learning conditions. We also performed a Kolmogorov–Smirnov test for normality four times (2 difficulty levels x 2 times of tests). The results showed that the immediate posttests maintained a normality for both the easy learning condition (p = .130) and the difficult learning condition (p = .200), while the delayed posttests were, for both learning conditions, not normally distributed (Figure 4.6.2.1). Therefore, we used a parametric test and a non-parametric test in accordance with the types of distribution, as well as a parametric test used only when both data maintained a normality, to determine the significance of differences.

Figure 4.6.2.1

Histograms on Recall Accuracy on the Immediate Posttest and the Delayed Posttest.

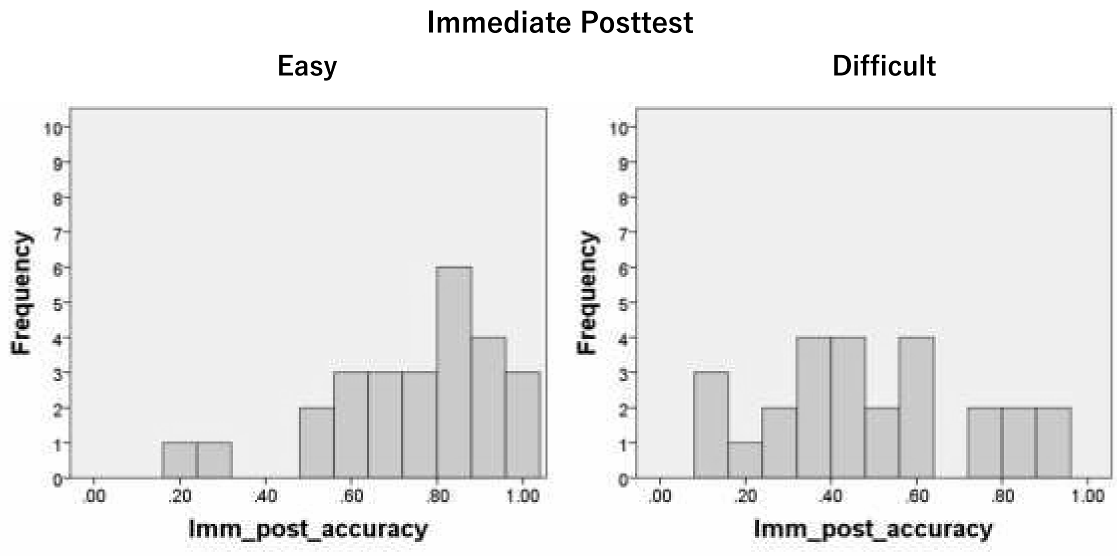
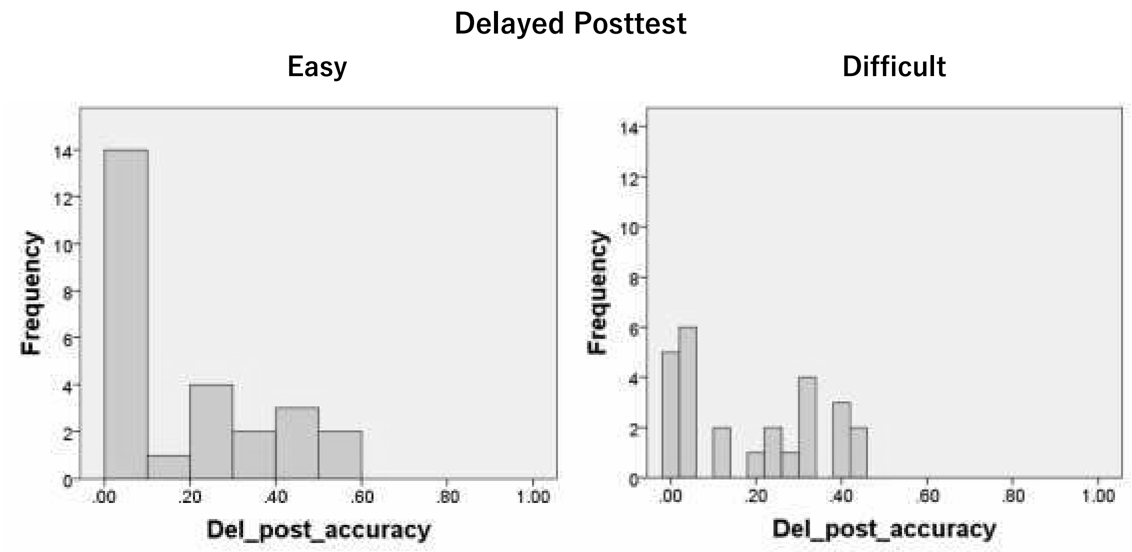


Figure 4.6.2.1 (Continued)

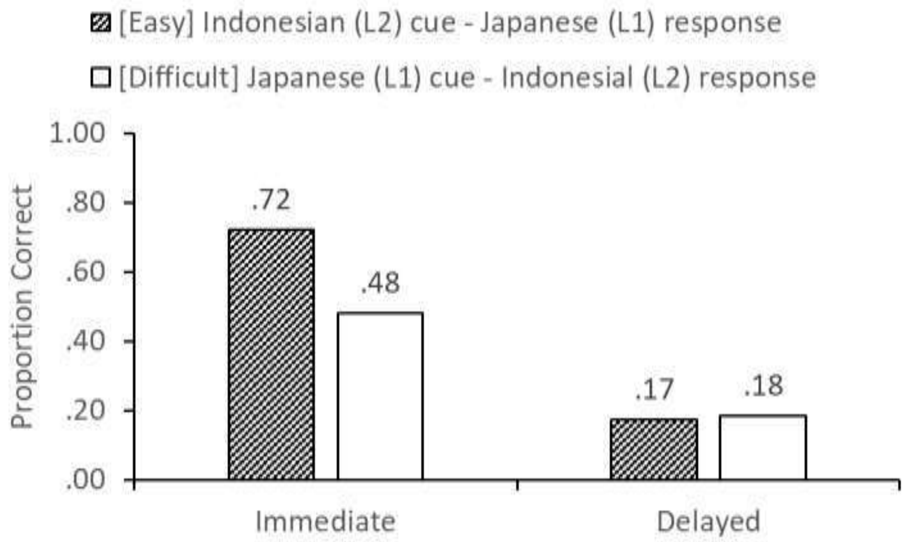


For the mean difference of performance on the immediate posttest, a parametric t-test showed that the participants in the easy learning condition recalled significantly more words (72 %) than those in the difficult learning condition (48 %), $t(50) = 4.01$, $p < .001$. However, a non-parametric Mann–Whitney test for differences of recall probabilities between these two learning groups on the delayed posttest did not reach significance (17% in the easy learning condition; 18% in the difficult learning condition). Furthermore, non-parametric Wilcoxon signed–rank tests confirmed that the decay of knowledge after a week-long retention interval was significant for the easy learning condition, $Z = –4.461$, $p < .001$, as well as for the difficult learning condition, $Z = –4.206$, $p < .001$.

The results indicated a long-term recall advantage associated with difficult learning, characterized in an inversed proportion of recall rates over a certain retention interval (Figure 4.6.2.2), as in the work of Schneider *et al.* (2002). Although the 1-week recall rates between the easy and difficult learning conditions were not statistically significant, the present research replicated one of the main findings in Schneider *et al.* (2002) with a similar pattern to the one they observed.

Figure 4.6.2.2

Proportions of Correct Recalls on the Immediate Posttest and the Delayed Posttest.



4.6.3 Research Question 3

The third research question asked whether easy learning and difficult learning would each show distinctive effects on subsequent recall. As mentioned in Section 4.1, Schneider *et al.*'s (2002) findings suggested that easy learning supported the retention of knowledge rather than its transferability, whereas difficult learning promoted the transferability of knowledge rather than its retention. Retention here means doing the same task completed in an initial learning process later again, and transfer refers to doing essentially the same, but a slightly different, task from the initial learning process at a later time.

To answer the question, we categorized the delayed posttest data into four categories, based on the possible combinations between the levels of learning difficulty and task types, and then calculated the recall probabilities for each case (Table 4.6.3). For Schneider *et al.*'s (2002) figures in these distinctions, see those provided in Table 4.1.1.

Table 4.6.3

Proportions of Correct Responses on the Immediate Posttests and the Delayed Posttests Based on Task Types.

Learning	Immediate	Delayed		Task Type
Indonesian(L2) cue	.72	Indonesian(L2) cue	.32	Easy learning & retention
[Easy]		Japanese(L1) cue	.02	Easy learning & transfer
Japanese(L1) cue	.48	Indonesian(L2) cue	.28	Difficult learning & transfer
[Difficult]		Japanese(L1) cue	.09	Difficult learning & retention

The points to look at are the four values in the Delayed column, each referring to the recall rate based on levels of learning difficulty and the posttest types. First, simply ranking these four figures, the best is the retention rate associated with easy learning (32%), the second is the transfer rate associated with difficult learning (28%), the third is the retention rate associated with difficult learning (9%), and the transfer rate associated with easy learning (2%) is the last. This rank order is exactly the same as that derived from Schneider *et al.*'s (2002) Table 5 (cf. Table 4.1.1). When initial learning was easy (translating from L2 to L1), participants doing the same task again a week after learning (translating again in the same direction) outperformed participants working on a slightly different task (translating the same word sets from L1 to L2). In contrast, when initial learning was difficult (translating from L1 to L2), participants doing a slightly different task (translating the same word sets from L2 to L1) performed better than participants doing the same task (translating L1 to L2) again. This pattern, by itself, shows evidence for this hypothesis (Hypothesis 3) that easy learning promotes retention, and difficult learning enhances transfer.

We further checked the statistical significance of these differences by performing non-parametric Mann–Whitney tests across the four types of delayed posttests. The results indicated that easy learning more significantly promoted retention compared to transfer, $U = 2.000$, $p < .001$, and difficult learning significantly enhanced transfer compared to retention, $U = 27.500$, $p < .001$. Moreover, in a comparison between the two learning levels, the retention rate was significantly better for easy learning than difficult learning, $U = 19.000$, $p <$ 001, and the transferability was significantly higher for difficult learning than easy learning, $U = 10.500$, $p < .001$. These results supported the present hypothesis that different levels of difficulty in learning each promote different aspects in subsequent recall.

4.7 Discussion

In Study 1, we attempted to replicate Schneider *et al.*'s (2002) Experiment 1 on long-term recall advantages derived from difficult learning by manipulating translation directions in paired-associate foreign-language vocabulary learning in participants' L1 and L2. Employing the same methodology but different languages and subject populations from the original study, the present study obtained results largely consistent with what Schneider *et al.* (2002) found on the basis of the contextual interference effect (Battig, 1972; Schneider *et al.*, 1995). The present findings reiterated the idea that a method that yields the fastest way to learn something is not necessarily the best way to maintain learned knowledge in the long run. In this replication study, we established three research questions that may correspond to key points of the findings of Schneider *et al.* (2002). For each research question, we will examine the present results in contrast with Schneider *et al.*'s results.

4.7.1 Rates of Learning in Easy and Difficult Learning Conditions

The present study clearly replicated Schneider *et al.*'s findings that, in paired-associate learning of foreign words, learning is faster and more efficient when people learn new foreign words by responding in their L1 from L2 cues than by responding in their L2 from L1 cues. The proportions of correct responses in each trial and condition of the present study are comparable to the corresponding figures of Schneider *et al.* (2002), even though a different set of languages and learner populations were employed (compare Table 4.7.1.1 with Table 4.7.1.2). Overall, this is an expected result when the mean correct responses from the final recall trial are considered. It is natural to think that participants who respond in their familiar L1 from L2 cues would outperform those who respond in an unfamiliar L2 from L1 cues in terms of accuracy.

Here, two things must be pointed out about the results. First, it appears that overall learning performance can be determined by the performance at the very first recall attempt. As seen in Table 4.7.1.1, the main difference between the easy task and the difficult task is the mean accuracy at the first recall trial. Learning gains from the first trial to the second trial, and the second to the third trial are relatively similar in both learning conditions. Therefore, the very first performance and the later performances can be separated in terms of quality. In other words, the first performance is the result of capitalizing on prior knowledge, and the later performances stem from pure learning.

Table 4.7.1.1

Mean Proportions of Correct Responses during Learning in the Present Study (Reshown from Table 4.6.1.2).

Participants' L1: Japanese	Trials		
Learning Condition (Translation Direction)	1	2	3
Easy (Indonesian [L2] to Japanese [L1])	.61	.80	.89
Difficult (Japanese [L1] to Indonesian [L2])	.23	.47	.60

Table 4.7.1.2

Mean Proportions of Correct Responses during Learning in Schneider et al. (2002) (based on Table 2, p. 426).

Participants' L1: English	Trials		
Learning Condition (Translation Direction)	1	2	3
Easy (French [L2] to English [L1])	.48	.69	.80
Difficult (English [L1] to French [L2])	.24	.44	.60

Second, the difficult learning condition actually does not appear to be inefficient. The statistics suggested that the mean recall accuracy for each trial was significantly different between the easy learning and the difficult learning condition. However, when focusing on learning gains between trials, the difficult learning condition was slightly better than the easy learning condition. The final gain from the first attempt was .28 for the easy learning condition and .37 for the difficult learning condition (Table 4.7.1.1). This was also the case for Schneider *et al.*'s results (.32 on easy condition vs. .36 on difficult condition; Table 4.7.1.2). Therefore, to put it in extreme terms, learning in a difficult way is more efficient than in an easy way, given that an initial advantage, such as those derived from prior knowledge, is neutralized.

4.7.2 Rates of Retention in Easy and Difficult Learning Conditions

The present study obtained a trend similar to that in the work of Schneider *et al.* (2002) for retention advantages from a difficult learning task (compare Figure 4.7.2.1 with Figure 4.7.2.2). The present study found that participants given Indonesian cues during learning had a significantly higher proportion of correct responses on the immediate posttest (i.e., the easy learning condition) than did participants given Japanese cues (i.e., the difficult learning condition). The opposite pattern was found, although it is not statistically significant, on the delayed posttest after a 1-week retention interval, in which participants given Japanese cues during learning showed a slightly higher proportion of correct responses than did participants given Indonesian cues during learning. In contrast to Schneider *et al.*'s (20002) results, the difference of performance between the two experimental groups on the delayed posttest was small, but the present results indicated, as Schneider *et al.* (2002) found, that knowledge learned on a difficult learning task was more durable than that learned on an easy learning task.

Figure 4.7.2.1

Proportions of Correct Recall on the Immediate Posttest and the Delayed Posttest in the Present Study (Reshown from Figure 4.6.2.1).

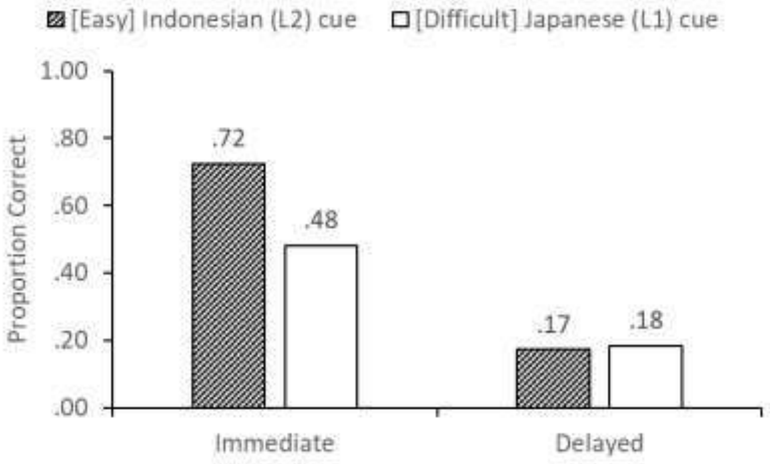
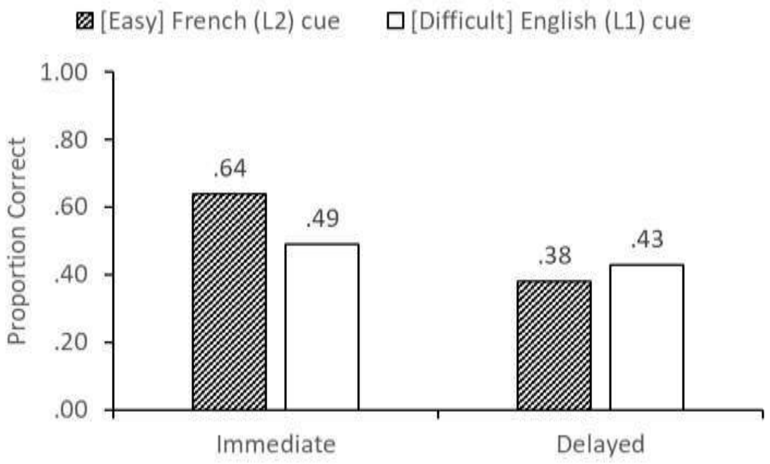


Figure 4.7.2.2

Proportions of Correct Recall on Immediate Posttests and Delayed Posttests in Schneider et al. (2002) (based on Table 5, p. 429) (Reshown from Figure 4.1.1).



Schneider *et al.* (2002) and other researchers might attribute this phenomenon to memory traces generated in cognitively shallow processing decay more easily than memory obtained by a cognitively demanding learning process, as the depth-of-processing theory predicts. However, from the viewpoint of the learning gains we looked at in the section above, another explanation is also possible. That is, participants on the difficult L1 cued learning might essentially have learned more than participants on the easy L2 cued learning in this particular condition. Such a situation can be expected because the overall learning gain was greater for the difficult learning condition than that of the easy learning condition. The results might seem impressive, especially when presented in a bar chart with an inverted proportion of correct responses before and after the retention interval, but they can also be simply understood as indicating that knowledge depends on the degree of learning.

4.7.3 Distinctive Effects of Easy Learning and Difficult Learning on Retention and Transfer

The final part of the present replication scrutinized what Schneider *et al.* (2002) did not clearly mention in their study is, on a breakdown of the abovementioned bar chart. In both Schneider *et al.*'s (2002) study and the present replication, half of the participants learned in the easy condition, and the remaining half did so in the difficult condition. Their memories were then tested immediately after learning in the same translation direction as in the learning process. A week later, their memories were tested again with half of the participants from each condition in the same translation direction and the other half in the opposite translation direction. Therefore, each of the four bars on the bar chart represents a sum of retention and transfer. Both the work of Schneider *et al.* (2002) and the present study employed "retention" for a recall probability measured in the same translation direction as in learning, and "transfer" to a recall probability measured in the opposite translation direction compared to the learning process. Surprisingly, the present study replicated Schneider *et al.*'s (2002) results almost perfectly in terms of rank accuracy on the breakdown performances on the delayed posttest, as summarized in Table 4.7.3.

Specifically, in the present study, the mean delayed retention rate was greater than the mean delayed transfer rate for participants who learned the word pairs in the easy learning condition, whereas participants in the difficult learning condition showed an opposite pattern on the delayed posttest. This is also true for Schneider *et al.*'s results. In both studies, the mean delayed retention rate associated with easy learning was the best, followed by the mean delayed transfer rate associated with difficult learning, the mean delayed retention rate associated with difficult learning, and finally, the mean delayed transfer rate associated with easy learning. This finding is important in two ways. First, it allows us to point out a theoretical discrepancy in what Schneider *et al.* (2002) argued. Second, it potentially points us toward a new assumption regarding retention and transfer in relation to task difficulty.

Table 4.7.3

Mean Proportions of Correct Response on Immediate Posttests and Delayed Posttests of the Present Study and Schneider et al. (2002) (Reshown from Table 4.6.3 and Table 4.1.1).

	Present Study			Schneider *et al.* (2002) Exp 1		
Learning	Immediate	Delayed		Immediate	Delayed	
Easy	.72	Retention	.32[1]	.64	Retention	.63[1]
		Transfer	.02[4]		Transfer	.13[4]
Difficult	.48	Retention	.09[3]	.49	Retention	.24[3]
		Transfer	.28[2]		Transfer	.62[2]

Note: The figures in brackets indicate a rank in recall accuracy on the delayed posttest.

In their paper, Schneider *et al.* (2002) wrote, "We found that the more difficult task yielded inferior learning and immediate retention but not inferior . . . delayed retention and transfer" (p. 438), adding, "It appears that retention and transfer suffer less when conditions of training are difficult, despite the fact that the learning process is slowed down" (p. 439). These remarks are, precisely, not appropriate. In their study, a rate of retention was reduced from the immediate posttest to the delayed posttest in the difficult learning condition (from .49 to .24), whereas little loss occurred for the easy learning condition (from .64 to .63). This was essentially replicated in the present study, suggesting that easy learning is better for delayed retention than difficult learning. For the delayed transfer performance, difficult learning outperformed easy learning in both studies. Consequently, these results might bring a new hypothetical perspective in which easy learning tends to enhance subsequent retention, whereas difficult learning promotes the transferability of learned knowledge more than easy learning does. In short, manipulations of task difficulty during learning might contribute to the subsequent retention and transfer in different ways.

The present study largely replicated Schneider *et al.*'s (2002) work, but we would also like to mention some differences. While both studies share similar patterns in their results, each recall probability shown in Table 4.7.3 differs considerably in terms of their absolute values. For example, compared to Schneider *et al.*'s (2002) results, recall probability on delayed retention in the present study's easy learning condition dropped sharply from that at the immediate posttest time. For example, compared to Schneider *et al.*'s (2002) results, recall probability on delayed retention in the easy learning condition fell sharply from that of the immediate posttest (from .72 to .32 in the present study; from .64 to .63 in Schneider *et al.*), even though our participants performed better than their participants on the immediate posttests (.72 vs. .64). A similar trend was also seen in the delayed retention performance in the difficult learning condition (from .48 to .09 in the present study; from .49 to .24 in Schneider *et al.*). The most extreme case was found in relation to the delayed transfer performances of the two studies. In the present study's results, the delayed transfer performance was not better than that measured during the immediate posttest a week earlier (from .72 to .02 in the easy condition; from .48 to .28 in the difficult condition). In contrast, in Schneider *et al.*'s (2002) findings, the delayed transfer performance in their difficult learning condition far surpassed the mean recall probability measured during the immediate posttest held a week earlier (from .49 to .62). This finding might entail some complications because those figures literally indicate that their participants suddenly knew what they did not know a week before. Moreover, now we have come to know that this controversial figure largely contributed to the superior delayed recall performance that Schneider *et al.* (2002) observed.

An explanation for this variability could be made in terms of the differences in the different languages used. In Schneider *et al.*'s (2002) study, French–English word pairs were employed for English-speaking participants; we used Indonesian–Japanese word pairs for native speakers of Japanese. Unlike Japanese and Indonesian, English shares many words with French, since tens of thousands of words were taken from French in the Middle English period (Baugh & Cable, 2013). Ethnologue, an online natural language database, suggests that 27% of contemporary English words are French cognates (retrieved January 30, 2020). One of the target items used in Schneider *et al.* (2002), "*avion*–airplane" is one such example. "Avi" in the French stimulus word "*avion*" is a prefix pertaining to birds or flying things in English. In addition, for the item "*lait*–milk," a participant with a keen sense of words would be able to guess the translation equivalents easily since "cafe au lait'" is an everyday word in English. In studies on psycholinguistic bilingual lexicons, Taylor (1976) found that French–English bilinguals produced more translation equivalents for similar words between the two languages than dissimilar words. Therefore, it is suggested that the effects of contextual interference manipulated in the translation directions vary depending on typological relationships between the two languages and the degree of participants' familiarity with those languages. This idea is consistent with Robinson's (2001a, 2001b, 2003) view that learning effects from language-learning tasks may vary according to interactions between task complexity factors and task difficulty factors. In the present case, the participants in the Schneider *et al.* (2002) study could basically be considered to be "+ proficiency" in terms of Robinson's task difficulty criteria as well as "+ prior knowledge" in task complexity due to the typological similarities of English and French, whereas the participants in the current study learned unfamiliar Indonesian words from scratch. In doing essentially the same tasks in the two studies, these conditional differences might affect subsequent recall performance, boosting retention and transfer in one condition and suppressing them in the other.

Chapter V

THE PRESENT STUDY—STUDY 2

5.0 Introduction

Study 1 confirmed Schneider *et al.*'s (2002) argument that foreign words learned in a difficult learning condition are retained better in the long run than those learned in an easy learning condition. For paired-associate foreign word learning tasks—comparable to those that Schneider *et al.* (2002) employed but different in languages used and participant population—Study 1 largely replicated their findings. Meanwhile, we arrived at a new assumption regarding the retention and transfer of knowledge according to task difficulty. Namely, easy tasks and difficult tasks might have different effects on retention and transfer, meaning that easy learning contributes to the retention of knowledge rather than promoting its transferability, and difficult learning facilitates the transfer of knowledge rather than enhancing its retention.

The purpose of Study 2 is to examine whether the findings in Study 1 also apply in typical language classroom settings. Therefore, this is a conceptual replication of Study 1. Research questions will be established in the same orientation as in Study1. However, because the training difficulty principle (Schneider *et al.*, 2002; Healy *et al.*, 2012) predicts that any type of difficulty during learning facilitates the subsequent retention and transferability of the learned knowledge, and because of our intention to simulate learning in practical language classrooms, the three following major changes were made in the present experiment. First, the implementation of task difficulty occurred during a post-learning training rather than at the time of encoding. It is safe to say that initial learning involves both memorizing new information and relevant practice right after learning. Reading was employed as a way to practice newly learned words. In typical language classrooms, teachers often introduce reading materials for practice after students learn new words in a deliberate manner. Second, the levels of task difficulty were operationalized by manipulating the +/– here-and-now variable of Robinson's task complexity (1995, 2001a, 2001b, 2003). This modification served as a test for the training difficulty principle. Finally, the transferability of word knowledge was tested using a picture-naming task, a method considered to involve more practically desirable use of language (e.g., saying what it is upon sight) than merely translating words in one language to another.

5.1 Research Questions and Hypotheses

Essentially the same research questions were adopted as in Study 1 for testing whether the findings in Study 1 could be replicated in different settings. That is, the hypothesis posits that foreign words learned in a difficult learning condition will be more likely to be retained than those learned in an easy learning condition. However, Study 2 takes learning as a set of activities involving both memorizing new words and practicing them, rather than as a single memorizing event. Therefore, the following research questions and hypotheses were formulated for the current study.

Research Question 1:

> Are post-hoc exercises more effective for an easy task than a difficult one? Specifically, when the initial memory levels of unfamiliar L2 words are equal, do participants reading passages containing the new words in a simple here-and-now training condition—operationalized with the presence of meanings of the new words in glossaries in the margins—respond to while-reading comprehension questions faster and

more accurately than those who read the passages without such glossaries in a complex there-and-then condition?

Three hypotheses are proposed for this question.

Hypothesis 1:

First, in terms of the training difficulty principle (Healy *et al.*, 2012; Schneider *et al.*, 2002), training in any easier conditions, as opposed to more difficult counterpart conditions, should yield more learning gains. Previous researchers argued that training in an easy way is advantageous in the short term. We initially hold this hypothesis in light of the findings in Study 1.

Hypothesis 2:

Second, in Robinson's view (2001a, 2001b, 2003), learners should perform more slowly but more accurately on a complex task than a less complex task. He asserted that cognitively complex tasks make learners pay attention to the target structures more often, thus leading to greater accuracy in production.

Hypothesis 3:

Finally, in the depth-of-processing account (Craik & Lockhart, 1972), there should be no effect because the theory predicts that the strength of a memory trace can only be determined at the time of encoding. Here, we will not manipulate task difficulty at the time of initial encoding but during a post-hoc exercise.

Research Question 2:

Does the complex there-and-then post-hoc reading exercise promote performance in the delayed recall of the learned words, in terms of retention and transfer as a whole? Specifically, when tested a week after learning—in which half the participants of each exercise condition work on the same translation task at the time of learning (retention) and the other half on a picture-naming task (transfer)—do participants in the complex there-and-then exercise condition recall more target words than participants in the simple here-and-now exercise condition?

Two hypotheses can be mentioned regarding this question.

Hypothesis 4:

In accordance with predictions by Healy *et al.* (2012), Schneider *et al.* (2002), and Robinson (2001a, 2001b, 2003), participants who practice learned words via more difficult or cognitively more complex task should produce more correct responses a week after learning than those who practice on easier or less complex task. We initially hold this hypothesis.

Hypothesis 5:

From the depth-of-processing view of human memory (Craik & Lockhart, 1972), there will be no effect in learning since the theory holds that the degree of retention can only be determined at the time of encoding.

Research Question 3:

Do the simple exercise and the complex exercise each show different effects on delayed retention and transfer, as observed in Study 1? Specifically, does the simple exercise lead to the maintained retention of knowledge rather than its transferability, whereas the complex exercise promotes the transfer of knowledge rather than its retention?

Hypothesis 6:

Unlike the questions mentioned above, only one hypothesis can be mentioned based on the replicated patterns of Schneider *et al.*'s results. That is, the simple here-and-now post-hoc exercise will promote retention of the learned words in the long term, whereas the complex there-and-then exercise will maintain transferability of knowledge.

5.2 Participants

Fifty-two students in the same participant pool of Study 1 participated in this study. Half of the participants were invited to the present experiment before conducting the Study 1 experiment sessions, while the remaining half participated in the reverse order. They received rewards in the same manner as in Study 1. However, eight students were excluded from later analyses because they took extraordinarily shorter or longer times to memorize target items at the very first phase of the present experiment, making them outliers situating out of the 95% confidence interval. We recognized that the initial encoding of target items should have been severely regulated because it would make some confounding effects on the later recall performance highly likely. Consequently, the number of participants became 43, with 22 being assigned to a simple training condition and 21 to a complex training condition.

5.3 Materials

Three paper-based materials and three computer-based materials were prepared for the present experiment. The following section describes each material.

5.3.1 English Vocabulary Size Test (Paper-based)

To assess participants' levels of English proficiency, a vocabulary size test was conducted. In contrast to vocabulary depth tests, which examine participants' detailed knowledge of individual words with many questions, vocabulary breadth or size tests are useful to measure the estimated size of the vocabulary a learner has with a small number of questions (Meara, 2009), which is economical. By this means, researchers can also roughly estimate learner's English proficiency based on standardized English tests such as TOEIC, TOEFL, or Eiken. For example, Chujo and Oghigian (2009) determined that a reader would need a minimum vocabulary size of 4,000 English words to gain 95% coverage on TOEIC, a 4,500-word vocabulary on TOEFL, and a 5,500-word vocabulary on Eiken Pre-1st Grade. Generally, it would be useful to provide participant information along with the levels of standardized English tests, especially when a major readership consists of teachers.

The question items of the present vocabulary size test were taken from the Vocabulary Size Test (bilingual Japanese version) available on Paul Nation's website. This is a test of receptive English vocabulary. The original version includes questions for the most frequent 1,000 words to a 14,000-word level in a multiple-choice format. The first 10 questions assess knowledge of the most frequent 1,000 words, the second

10 questions evaluate knowledge of the second frequent batch of 1,000 words, and so on up to the 14,000-word level. Among these questions, for this study, the questions ranged from the 2,000-word level to the 9,000-word level. Therefore, the test measured receptive English vocabulary knowledge up to 9,000 words, postulating that all participants had already known the most basic 1,000 words. Considering that all participants were undergraduate students at a key university in the country, the most frequent 1,000 words were certainly too easy and did not require assessment to save time.

Figure 5.3.1

Format of the Vocabulary Size Test.

21. He is a **soldier**.
    a.実業家　b.学生　c.金属細工人　d.兵士

22. It has been **restored**.
    a.繰り返された　b.移転された　c.値引きされた　d.復元された

The test employed a multiple-choice format for the Japanese meaning of a target English word provided in bold letters in a sentence (Figure 5.3.1). An estimated vocabulary size for each participant was calculated in the formula [the number of correct answers × 100 + 1,000], with a possible range from 1,000 to 9,000. All the questions on the test are provided in Appendix E.

5.3.2 Target Word-Pair Sheets (Paper-based)

Two stapled A4-size sheets of papers were used to introduce the target word pairs. The first page included very unfamiliar English words that even native speakers might not know, which served as nonwords in this study. To ensure a certain degree of linguistic naturalness, the target words were carefully selected from *A Dictionary of Hiberno-English* (Dolan, 1999), an Irish dialectal English dictionary, but they are not listed in common English dictionaries. In addition, to equalize their readability, only those of six-letter words were randomly employed.

The second page contained meanings for those nonwords listed on the first page in common English. The meaning for each nonword was not an actual meaning shown in the Irish English dictionary, but the researcher artificially assigned an easy concrete English noun that appeared in the reading passages in the treatment task (described in a later section) to each nonword.

A set of the sheets contained 12 nonwords on the first page and 12 counterpart meanings on the second page. There were two sets, versions A and B, so the participants need to learn 24 word pairs in total. The association between a particular nonword and its meaning could be identified by placing these pairs of words in the same positions on each page as shown in Figure 5.3.2. That is, participants were instructed, for example, that the word "ponger" meant "hand" and that the word "gistra" meant "bag." All the words were printed in black, 20-point Times New Roman font. Participants memorized these associations by flipping pages back and forth at their own pace in the initial rote learning session. The elapsed memorization time was measured with a stopwatch and recorded. A full list of the target word pairs is provided in Table 5.3.2.

Figure 5.3.2

A Set of Word-pair Sheets.

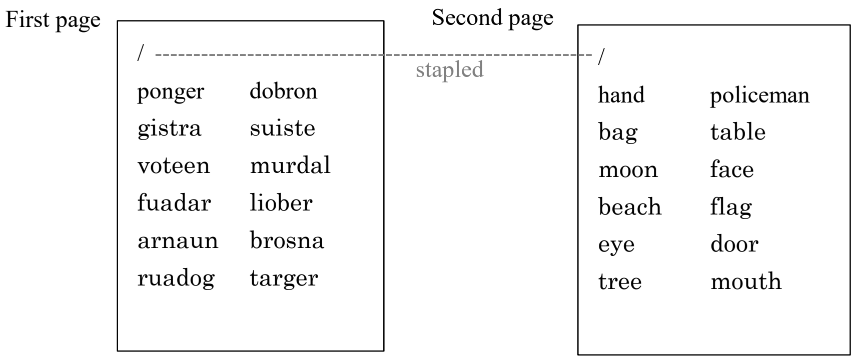


Table 5.3.2

A Full List of Target Word Pairs.

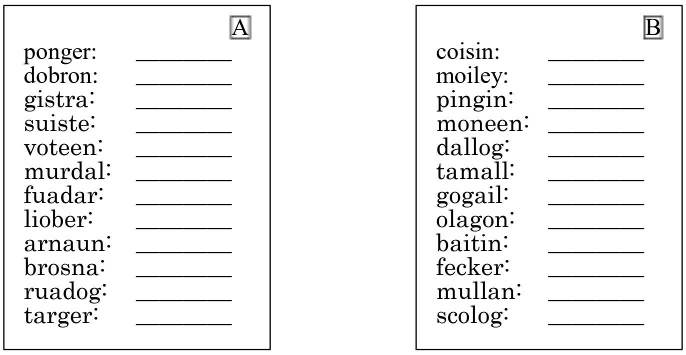
Target Nonword–Meaning	
Version A	Version B
ponger – hand	coisin – hat
dobron – policeman	moiley – wing
gistra – bag	pingin – telephone
suiste – table	moneen – cafe
voteen – moon	dallog – newspaper
murdal – face	tamall – sun
fuadar – beach	gogail – photograph
liober – flag	olagon – neck
arnaun – eye	baitin – clothes
brosna – door	fecker – boy
ruadog – tree	mullan – town
targer – mouth	scolog – tower

5.3.3 Answer Sheets for Initial Learning (Paper-based)

Immediately after memorizing 12 word associations on each word-pair sheet (version A or B), participants were asked to write down the meaning for each nonword in English on an answer sheet (Figure 5.3.3). Participants were allowed to proceed to the next task, a simple- or a complex-conditioned post-hoc treatment task, only after demonstrating 100% accuracy for the meanings of all nonwords. If not, they had to memorize the items again and retake the test until reaching that criterion. This step ensured that all participants had an equal level of knowledge of the target items, in other words, to equalize their baseline memory for subsequent experiment sessions.

Figure 5.3.3

Format of the Answer Sheets.



5.3.4 Reading Tasks (Computer-based)

The researcher developed treatment English reading tasks on SuperLab. The purpose of this reading task was to make the participant use the memorized words through reading, similar to method employed in typical language classrooms. Participants saw the 24 target nonwords in two different reading passages. Half of the nonwords, those provided on the version A word-pair sheets, were embedded in passages of *Tinkers Island* (Rabley, 1998a), a Penguin graded reader for beginners, and the other half, those on the version B word-pair sheets, were incorporated in *Flying Home* (Rabley, 1998b), another Penguin graded reader of the same level. Both readings consisted of approximately 700 words, with the easiest 200 headwords in English. On a computer screen, they were shown in black, 20-point Georgia font.

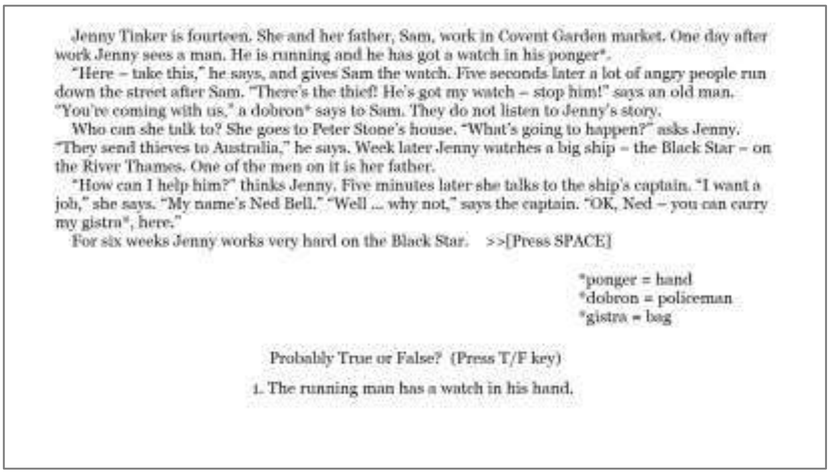
The reading passages were presented by dividing them into four parts of equal length. Each of the parts occupied one screen and contained three target nonwords. Participants read the passages at their own pace. On each screen, when pressing the spacebar to proceed, a true–false question for the passage at the present moment additionally appeared at the bottom of the screen. Three questions appeared per screen, and they were always targeted at one of the nonwords on the present screen. The implementation of while-reading comprehension questions was intended to provide participants opportunities to recall the learned words. Participants answered these questions by pressing either the “T” or “F” button on the keyboard. Response times were measured for every single keystroke during the task so that fluency could be analyzed retrospectively together with their responses.

Participants worked on this task in either a simple training condition or a complex training condition. The two levels of task complexity were operationalized by manipulating the +/– here-and-now parameter of Robinson’s (2001a, 2001b, 2003) task complexity framework. The + here-and-now condition featured everything necessary to complete the task being present at sight, suggesting no deep processing in memory during the task, while the – here-and-now condition (or the there-and-then condition in another term;) emphasized the need for memory of the learned items because critical information was not present during the task. Therefore, the only difference between the two levels of task complexity in the present task was whether or not the meanings of the target nonwords appeared in glossaries in the margins (Figure 5.3.4).

Figure 5.3.4

Format of the Reading Tasks.

In the simple here-and-now reading condition



Jenny Tinker is fourteen. She and her father, Sam, work in Covent Garden market. One day after work Jenny sees a man. He is running and he has got a watch in his ponger*.

"Here – take this," he says, and gives Sam the watch. Five seconds later a lot of angry people run down the street after Sam. "There's the thief! He's got my watch – stop him!" says an old man. "You're coming with us," a dobron* says to Sam. They do not listen to Jenny's story.

Who can she talk to? She goes to Peter Stone's house. "What's going to happen?" asks Jenny. "They send thieves to Australia," he says. Week later Jenny watches a big ship – the Black Star – on the River Thames. One of the men on it is her father.

"How can I help him?" thinks Jenny. Five minutes later she talks to the ship's captain. "I want a job," she says. "My name's Ned Bell." "Well ... why not," says the captain. "OK, Ned – you can carry my gistra*, here."

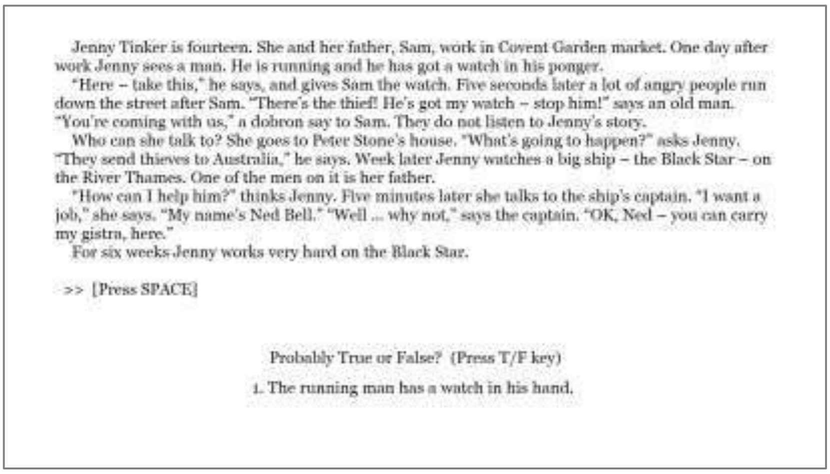
For six weeks Jenny works very hard on the Black Star. >>[Press SPACE]

*ponger = hand
*dobron = policeman
*gistra = bag

Probably True or False? (Press T/F key)

1. The running man has a watch in his hand.

In the complex there-and-then reading condition



Jenny Tinker is fourteen. She and her father, Sam, work in Covent Garden market. One day after work Jenny sees a man. He is running and he has got a watch in his ponger.

"Here – take this," he says, and gives Sam the watch. Five seconds later a lot of angry people run down the street after Sam. "There's the thief! He's got my watch – stop him!" says an old man. "You're coming with us," a dobron say to Sam. They do not listen to Jenny's story.

Who can she talk to? She goes to Peter Stone's house. "What's going to happen?" asks Jenny. "They send thieves to Australia," he says. Week later Jenny watches a big ship – the Black Star – on the River Thames. One of the men on it is her father.

"How can I help him?" thinks Jenny. Five minutes later she talks to the ship's captain. "I want a job," she says. "My name's Ned Bell." "Well ... why not," says the captain. "OK, Ned – you can carry my gistra, here."

For six weeks Jenny works very hard on the Black Star.

>> [Press SPACE]

Probably True or False? (Press T/F key)

1. The running man has a watch in his hand.

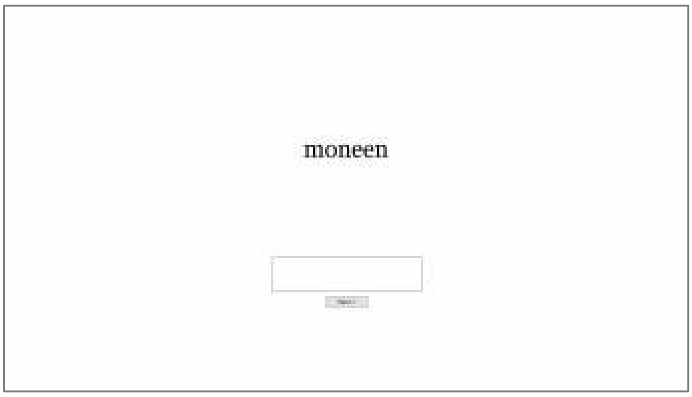
Participants in the simple here-and-now reading condition were able to read passages, as well as answer while-reading comprehension questions, with the assistance of glossaries in the margin. In contrast, participants in the complex there-and-then reading condition had to rely on the recall of the target nonwords. Therefore, it was assumed that the complex there-and-then condition imposed deeper processing on participants than did the simple here-and-now condition. All the reading passages and corresponding comprehension questions are provided in Appendix F for version A reading and in Appendix G for version B.

5.3.5 Retention Posttest (Computer-based)

This retention test required participants to type in the meaning of each target nonword in English on the computer screen (Figure 5.3.5). They typed in the corresponding meanings in response to the target nonwords that appeared randomly one by one in the middle of the screen in black, 36-point Georgia 36 font. This constituted a retention test because participants needed to recall words in the same way as they learned. The self-paced task ran on SuperLab. All the test items are provided in Appendix H.

Figure 5.3.5

A Sample Screen of the Retention Posttest.



5.3.6 Transfer Posttest (Computer-based)

In the transfer posttest, participants saw a picture related to one of the target word pairs and then typed in the corresponding nonword. For example, in the case shown in Figure 5.3.6, the participant should type in “dobron” in response to the picture of a policeman. Twenty-four different pictures depicting each word pair were presented one by one in a randomized order (see Appendix I for all the test items).

Unlike the transfer posttest in Study1 in which participants saw a word and answered the translation equivalent in the opposite translation direction from the learning process, this test provided a picture as a cue to test acquired knowledge in more practical situations. Bourne and Healy (2012) mentioned that the transferability of acquired knowledge refers to the generalizability of that knowledge. In this sense, measuring knowledge learned through printed words in another modality (i.e., via vision) would be preferable to merely measuring this aspect within the same modality. Moreover, in a common second-language environment, it is certainly desirable for students to have the capacity to name things in that language rather than merely translate words from one language to another. Similar to the retention posttest, the task was self-paced and ran on SuperLab.

Figure 5.3.6

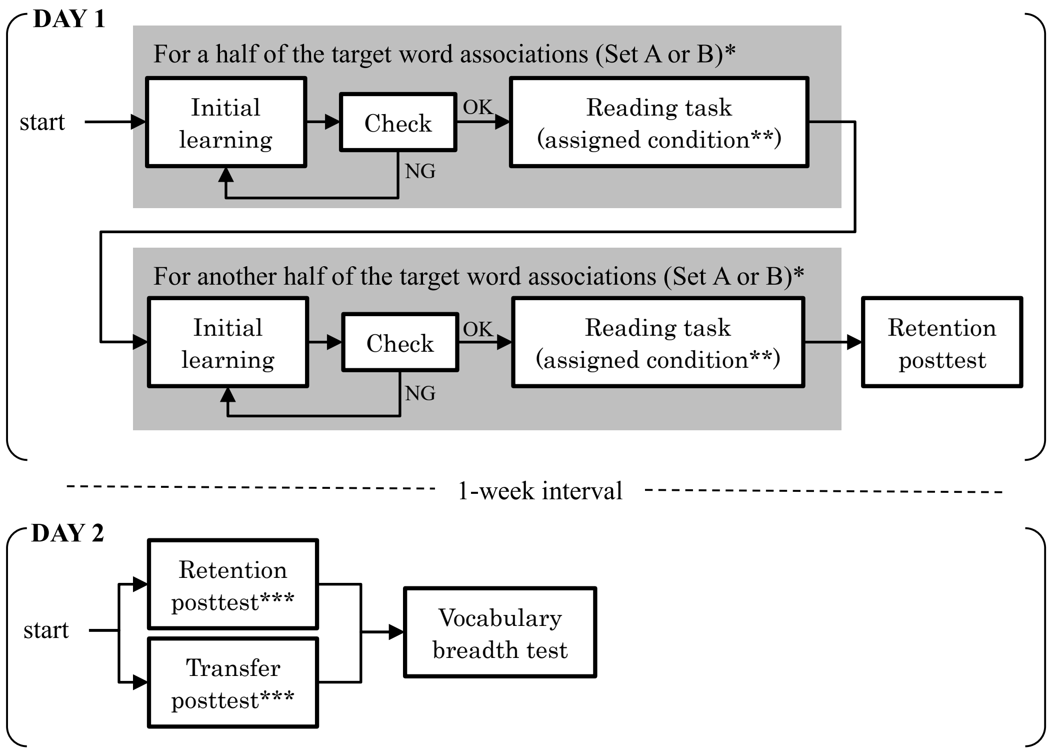
A Sample Screen of the Transfer Posttest.

5.4 Procedure

The experiment sessions of Study 2 required two separate days, the first day including learning sessions, treatment post-hoc reading tasks with two different complexity levels, and an immediate retention posttest. The second day involved a delayed posttest and a vocabulary size test. There was a one-week interval between the two days. Participants were examined one by one in a lab-based environment. The sequence of the experiment is summarized in Figure 5.4.

Figure 5.4

A Flowchart of Study 2.



* The presentation order of set A and set B is altered for each participant.

** The same participant always reads in the same condition (easy or difficult).

*** Half of participants take the retention posttest, and the other half take the transfer posttest.

At the beginning of day 1, participants learned half of the 24 target word pairs provided on either version A or version B of the word-pair sheet at their own pace. The researcher recorded time spent learning. Participants received a pencil and an answer sheet to write down the meanings of the target nonwords. The researcher checked the answers quickly based on prepared answer keys and introduced a post-hoc reading task only when all the answers were correct. Participants who made errors received the word association sheet to memorize again, the encoding time counted again, and participants wrote down the meanings of all the target nonwords on another blank answer sheet. This process continued until each participant demonstrated 100% accuracy on all items.

After memorizing half of the target word associations correctly, participants read passages containing the target nonwords they had just memorized in either a simple or a complex reading condition. As mentioned earlier, the only difference between these two conditions was whether or not glossaries for the meanings of the target nonwords were provided during reading. This procedure was also done for the remaining half of the target word pairs.

After two sets of initial learning sessions followed by the reading task were completed (i.e., participant had been exposed to all 24 target nonwords and practiced them in the post-hoc reading task in either the simple here-and-now or the complex there-and-then training condition), participants took an immediate retention posttest.

A week later, on day 2, participants returned and took a delayed posttest. Half of the participants in each treatment reading condition on day 1 took a delayed retention test, completing the same task they did on the immediate posttest. The other half of each treatment reading condition on day 1 took a delayed transfer test in the form of the picture-naming task described earlier. Finally, an English vocabulary size test was conducted to collect background information (i.e., a rough estimate of English proficiency) on the participants.

5.5 Results

5.5.1 Vocabulary Sizes of the Participants

Receptive English vocabulary sizes of the participants ranged from 4,800 to 7,600, with a mean of 6,074. A Kolmogorov–Smirnov test of normality showed that they maintained a normal distribution. Moreover, a *t*-test for the differences in vocabulary sizes between the simple reading group (M = 5,990) and the complex reading group (M = 6,161) revealed no significant difference between them, $t(41) = -.880$, $p = .384$. In addition, correlation analyses of vocabulary size and accuracy on while-reading comprehension questions indicated that there were no particular correlations between them, $r = .190$, $p = .222$, suggesting that individual differences in vocabulary size were not related to their reading performances.

5.5.2 Encoding Time

Participants needed, on average, 523 seconds to memorize all the target words. After the exclusion of eight unreliably fast or slow memorizers, the fastest participant completed in 353 seconds, and the slowest one required 860 seconds. Separate correlation analyses showed that encoding time was not correlated with accuracy on while-reading comprehension questions ($r = .022$, $p = .890$), accuracy on the immediate posttest ($r =.113$, $p = .471$), or accuracy on the delayed posttest ($r = -.019$, $p = .905$). However, encoding time was weakly correlated with vocabulary size at a near significant level, $r = .291$, $p = .058$. Therefore, it was confirmed that variabilities in encoding time between participants were safely regulated.

5.5.3 Research Question 1

The first research question asked whether a simple here-and-now reading exercise would prove more effective than a complex there-and-then reading exercise as a part of an initial learning phase. The simple condition had 22 participants, and the complex condition had 21. Surprisingly, descriptive statistics revealed that the accuracy on while-reading comprehension questions was better for the complex there-and-then reading group than the simple here-and-now reading group on average, even though the simple reading group had opportunities to refer to the meaning of nonwords in glossaries (Table 5.5.3).

Table 5.5.3

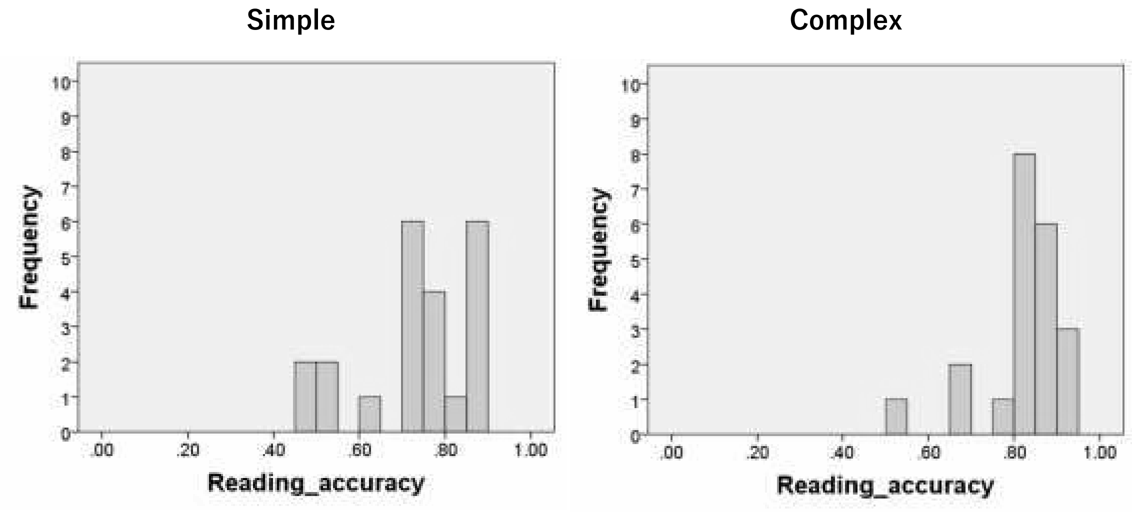
Descriptive Statistics on Response Accuracy on While-reading Questions.

	Simple Here-and-now Condition	Complex There-and-then Condition
Mean	.729	.824
Median	.730	.830
SD	.134	.100
Min	.460	.500
Max	.880	.920
Skewness	–.680	–2.805
Kurtosis	–.314	4.910

The minimum and maximum in accuracy were both higher for the there-and-then reading group than the here-and-now reading group, indicating that the complex task condition raised the minimum level of learning as a whole. Furthermore, the *SD* was smaller for the there-and-then reading group than the here-and-now reading group, suggesting an overall stable performance in the group. We examined distributions of data for calculations of statistical significance. Two separate tests of normality showed that the accuracy on while-reading questions was not normally distributed for either experiment group (see Figure 5.5.3.1). Therefore, non-parametric tests were employed.

Figure 5.5.3.1

Histograms on Response Accuracy on While-reading Questions.

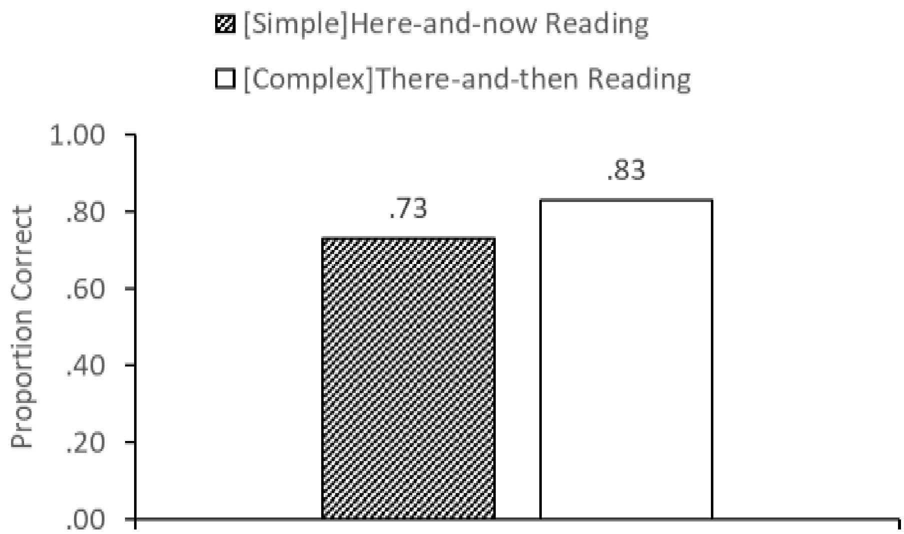


A non-parametric Mann–Whitney test was performed for the intensity of the median difference in accuracy on while-reading comprehension questions between the two experiment groups. The results demonstrated that the complex there-and-then reading group responded to the comprehension questions requiring knowledge of the learned words significantly more accurately (MD = .83) than did the simple here-and-now reading group (MD = .73), U = 131.50, p = .014 (Figure 5.3.3.2). Furthermore, the mean elapsed time to answer each question was smaller for the complex reading group (M = 18.4 sec) than the simple reading group (M = 22.7 sec), although the difference did not reach a significant level (p = .190).

Together, the results raised a question regarding the training difficulty principle (Healy *et al.*, 2012; Schneider *et al.*, 2002), which holds that training in any easy condition is advantageous in terms of yielding short-term gains. Meanwhile, the results indicated partial support for Robinson's view of task complexity (2001a, 2001b, 2003), which maintains that performance associated with existing knowledge (i.e., on a learner's present interlanguage system) on a complex task enhances production accuracy at the cost of fluency.

Figure 5.3.3.2

Proportions of Accurate Responses on While-reading Comprehension Questions.



5.5.4 Research Question 2

The second research question asked whether the complex there-and-then post-hoc reading exercise would promote performance in the delayed recall of the learned words. We first calculated descriptive statistics on recall accuracies on the immediate posttest and the delayed posttest for the two experiment groups and then tested the normality of the data distributions of each experiment group.

Table 5.5.4

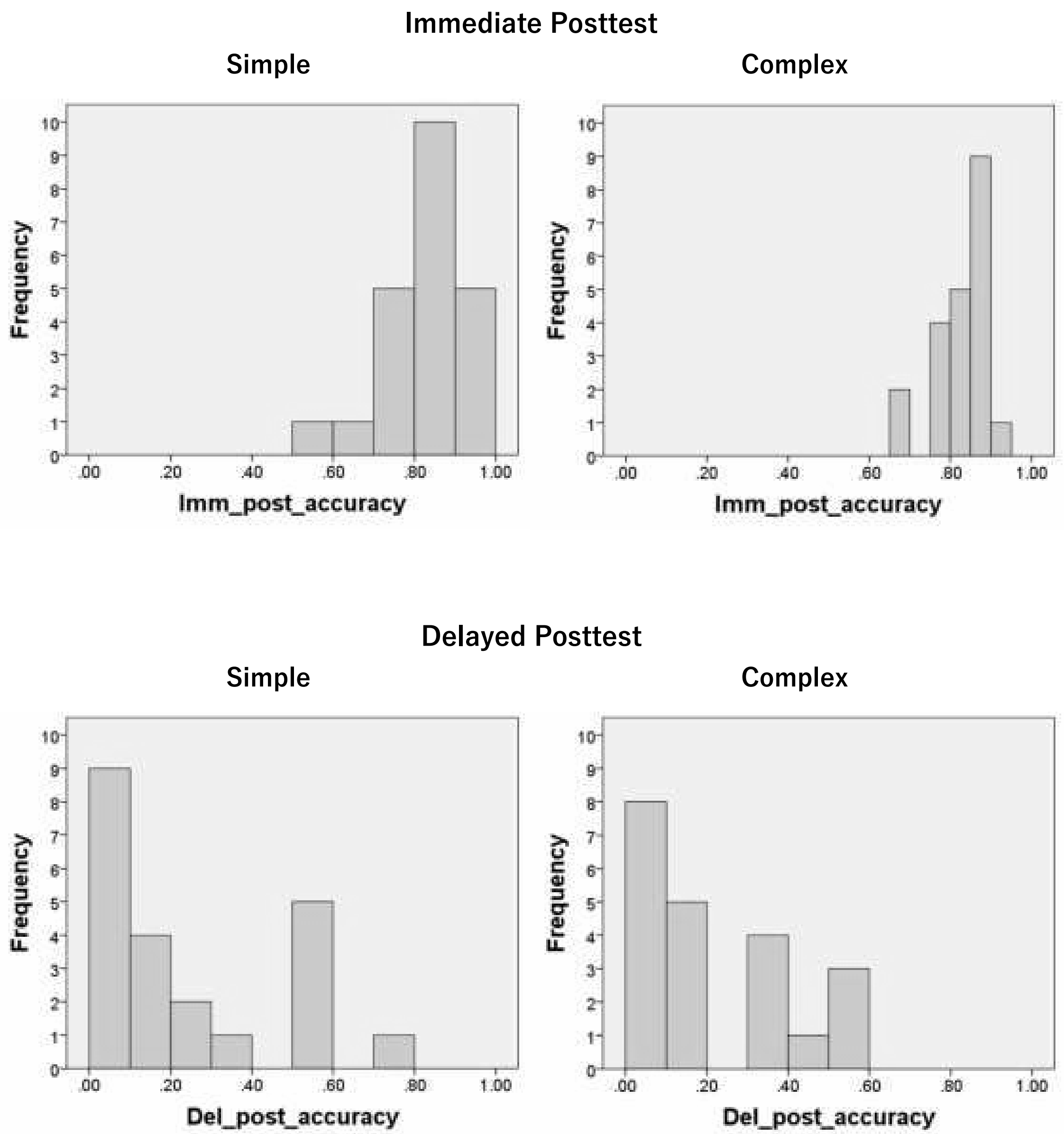
Descriptive Statistics on Response Accuracy on the Immediate Posttest and the Delayed Posttest.

	Simple Here-and-now Reading Condition		Complex There-and-then Reading Condition	
	Immediate	Delayed	Immediate	Delayed
Mean	.814	.222	.827	.209
Median	.830	.150	.830	.170
SD	.107	.231	.071	.202
Min	.500	0	.670	0
Max	.960	.710	.920	.580
Skewness	–1.254	.744	–1.064	.558
Kurtosis	2.440	–.898	.283	–1.029

As shown in Table 5.5.4, descriptive statistics for the immediate posttest showed that the complex there-and-then post-hoc reading group slightly outperformed (M = .827) the here-and-now post-hoc reading group (M = .814) in the mean accuracy, reflecting some savings in retention from learning. Moreover, the minimum performance was higher for the there-and-then reading group (Min. = .670) than the here-and-now reading group (Min. = .500). Furthermore, the smaller SD value of the there-and-then reading group (SD = .071) in comparison with that of the here-and-now reading group (SD = .107) suggested that complex post-hoc exercises contributed to an invariant subsequent performance to some extent. However, the results for the delayed posttest, showing participants' recall performance after a 1-week retention interval, revealed that the here-and-now reading group slightly outperformed (M = .222) the there-and-then reading group (M = .209) in terms of the mean recall accuracy. Some participants of each group forgot all the target items.

Figure 5.5.4.1

Histograms on Recall Accuracy on the Immediate Posttest and the Delayed Posttest.



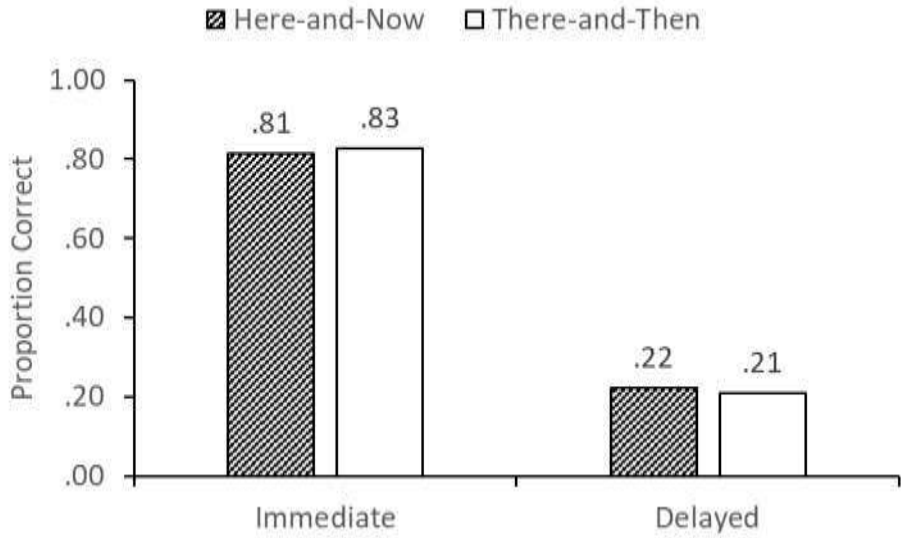
Four separate Kolmogorov–Smirnov tests were performed to determine the normality of the data distributions for the test types (immediate vs. delayed) and independent variables (simple vs. complex). The results showed that all of the four histograms did not maintain a normality, with a trend of negative skewness for the immediate posttest and a positive skewness for the delayed posttest (Figure 5.5.4.1).

Regarding the intensities of the differences in the recall accuracy between the here-and-now reading group and the there-and-then reading group on both the immediate test and the delayed posttest, two non-parametric Mann–Whitney tests were performed. According to the results, the recall performances of the two experiment groups were not statistically different for either the immediate posttest, $U = 214.00$, $p = .672$, or the delayed posttest, $U = 226.00$, $p = .902$. Concerning another facet, non-parametric Wilcoxon signed-rank tests for the within-participant median difference between the two occasions of the posttests revealed that the learned knowledge had decayed significantly from the time of the immediate posttest to the delayed posttest in both groups ($Z = -4.109$, $p < .001$ for the here-and-now group; $Z = -4.021$, $p < 001$ for the there-and-then group).

Overall, the learning gains of the complex there-and-then post-hoc reading group disappeared for the times of posttests. The recall probabilities of both experimental groups almost leveled off, even at the time of the immediate posttest. The delayed recall performances deteriorated significantly from those on the immediate posttest regardless of the task conditions at learning, reflecting a function of time (Figure 5.5.4.2). The findings did not support the initial hypothesis that practice on learned words on a cognitively complex learning task would promote the subsequent recall of knowledge more than working on a cognitively simple learning task. Meanwhile, the results suggested support for the depth-of-processing framework of human memory (Craik & Lockhart, 1972), holding that the same levels of encoding yield memory traces at the same strength.

Figure 5.5.4.2

Proportions of Correct Recalls on the Immediate Posttest and the Delayed Posttest.



5.5.5 Research Question 3

On the grounds of the replicated findings in Study 1, we asked whether the simple and complex post-hoc exercises would each show different effects on the delayed retention and transfer of the learned words in the third research question. Here, the present research results identified that retention involved participants' capacity to reproduce what they had done during the learning process in exactly the same way, and transfer referred to utilizing learned knowledge on a task in a slightly different way. Therefore, the present study employed the same translation task as that presented on the immediate posttest to examine knowledge retention, and a picture-naming task was adopted to measure knowledge transfer.

Analyses for this research question were conducted in the same way as for Study 1. That is, we first categorized the delayed posttest data into four categories, based on possible combinations between levels of task complexity and task types, and then calculated the recall probabilities for each case. The results indicated that no such hypothesized effects occurred on the delayed recall performances, as indicated in Table 5.5.5.

Table 5.5.5

Proportions of Correct Responses on the Immediate Posttests and the Delayed Posttests Based on Task Types.

Post-hoc Reading Condition	Delayed Posttest	
Here-and-now	Retention	.39
	Transfer	.06
There-and-then	Retention	.33
	Transfer	.10

The breakdown figures on the delayed recall performances showed a similar trend for both experiment groups: Overall, the retention rate was higher than that of the transfer. Furthermore, non-parametric Mann–Whitney tests confirmed that the delayed retention performances were significantly better than the delayed transfer performances for both the here-and-now post-hoc reading group (.39 vs. .06, $U = 11.00$, $p = .001$) and the there-and-then group (.33 vs. .10, $U = 20.50$, $p = .013$). In contrast, when comparing the delayed retention performances between two groups, they were not significantly different (.39 vs. 33, U = 45.00, p = .512), which was also the case for the delayed transfer performances (.06 vs. .10, U = 54.500, p = .699). As a whole, the results did not support the hypothesis.

5.6 Discussion

In Study 2, we extended the previous replication study on the effects of task difficulty during learning on the subsequent retention and transfer of knowledge of newly learned words in three ways. First, manipulations of task complexity, as a synonym for task difficulty, were implemented at the time of a post-hoc exercise after memorizing new words rather than at the time of encoding those words. On the grounds that the introduction of new words and practice with them are often set activities done within the same class time, it would be safe to say that post-hoc exercises form part of initial learning. Second, we operationalized levels of task complexity by manipulating the +/– here-and-now parameter of Robinson's framework of task complexity using English rather than manipulating the translation directions between the two languages. This change also simulated monolingual second-language learning environments, which are now employed in a wide variety of English courses at many universities. Finally, the transferability of word knowledge was measured via a picture-naming task, a method considered to involve more practically desirable use of language than merely translating words from one language to another. We will discuss the results obtained in Study 2 focusing on these changes.

5.6.1 Learning Performances on Simple and Complex Post-hoc Exercises

Based on our replicated findings in Study 1, we did not expect that a less complex task would hinder performance during learning. Although reading after learning new words is essentially a technical task compared with recalling those words straightforwardly, it is still problematic to grasp the results because students in the simple here-and-now reading condition were able to look at meanings of the nonwords in glossaries during the task. A possible explanation may be that the here-and-now reading group might have been somewhat disrupted while reading. Because participants in the here-and-now reading condition had overall smaller vocabulary sizes than the there-and-then reading group, their reading proficiency could have had potential negative effects on their performance. It is also a conceivable grounds for why the here-and-now reading group actually took longer to answer the while-reading comprehension questions than the there-and-then group did.

The here-and-now reading group's inaccurate responses on the while-reading questions were probably not because they forgot the meanings of the target nonwords, as indicated by their equally accurate responses to the other group on the immediate posttest. An important point here is to determine what was actually measured on the post-hoc reading exercise. We initially anticipated that answering questions with clues would be easier than answering while remembering the words and that this might cause a variability in depth-of-processing during memorization. Therefore, we expected that the trends in performance between this post-hoc exercise and an immediate posttest were congruent. However, the results did not exhibit that pattern. Considering the fact that their knowledge of the learned words had not been lost by the time of the immediate posttest, it is suggested that we manipulated a different aspect other than memorization at the level of initial learning. We considered that what we had manipulated during the post-hoc exercise were the levels of the immediate usability of the newly learned words and that this aspect was relatively unrelated to the encoding itself. The complex there-and-then reading condition created a situation that supported learners' reference to newly learned information. Assumptions in this direction, as a result, explain the results nicely. Nevertheless, a question remains in terms of the generation effect of skill acquisition theories, which suggests that learning in a productive way enhances the retention of learned information.

5.6.2 Subsequent Recall Probabilities via Simple and Complex Post-hoc Exercises

The two experiment groups practiced memorized words via two tasks of different complexity, and each showed similar levels of recall probabilities on subsequent recall tasks. Therefore, the second hypothesis (that learners in more complex there-and-then training conditions would show greater subsequent recall probabilities than those in less complex here-and-now training conditions) was rejected. Explanations for these results can be made on several theoretical grounds. The most straightforward understanding would be that the experiment obtained results consistent with the classic depth-of-processing framework (Craik & Lockhart, 1972). This theory suggests that the degree of the retention of new information can only be determined depending on how deep the learner analyzes the information during encoding. Actually, all the participants encoded target nonwords in rote memorization at the very first phase of learning. Therefore, it is rationally conceivable that the same levels of encoding yielded memory traces at the same strength.

Another interpretation could be that the superior learning performance on the complex there-and-then post-hoc exercise might be in effect only during the task. This view, at the same time, implies that gains during learning do not carry over into subsequent retention automatically. Therefore, as suggested by Robinson (2001a, 2001b, 2003), the consolidation of knowledge requires more learning trials through, for example, tasks in which the resource-depleting dimensions of his framework of task complexity (e.g., more prior knowledge) are maintained. Actually, in the present experiment, participants were allowed to memorize new words and practice them only once.

5.6.3 Effects of Simple and Complex Post hoc Exercises on Subsequent Retention and Transfer

The last point concerned the replicated findings in Study 1 on distinct effects at two levels of task complexity on subsequent retention and transfer. We hypothesized that the simple here-and-now post-hoc exercise would lead to the retention of learned words but not transfer and that the complex there-and-then post-hoc exercise would lead to the transfer of learned words but not retention. Contrary to the hypothesis, this study revealed that the two post-hoc exercises at different task complexity levels led to similar patterns in delayed recall performance. That is, retention performance was always higher than transfer performance regardless of the differences in task complexity after encoding. The results were, again, consistent with the depth-of-processing theory (Craik & Lockhart, 1972). Meanwhile, the results indicated that our view on the cause of superior transfer performance in Study 1 might not be so misdirected. We suspected, as far as retention and transfer are tested in translation directions between L1 and L2, that the levels of participants' familiarity with those languages would necessarily be involved in transfer performance. To put it in extreme terms, production in an L1 is always more advantageous than in an L2, especially at the early stages of learning. Therefore, we can now say that the replicated results of Schneider *et al.* (2002) in the previous chapter were considered to be a distinctive case when the translation directions were manipulated. Yet, in terms of absolute values, the simple here-and-now training group slightly outperformed the complex there-and-then training group in terms of delayed retention performance (.39 vs. .33), and the opposite was true for transfer performance (.06 vs. .10). Therefore, the results revealed the hypothesized pattern with a small sample size, although the extent of the effects was scarce. While this study failed to confirm this hypothesis, we would prefer not to dismiss this idea because it could serve as a further line of research in studies on task difficulty.

Chapter VI

THE PRESENT STUDY—ADDITIONAL INVESTIGATIONS ON THE PRESENT TASKS

6.0 Introduction

In this chapter, the tasks used in the present study are examined with a particular emphasis on perceived difficulty of, as well as the similarities and differences between, the Study 1 tasks and Study 2 tasks. Thus far, we have looked at experiments based on task difficulty or task complexity, but one might ask whether the participants in this study really found those tasks difficult. Indeed, this is a crucial point for the reliability of the present research. In this sense, it could be said that a methodological flaw in Schneider *et al.* (2002) was that the actual difficulty of each of their tasks was not known objectively. They presupposed that participants would find their tasks difficult to complete based on Kroll and Stewart's (1994) account of bilingual lexicons, which holds that the mental lexical association of a word from an L2 to an L1 translation equivalent is stronger than that in the opposite direction because initial learning usually involves L2-to-L1 learning. Therefore, theoretically, learning in an L1-to-L2 direction must be more difficult than learning in the opposite direction. We agree with this assessment. However, such a view remains within a theoretical realm, and we actually will not know the difficulty of a particular task without confirming it objectively, and we had an opportunity to do so. Therefore, we conducted a small-scale study and asked students about the difficulty of the tasks used in the present study.

The other investigation of this chapter concerns a contrast between the Study 1 and Study 2 tasks of the present study. As mentioned in previous chapters, the overall results of Studies 1 and 2 differed considerably, even though both studies were conducted along the same lines and concerned retention and transfer at different levels of task difficulty, and the same individuals participated in both studies. Therefore, we conducted correlational analyses on participants' performances in the different tasks to identify the similarities and differences of those tasks by re-examining the collected data. In the following sections, we will look at a study for the perceived difficulty of the present tasks and then proceed to some correlational analyses of them.

6.1 Perceived Difficulty of the Present Tasks

6.1.1 Participants

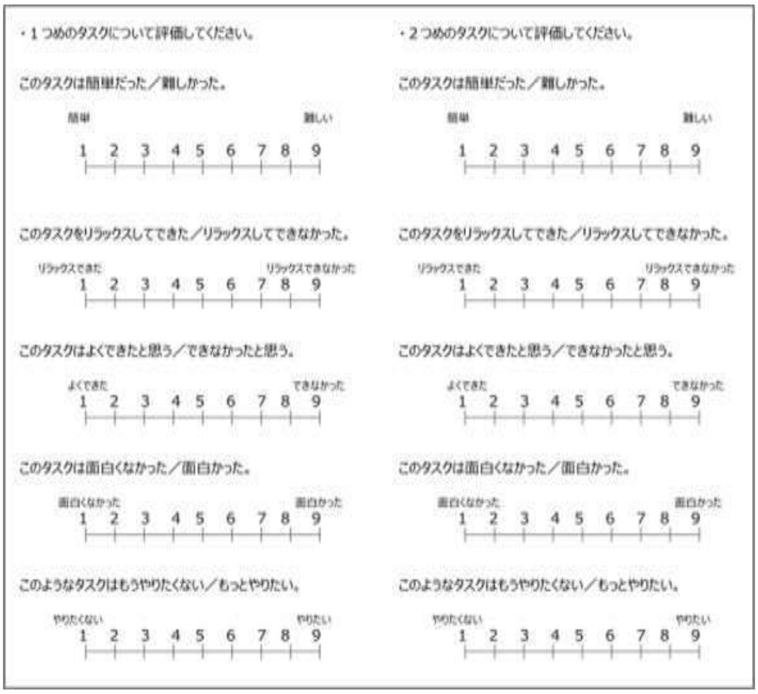
Twelve Japanese-speaking students at Aoyama Gakuin University, who registered in a different participant pool from the main parts of the present study, participated in the present investigation. They were all female undergraduates in the English department. The participants received 1,000 JPY for a session when it was held on weekdays or 1,500 JPY when conducted on weekends or during a vacation period.

6.1.2 Material

A Japanese version of the questionnaire employed in Robinson (2001a, p. 41) was used to assess participants' perceptions of the task difficulty. We translated the original questions from English into Japanese. The survey consisted of five questions on a 9-point Likert scale to evaluate participants' feelings about performing the tasks: (1) "I thought this task was easy/I thought this task was hard" (1: *not difficult*, 9: *very difficult*), (2) "I felt relaxed doing this task/I felt frustrated doing this task" (1: *no stress*, 9: *a lot of stress*), (3) "I didn't do well on this task/I did well on this task" (1: *not confident*, 9: *very confident*), (4) "This task was not interesting/This task was interesting" (1: *not interesting*, 9: *very interesting*), and (5) "I don't want do more tasks like this/I want to do more tasks like this” (1: *do not want*, 9: *would like to*). Two sets of the five questions were printed on a paper sheet for the comparison of the two individual tasks (Figure 6.1.2).

Figure 6.1.2

Questionnaire Sheet for the Perception of Task Difficulty.


・1つめのタスクについて評価してください。

このタスクは簡単だった／難しかった。

簡単 1 2 3 4 5 6 7 8 9 難しい

このタスクをリラックスしてできた／リラックスしてできなかった。

リラックスできた 1 2 3 4 5 6 7 8 9 リラックスできなかった

このタスクはよくできたと思う／できなかったと思う。

よくできた 1 2 3 4 5 6 7 8 9 できなかった

このタスクは面白くなかった／面白かった。

面白くなかった 1 2 3 4 5 6 7 8 9 面白かった

このようなタスクはもうやりたくない／もっとやりたい。

やりたくない 1 2 3 4 5 6 7 8 9 やりたい

・2つめのタスクについて評価してください。

このタスクは簡単だった／難しかった。

簡単 1 2 3 4 5 6 7 8 9 難しい

このタスクをリラックスしてできた／リラックスしてできなかった。

リラックスできた 1 2 3 4 5 6 7 8 9 リラックスできなかった

このタスクはよくできたと思う／できなかったと思う。

よくできた 1 2 3 4 5 6 7 8 9 できなかった

このタスクは面白くなかった／面白かった。

面白くなかった 1 2 3 4 5 6 7 8 9 面白かった

このようなタスクはもうやりたくない／もっとやりたい。

やりたくない 1 2 3 4 5 6 7 8 9 やりたい

6.1.3 Procedure

Data were collected in the same lab-based environment in which Study 1 and Study 2 were conducted. After filling in a participation agreement form (see Appendix A), participants performed both the easy learning task and the difficult learning task used in Study 1. The presentation order of the two tasks was counterbalanced for each participant. Upon completion of the two tasks, each participant was asked to mark on the questionnaire sheet by making a relative judgment on each item between the two tasks.

After a brief intermission, participants were instructed to memorize the nonwords used in Study 2 in the same manner as in that study. This time, for time constraints, they were allowed a maximum of 5 minutes to memorize the words. Moreover, the verification of whether or not each participant completely memorized the nonwords was skipped. Soon after, participants performed post-hoc reading tasks in both the simple here-and-now and complex there-and-then conditions. The presentation order of the two tasks was also counterbalanced for all participants. After completing the two tasks, participants marked on another questionnaire sheet based on a relative judgment between them.

The order of the overall presentation was counterbalanced. Therefore, some participants might evaluate on Study 1 tasks first and then Study 2 tasks, while others completed the form in the reverse order. Furthermore, tasks irrelevant to the purpose of this study, such as an immediate posttest, were not assigned, and task performances were not recorded. Only the responses to the questionnaires were recorded. At the end of the experiment session, we asked participants to identify the differences between the two tasks of the Study 1 task set and Study 2 task set separately.

This experiment employed within-participant repeated measures, in which the same individual compared two tasks that differed in terms of task difficulty for the two sets used in Study 1 and Study 2. Robinson (2001a) mentioned that task difficulty was a within-learner variable, in which the difference in difficulty between two tasks would become apparent only by the same learner comparing two tasks.

6.1.4 Results

6.1.4.1 Answers to Debriefing Questions

For the question about the difference between two tasks of Study 1, all participants reported that they saw Indonesian words and typed their translation equivalents in Japanese on one task, and then they saw Japanese words and typed the corresponding Indonesian words on the other. For the question about the difference between the two tasks of Study 2, all participants reported that they saw meanings of words while reading for one task but that those were not provided for the other task. It was thus confirmed that all participants were aware of the experimental manipulations.

6.1.4.2 Perceived Difficulty of Study 1 Learning Tasks

We first calculated descriptive statistics on participants' ratings of the Study 1 learning tasks. Of those replication tasks of Schneider *et al.* (2002), we assumed that tasks requiring participants to respond in Japanese from Indonesian cues were easier in comparison to those in the opposite direction. Therefore, the distinction between "easy" and "difficult" shown in the first row of Table 6.1.4.2 was made on that norm.

Table 6.1.4.2

Participants' Ratings of the Easy and Difficult Learning Tasks Used in Study 1.

	Study 1 Easy Learning Task					Study 1 Difficult Learning Task				
	D	S	C	I	M	D	S	C	I	M
Mean	4.33	3.92	6.00	6.75	5.75	6.83	5.50	3.83	5.83	4.58
Median	4	3	6	7	6	7	7	3	5	3
SD	2.53	2.53	2.83	1.88	1.64	2.23	3.10	2.03	2.27	2.47
Min	1	1	1	2	3	2	1	1	1	1
Max	9	8	9	9	8	9	9	7	9	9

D: Difficulty (1: not difficult - 9: very difficult), S: Stress (1: no stress – 9: a lot of stress), C: Confidence (1: not confident – 9: very confident), I: Interest (1: not interesting – 9: very interesting), M: Motivation (1: do not want – 9: would like to).

Based on the mean and the median of the perception item for difficulty, participants actually felt that the easy learning task of Study 1 was easier ($M = 4.33$, $MD = 4$) than the counterpart difficult learning task ($M = 6.83$, $MD = 7$). Furthermore, they felt less stress on the easy learning task ($M = 3.92$, $MD = 3$) than the difficult learning task ($M = 5.50$, $MD = 7$), and were more confident about the easy learning task ($M = 6.00$, $MD = 6$) than the difficult learning task ($M = 3.83$, $MD = 3$). Moreover, the easy learning task was more interesting ($M = 6.75$, $MD = 7$) and motivating ($M = 5.75$, $MD = 6$) than the difficult learning task ($M = 5.83$, $MD = 5$ for interest; $M = 4.58$, $MD = 3$ for motivation).

To evaluate the statistical significances of the differences of each perception measure, we performed five non-parametric Wilcoxon signed-rank tests for each type separately. The results showed that the easy learning task was perceived as being significantly easier than the difficult task, $Z = –2.128$, $p = .033$. Furthermore, participants were significantly more confident during the easy learning task than the difficult learning task, $Z = –2.015$, $p = .044$. The difference in stress also approached a statistically significant level, $Z = –1.786$, $p = .074$. In this way, we objectively confirmed that the task we intended to make easy was actually easier than the other task we thought would be difficult.

6.4.3 Perceived Difficulty of Study 2 Learning Tasks

We analyzed participants' responses to the perceived difficulty of the learning tasks used in Study 2 in the same way as that done above. We believed, in Study 2, that answering while-reading questions with the meanings of nonwords in glossaries should be easier or cognitively less complex than doing so without such reference materials. The distinction between "simple" and "complex" shown in the first row of Table 6.1.4.3

was made following that norm.

Table 6.1.4.3

Participants' Ratings of the Simple and Complex Training Tasks Used in Study 2

	Study 2 Simple Training Task					Study 2 Complex Training Task				
	D	S	C	I	M	D	S	C	I	M
Mean	2.83	2.25	6.25	6.08	5.58	5.08	3.92	4.67	6.58	6.00
Median	2	1	7	6	5	5	4	5	6	5
SD	1.67	1.83	2.01	2.36	2.14	2.06	2.25	1.49	1.93	1.83
Min	1	1	3	1	1	2	1	2	4	3
Max	7	7	9	9	9	8	9	7	9	9

D: Difficulty (1: not difficult - 9: very difficult), S: Stress (1: no stress – 9: a lot of stress), C: Confidence (1: not confident – 9: very confident), I: Interest (1: not interesting – 9: very interesting), M: Motivation (1: do not want – 9: would like to).

A descriptive analysis showed a different overall trend from that found in the analysis of the Study 1 tasks. Similarities were found in perceptions of difficulty, stress, and confidence, while trends in interest and motivation differed. Participants actually considered the simple training task of Study 2 easier (*M* = 2.28, *MD* = 2) than the counterpart complex training task (M = 5.08, Md. = 5). They felt less stress on the simple training task (*M* = 2.25, *MD* = 1) than the complex training task (*M* = 3.92, *MD* = 4), and they were more confident about the simple training task (*M* = 6.25, *MD* = 7) than the complex training task (*M* = 4.67, *MD* = 5). In contrast to patterns found for the Study 1 tasks, the complex training task was more interesting (*M* = 6.58, *MD* = 6) and more motivating (*M* = 6.00, *MD* = 5) than the simple training task (*M* = 6.08, *MD* = 6 for interest; *M* = 5.58, *MD* = 5 for motivation).

Non-parametric Wilcoxon signed-rank tests indicated that only the difference in the measure of difficulty was statistically significant, $Z = -2.096$, $p = .036$. The difference in the measure of stress was near significant, $Z = -1.916$, $p = .055$. Thus, it was confirmed that the simple training task actually felt easy and that the counterpart complex training task was actually difficult.

6.1.5 Discussion

We discovered that participants actually felt more difficulty on one task than the other for both the Study 1 set and Study 2 set of word-learning tasks that were manipulated in terms of task difficulty. Meanwhile, other perceptual measures varied between these sets of tasks, which we expected because the manipulations in task difficulty were made on different aspects in the different studies. Furthermore, different studies on task complexity have reported statistical significances for different aspects of perceived difficulty. For example, Robinson (2001b) manipulated levels of task difficulty in +/– prior knowledge and found that his participants who gave directions from one point to another on an unfamiliar city map considered the task more difficult and stressful than doing so on a familiar campus map. Sanajou, Zohali, and Zabihi (2017) manipulated +/– dual task components, asking participants to write about stories on cartoon strips, one sequenced correctly and the other disordered. Story writing on a disordered cartoon strip served as "+ dual task" because participants needed to write about the story while thinking about a correct sequence for the scenes. The researchers found that their participants perceived more difficulty and stress in the complex task condition but also found it more interesting than in the simple story condition. The results of those studies and the present results could be comparable because they all have employed the same questionnaire sheet derived from Robinson's (2001b) study. (Table 6.1.5).

Table 6.1.5

Ratings of Perceived Difficulty Measured on the Robinson Questionnaire.

Studies	Task	Statistical Significance				
		D	S	C	I	M
Robinson, 2001[b]	Giving directions on familiar/unfamiliar maps (+/– prior knowledge)	✓	✓			
Sanajou *et al.*, 2017	Story writing on ordered/disordered comic strips (+/– dual task)	✓	✓		✓	
Present Study 1	Translating words in L2-to-L1/L1-to-L2 directions	✓		✓		
Present Study 2	Answering while-reading questions with glosses/without glosses (+/– here-and-now)	✓				

D: Difficulty, S: Stress, C: Confidence, I: Interest, M: Motivation

Overall, 6.1.5 suggests that the questionnaire items other than those on difficulty are relatively unstable, as they can vary depending on what aspects are manipulated. Yet, it is certainly useful and important to measure participants' perceived difficulty regarding tasks by using a questionnaire such as like this one.

6.2 Correlational Analysis of Performance between Study 1 and Study 2 Tasks

6.2.1 Procedure

To look into the similarities and differences between the Study 1 tasks and Study 2 tasks, non-parametric Spearman rank correlation coefficients were calculated on participants' performances for four tests of each study task, namely, performance during learning, as well as accuracy on the immediate retention posttests, delayed retention posttests, and delayed transfer posttests of each study. This time, only the data of participants who participated in both Study 1 tasks and Study 2 tasks in the same ways conceptually were chosen for analysis. Specifically, data on participants who participated in Study 1 tasks in the easy learning condition and did Study 2 tasks in the simple training condition, as well as those who carried out Study 1 tasks in the difficult condition and did the Study 2 tasks in the complex training condition, were selected. This choice was made to extract the differences attributed only to the characteristics of those tasks. As a result of data selection in this way, the sample size for the easy–easy comparison became 13, and that for the difficult–difficult comparison was 9. These unequal sizes of data stemmed from the fact that some outliers were excluded in Study 2 and that they could not be included in this analysis suddenly without a good excuse. Receptive English vocabulary sizes of those qualified participants, obtained in Study 2, were also incorporated into this analysis as a factor of individual differences. We had no hypotheses for this analysis.

6.2.2 Results and Discussion

Overall results are provided in Appendix J due to the limited space here. For both the easy–easy and difficult–difficult correlation tables, strong correlation were seen, but most of them were not significant, probably because of the small sample sizes and the non-parametric test used. Regarding the intercorrelations between the Study 1 tasks and Study 2 tasks, a moderate positive correlation was found at a near significant level (r_s = .55, p = .051) between the learning performance of Study 1 and the immediate recall accuracy of Study 2 in the easy–easy task comparison. An explanation for this similarity would be that both tests involved processes in an L2-to-L1 direction. Learning in the Study 1 easy condition required participants to respond in Japanese from Indonesian cue words. Similarly, recall on the immediate posttest in Study 2 was based on remembering the meanings of nonwords, which can be understood as recall in L1 from L2 cues in terms of a distinction between known and unknown languages. This was the only noticeable correlation in comparisons between the Study 1 and Study 2 tasks. Therefore, these two kinds of tasks constitute different measures of task complexity.

Meanwhile, several correlation values on individual differences were found to be at a significant level. Specifically, the delayed transfer performance in the easy learning condition of Study 1 positively and significantly related to English vocabulary size ($r_s = .86$, $p < .05$). This is an interesting relationship because knowledge of English was not required in the recall of Indonesian words from Japanese cues that the task required. This finding can be explained in terms of the effects of L2 proficiency in learning a third language. The most salient similarity between Indonesian and English is that both languages share the same alphabetic characters (i.e., A–Z). Therefore, it is conceivable that learners with a certain level of English proficiency view Indonesian vocabulary words similar to new English words. That is, existing knowledge of languages can affect the learning of another language both positively and negatively depending on typological similarities between them. The present results are considered to be the former case (i.e. a positive case). Furthermore, such positive transfer occurred only on the delayed transfer task in the easy learning condition of Study 1, even though the difficult learning group also performed the same task during the immediate test and the delayed retention test. This finding could be because recalling Indonesian words as a measure of delayed transfer was the most difficult task among the tasks that required

Indonesian responses in Study 1. It is safe to say that individuals' abilities influenced their performance the most when the task was very difficult. Therefore, this strong correlation between two seemingly unrelated factors (i.e., knowledge of English and Indonesian) reflected a relationship through a secondary factor (i.e., difficulty).

We can see that accuracy on the while-reading questions in the simple training group in Study 2 correlated positively with English vocabulary size at a near significant level (r_s = .55, p =.052). This finding can be explained in the view that the simple training group who read passages with glossaries could access their mental lexicons more easily than those who needed to process nonwords during learning in the complex training condition. Furthermore, a relatively significant correlation was found (r_s = .64, p = .063) between accuracy on the immediate retention test in the complex training group of Study 2 and their English vocabulary sizes, which could be explained in relation to the hypothesized effects of the there-and-then reading condition. Among those who read passages without glossaries, participants with a high level of English proficiency could allocate their cognitive resources to process nonwords, while participants with a low level of English proficiency needed to process both English words and nonwords at equally explicit levels. Consequently, these individual differences led to different levels of processing during training. Therefore, the degrees of subsequent retention differed.

In conclusion, correlational analyses on participants' performances on the Study 1 tasks and Study 2 tasks suggested that these two types of tasks measured different aspects of task difficulty even though manipulations were made on the same conceptual axis, namely, the difficulty or complexity of learning tasks. By including an individual difference factor, English vocabulary size, into these analyses, it was found that the same learner factor affects task performance differently by the task characteristics.

Chapter VII

GENERAL DISCUSSION AND CONCLUSION

7.0 Introduction

In this chapter, the findings of the present research are reported and explored. In addition, the implications for task difficulty research are presented, as well as limitations of the present research and future research directions. Finally, conclusions are drawn from the present findings.

7.1 Summary of the Findings

Motivated by Schneider *et al.*'s (2002) findings, the two studies presented in this paper investigated whether learning in a difficult condition is more effective in the long run than learning in an easy condition within the context of second-language vocabulary acquisition. Hypotheses were consistent across the two studies, as follows. (1) The rates of learning will be greater in an easy learning condition than a difficult learning condition; (2) learning in a difficult condition will promote subsequent recall more than learning in an easy learning condition; and (3) learning in an easy condition will facilitate retention, but learning in a difficult condition promotes transfer. Study 1, a direct replication of Schneider *et al.*'s (2002) work on this topic, uncovered results in line with those hypotheses. In contrast, Study 2, a conceptual replication of Study 1, obtained results inconsistent with the hypotheses. Despite these different results, the primary factor manipulated in both studies was task difficulty, in which one task was easy and the other was difficult, as indicated in the examination of the perceptual difficulty of the present tasks. Therefore, we will now consider how those results could be explained from a theoretical viewpoint.

Result 1: Regarding the learning tasks in Study 1 (a direct replication of the work of Schneider *et al.*), participants in the easy learning condition showed superior rates of learning than those in the difficult learning condition. Therefore, the findings of Schneider *et al.* (2002) were replicated.

In this task, participants in the easy learning condition learned translation word pairs in an L2-to-L1 direction, by responding in Japanese from Indonesian cue words. Participants in the difficult learning condition, in contrast, learned the same translation word pairs in an L1-to-L2 direction, in which they responded in Indonesian based on Japanese cue words. Therefore, the easy learning condition was largely associated with data-driven processing, while the difficult learning condition involved conceptually driven processing to a large extent. Here, we employed the terms data-driven and conceptually driven processing (Jacoby, 1983; Roediger, 1990) because these terms establish a stark contrast between the two types of learning manipulated according to their translation directions.

Data-driven processes depend on perceptual information. In the case of processing written words, for example, data-driven processing starts with participants identifying each letter of a word and combining those letters in order, and the process ends with them accessing a concept of the word (Fujita, 1994). Conceptually driven processes involve accessing a meaning to support task performance. For example, recalling associated words from a stimulus word (e.g., free recall from a word) is a typical conceptually driven process (Fujita, 1994). Human information processing involves both types of processing, but one type dominates most of the time depending on the information type (Donnelly, 1994). We have summarized the types of processes involved in each experiment task on the table in Appendix K.

Learning words in an L2-to-L1 translation direction constitutes a data-driven process because it requires the recognition of unknown words. When we harness unknown L2 words as "data" and known L1 words as "concepts," a question such as "L2 → ?" should be data driven. Furthermore, in the easy learning condition, participants' responses must have been made largely implicitly because they answered in their L1, Japanese (i.e., presumably, no one exerted great efforts to recall the spelling of the word in Japanese). In contrast, learning words in an L1-to-L2 translation direction is considered to be conceptually driven because this process involves the addition of a new concept to a given L1 word. For instance, the experiment item "王(king)–raja" required participants to add a new concept (raja) to the known word (king). Moreover, in this condition, participants had to type in answer words on a computer explicitly because those words were new to them. From these viewpoints, we can draw the following conclusion: On the paired-associate word-learning phase in Study 1, the easy learning group had only one thing to do consciously, namely, to make connections between L2 and L1 words, whereas the difficult learning group had two things to do consciously, that is, to make an association between L1 and L2 words and to remember the spellings of L2 words. Therefore, the easy learning group outperformed the difficult learning group in the rate of learning.

Result 2: During posttests a week after learning in Study 1, we identified a trend that the retention rates of the words of the difficult learning group were greater than those of the easy learning group. The findings of Schneider *et al.* (2002) were thus replicated.

Explanations for this result can be given consistent with the viewpoints above. As mentioned earlier, in the learning phase, the amount of information that participants needed to pay attention to differed between the two groups. The difficult learning group had to process both associations of words and forms of L2 stimuli, while the easy learning group could complete a task by learning L2–L1 associations. Therefore, the difficult learning condition imposed elaborate processing on participants to a greater degree than the easy learning condition did. In other words, the difficult condition made it possible for participants to process more information during the task than the easy learning condition did. Consequently, the difficult learning group recalled more words than the easy learning group a week later, in accordance with the hypothesized effects of the depth-of-processing theory.

However, the present study revealed only a very slight difference between the two groups on subsequent recall in comparison to a larger difference found in the work of Schneider *et al.* (2002). Considering the fact that we employed two languages that were typologically more distant from each other (i.e., Japanese–Indonesian) than those used by Schneider *et al.* (i.e., English–French), it was suggested that prior knowledge might have worked as a moderator variable to the levels of elaboration in processing (cf. Leow, 2015). It is possible that elaboration in processing was at a limited level because there was little relevant information stored in the long-term memory.

Result 3: At posttests a week after learning in Study 1, a trend was found that learning in an easy learning condition led to subsequent retention, but learning in a difficult learning condition led to transfer. This finding was consistent with the trend shown in Schneider *et al.*'s (2002) study.

As we observed in Study 1, the delayed retention accuracy, measured by responding in the same translation direction as in learning, was greater for the easy learning group than the difficult learning group. In contrast, the delayed transfer accuracy, measured by responding in the opposite translation direction to that in the learning process, was greater for the difficult learning group than the easy learning group. This finding constitutes evidence for the hypothesis for the third research question of Study 1. Nevertheless, a counterargument against the hypothesis is possible because we can see a noticeable pattern in which delayed recall responses in an L2-to-L1 direction were always more accurate than those in an L1-to-L2 direction regardless of training conditions (see Table 4.6.3 in Chapter 4). Therefore, the point in question is whether or not the effects of the translation direction influence transfer performances in an unwanted way. However, the answer remains elusive at the present moment because we did not conduct any further investigations on this issue in Study 1 for the sake of conducting a replication of Schneider *et al.*'s (2002) study. We will look at this issue again in the coming section concerning the third research question of Study 2, which did not involve translation direction.

Result 4: On the post-hoc training tasks in Study 2 (conceptual replication of Study 1), participants in the complex training condition showed superior rates of learning compared to those in the simple training condition. This pattern was opposite to that replicated in Study 1.

In Study 2, participants pre-learned associations of unknown words and familiar English words regardless of experimental conditions and then read passages containing the unknown words in either a simple training condition or a complex training condition.

In the simple training condition, participants answered while-reading questions with the assistance of the meanings of those nonwords provided in glossaries in the margins. In the complex training condition, participants answered the while-reading questions without such glossaries. Therefore, in Study 2, all participants used data-driven processing for the initial learning of words, and learned knowledge was trained via a conceptually driven activity. The initial learning of words was considered to be data driven because participants saw nonwords first and then corresponding English words. As those nonwords consisted of unknown L2 words and English definitions for them in terms of familiar L1 words, the format used in the initial learning "L2 → ?" must have encouraged data-driven processes. In contrast, processes required in the post-hoc training were dominantly conceptually driven because the task involved reading passages. Moreover, in Study 2, the task difficulty was manipulated at this phase of learning by providing or withholding definitions of the target nonwords in glossaries. Therefore, generation demands were present in the complex training condition but not in the simple condition (see "generation effect" in Chapter 2). A general assumption of the generation effect is that self-generation demands during learning deepen the level of processing, thus, strong memory traces are produced. We also believe that the processes during the reading of the two experimental groups also differed, as the simple training group employed a mixture of data-driven and conceptually driven processing, whereas the complex training group predominantly involved conceptually driven processing. This difference emerged because relearning target nonwords was actually possible during reading in the simple training condition due to the presence of definitions, but it was not for the complex training group. Therefore, uniformity in the processing types was higher for the complex training group than the simple training group. Put simply, the processes of the simple training group were messy in comparison to the tidier processes of the complex training group.

In this way, it is considered that the superior performance of the complex training group in the post-hoc training tasks was attributed to the uniformity of processes involved. Still, it is interesting to think that participants considered a reading task with a glossary to be easier than one without despite the complex nature of its internal processes.

Results 5: In Study 2, learning gains of the complex training group disappeared soon after learning. The recall probabilities of both experimental groups were almost equal at the time of an immediate posttest. The same trend was also observed during delayed posttests a week later. The results contradicted the trend replicated in Study 1.

The results could be explained in two ways. First, generation demands imposed in a context-based training task (i.e., reading) could result in only limited effects on subsequent retention compared to those initially expected. In fact, beneficial effects on subsequent retention from generation demands that have been reported so far come from results on decontextualized word learning tasks (see "generation effect" in Chapter 2). Researchers sometimes choose this approach to identify the basic effects in purified conditions. In the present study, it was tested in a complex contextualized condition in which the expected effects could be contaminated by other factors. Regardless, this approach does not represent a methodological flaw because the experiment reflects what students actually do in classes. However, although less likely, it is also possible that generation demands in the complex training condition and relearning effects in the simple training condition were equally beneficial for the subsequent retention of learned words. This possibility can be explored in future research.

Second, it is also theoretically conceivable that the effects of the post-hoc reading training were canceled due to a mismatch of study-test overlaps in processing. As mentioned earlier, the initial learning of words was considered to involve data-driven processing, but the post-hoc reading task was predominantly conceptually driven. However, data-driven processing was again required in the immediate recall test. That is, a conceptually driven training task was sandwiched between a data-driven pre-learning task and a data-driven test. The manipulation of task complexity occurred in that training task in between those two events. Similar to the transfer-appropriate processing theory (Bransford *et al.*, 1979; Morris *et al.*, 1977), the framework of data-driven and conceptually driven processing also holds that the similarity in operations between learning and testing influences subsequent recall (Jacoby, 1983; Roediger, 1990). For example, information from data-driven learning is readily accessible in data-driven testing and vice versa. If this was the case, the equal levels of recall on the posttests could be easily explained in terms of their equal levels of initial learning. At the same time, it is possible that training performance would not necessarily be associated with subsequent test performance depending on the degree of similarity between these two processes.

Result 6: When learning and testing took place within the same language (English) in Study 2, as opposed to manipulating the translation directions between two languages (Indonesian–Japanese) in Study 1, the rates of retention were always higher than those of transfer regardless of training complexity. The results rejected the hypothesis that easy learning and difficult learning each contribute uniquely to subsequent recall.

In Study 2, a conceptual replication of Study 1, the transfer of learned word knowledge was assessed via a picture-naming task. Although the task used differed from that in Study 1, it essentially inherited the same concept to assess the transferability of learned words because a picture represented an existing concept (i.e., L1), and responses were made regarding the learned words (i.e., L2), which formed a conceptually driven "L1 ➔ ?" question. In line with a replicated pattern in Study 1, we hypothesized that a simple training task would benefit retention, but a complex training task would promote transfer. However, the same trend was found for both experimental groups: The retention performance was better than the transfer performance. This phenomenon may be explained in the way mentioned in the previous section. That is, the generation effects in the complex training group could be diluted on a context-based training task, or conceptually driven training itself could be ineffective because perception-based, data-driven learning and testing were conducted in advance and afterward, or both.

While inconsistent results were obtained between Study 1 and Study 2 in terms of a distinction between retention and transfer, there was a pattern common to both studies in terms of an "L1–L2" distinction. In the delayed posttests of Study 1, the accuracy of responses was greater for participants who responded in L1 (Japanese) from L2 (Indonesian) cues than those who responded in the opposite direction. Similarly, in Study 2, participants who responded with familiar English words (taken as L1) from nonword (taken as L2) cues showed greater accuracy than those who responded with nonwords (taken as L2) from picture cues (taken as L1). Furthermore, this trend was also common in the findings of Schneider *et al.* (2002). Based on their delayed posttests, responses in L1 (English) from L2 (French) cues outperformed those in L2 from L1 cues in accuracy regardless of training conditions, and they also observed a distinction between retention and transfer. We can simply conclude that performances in which the “forms” of responses were derived from long-term memory were basically advantageous in comparison to those that had to be provided in newly learned forms. Therefore, the superior transfer performances in the difficult learning conditions of both Study 1 and Schneider *et al.*’s (2002) study could be viewed as the result of capitalizing on advantage. This view may also explain Schneider *et al.*’s (2002) mysterious transfer performance in their difficult learning condition. In Chapter 4, we pointed out ~~a~~ complication in which participants in their difficult learning condition recalled more words on a delayed transfer posttest than on an immediate posttest against the principle of forgetting by time. This probably happened because the restrictions in responding in an L2 on the immediate posttest were lost on the delayed posttest, where L1 responses were required.

In Study 2, those undesirable effects from L1 were somewhat reduced because tasks were carried out in monolingual L2 settings (i.e., only English was used). Therefore, the results emphasized the recall of learned items without strong supportive sub-processes, such as those originating in L1 knowledge. Under such a restricted condition, participants in the complex training condition correctly spelled 10% of target nonwords on a picture-naming task, a measure of knowledge transfer, while the transfer rate of the simple training group was 6%. This acquisition rate could be seen as too low or maybe fair depending on one’s perspective. However, this showed that 10% of students in the study achieved a practically desirable use of new words, naming things in a target language, after only one opportunity to study and practice them. By spelling English words for the meaning of the learned nonwords, participants in the simple training condition, who also only had an opportunity to study and practice once, showed that about 40% of the nonwords they had learned were receptively available. These results suggested that learners have difficulty obtaining transferability of newly learned second-language words in a week or so, but they can use that knowledge receptively. Further research on the roles that easy and difficult learning tasks play for subsequent retention and transfer is needed. By revealing the effects of easy tasks in addition to those of difficult tasks, we will be able to see the nature of task difficulty in detail.

7.2 Pedagogical Implications

Typical language classes can be classified into two types depending on whether a teacher shares the same native language with students or not. Therefore, Study 1 represents classes in which translation between a target language and a native language is available. Based on the findings, word learning in an L1-to-L2 direction is desirable in this type of class. When vocabulary acquisition activities are given in class, all students' responses should be made in L2 because, according to the result of the present study, forms (i.e., spellings) of L2 words influence learning gains more than conceptual connections between L1 and L2. Throughout the present research, probabilities of delayed recall of learned words were superior when participants recalled L1 words from L2 cues rather than recalling L2 words from L1 cues regardless of task difficulty. Meanwhile, difficult training tasks (i.e., generating L2 words from L1 cues) enhanced the probabilities of successful delayed word recall in an L1-to-L2 direction, even though the effects were statistically not significant. Therefore, from the viewpoint of how learning can be optimized, activities emphasizing forms of L2 words are recommended.

Study 2 represents language classes conducted by native speakers of a target language, where translation into students' L1 is less likely. Consequently, task difficulty must be implemented by using component factors other than the translation direction in such classes. Study 2 implied that incorporating too many types of activities into a training set might impede optimal learning. As mentioned in the previous chapter, a reason why learning gains in post-learning training were not reflected in performance on posttests in Study 2 might be the mismatch of the processes required between the training task and the posttests. In Study 2, participants performed data-driven rote learning of word associations. Then, they received reading passages for training on learned words that involved conceptually driven processing. Finally, again, data-driven posttests on the word associations were administered. We might intuitively think that these activities were related to each other, but the results indicated that training did not affect participants' subsequent retention. Therefore, the uniformity or similarity of types of processes in a training set might constitute a key issue for optimal learning.

7.3 Limitations of the Present Study

Several limitations can be mentioned in relation to Study 2. First, because an emphasis was on simulating classroom environments, the time allowed for the initial learning of words was not strictly regulated. Participants learned a set of target words at their own pace. Thus, we had to exclude a few participants from the analysis as outliers, making the sample size smaller than initially planned. Moreover, the possibility cannot be ruled out that a self-paced initial learning session might have influenced the strength of the memory traces that participants initially generated. Second, in relation to this idea, the initial learning of words and subsequent practice with them occurred only once, although we could have repeated these phases several times, as in Study 1. Naturally, repeated practice would enhance the effects of learning, in this way, we could most likely detect differences between the different conditions. However, this issue always revolves around a basic question; Exactly how much practices enough? We believed that the clearest justification could only be made with one opportunity. However, other options could work if researchers have a good reason to do so. Finally, because of time constraints, only English vocabulary size was measured as an individual difference factor. By measuring other learner factors such as working memory capacity, we could examine the effects of task difficulty further.

7.4 Future Research

The present research was conducted on the basis of task difficulty, a component factor of the framework for optimal second-language learning. It is not much say that all studies on optimal second-language learning can be traced back to the desirable difficulty framework (Bjork, 1994a, 1994b; Schmidt & Bjork, 1992). The framework states that learners' post-practice retention and transfer become maximized when optimal levels of difficulty exist during practice. Therefore, researchers are now testing the effects of task difficulty by using various component factors of difficulty (e.g., Kasprowicz *et al.*, 2019; Li & Dekeyser, 2019 on distributed practice schedule; Strong & Bores, 2019 on generative learning; Kim & Godfroid, 2019 on modalities of input). The main aim of research is concerned with identifying the beneficial effects of these component factors on the long-term retention of knowledge. However, it is important to identify the suitable level of difficulty for a particular learner because the theory essentially states that retention will be maximized if an "optimal" level of difficulty is provided. Actually, when previous papers were reviewed for the present research, we discovered that this type of issue has rarely been touched on in the literature. Therefore, research on objective indicators of levels of task difficulty matched with learners' proficiency, and other individual factors is needed in parallel with the current line of research.

7.5 Conclusion

The present findings taken together, the effects of task difficulty on the retention and transfer of newly learned words become evident under a condition in which the learner can make use of existing knowledge for input processing to be elaborated. Study 1, in which a participant's L1 was involved in input processing, represents an extreme case of this phenomenon. Study 2, in which the entire learning process occurred within a participant's developing L2, constitutes a minor case. The differences between the two cases result from the different degrees of strength of this knowledge. The results also suggest that successful orthographic responses with newly learned words are the keys to improving test performances, especially regarding transfer.

Overall, the present research supports the position of the training difficulty principle because no strong negative evidence against the theory was found. If the goal of second-language acquisition consists of developing sustainable knowledge that can also be adapted to a new context, training in difficult learning conditions is recommended, as Schneider *et al.* (2002) mentioned. However, it might also be true that the beneficial effects of task difficulty on later retention and transfer are extremely small on factors other than the translation direction. Therefore, in this particular field of research, it is important to develop further factors of task difficulty that are commonly effective in various second-language learning situations.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control process. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation Vol. 2*. pp.89-195. New York: Academic Press.

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

Anderson, J. R. (2009). *Cognitive Psychology and Its Implications*. New York: Worth Publishers.

Baddeley, A. D. (1966). Short-term Memory for Word Sequences as a Function of Acoustic, Semantic, and Formal Similarity. *Quarterly Journal of Experimental Psychology, 18*, 362-365.

Baddeley, A. D. (1978). The Trouble with Levels: A Reexamination of Craik and Lockhart framework for memory research. *Psychological Review, 85*, 139-152.

Baddeley, A. (1979). Levels of Processing and Levels of Explanation: Discussion of the Papers by Treisman and Bransford, Franks, Morris, and Stein. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory.* pp.355-360. New Jersey: Lawrence Erlbaum Associates, Inc.

Baddeley, A. D. (1990). *Human Memory: Theory and Practice*. Massachusetts: Allyn and Bacon.

Baddeley, A. (2000). Short-Term and Working Memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*. pp.77-92. New York: Oxford University Press.

Baddeley, A. D., Grant, S., Wight, E., & Thomson, N. (1975). Imagery and Visual Working Memory. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and*

Performance, V. pp.205-217. London: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation*. pp.47-89. New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (2017). Is the Levels of Processing Effect Language-limited? *Journal of Memory and Language, 92*, 1-13.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., and Bahrick, P. E. (1993). Maintenance of Foreign Language Vocabulary and the Spacing Effect. *Psychological Science, 4*(5), 316-321.

Battig, W. F. (1972). Intratask Interference as a Source of Facilitation in Transfer and Retention. In R. F. Thompson & J. F. Voss (Eds.), *Topics in Learning and Performance*. pp.131-159. New York: Academic Press.

Baugh, A., & Cable, T. (2013). *A History of the English Language (Sixth Edition)*. London: Routledge.

Bird, S. (2010). Effects of Distributed Practice on the Acquisition of Second Language English Syntax. *Applied Psycholinguistics, 31*, 635-650.

Bjork, R. A. (1994a). Memory and Metamemory Considerations in the Training of Human Beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing*. pp.185-205. Cambridge: MIT Press.

Bjork, R. A. (1994b). Epilogue: Institutional Impediments to Effective Training. In D. Druckman & R. A. Bjork (Eds.), *Learning, Remembering, Believing: Enhancing Human Performance*. pp.295-306. Washington D.C.: National Academy Press.

Bjork, R. A. (1999). Assessing Our Own Competence: Heuristics and Illusions. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII*. pp.435-459. Cambridge: MIT Press.

Bjork, R. A., & Kroll, J. F. (2015). Desirable Difficulties in Vocabulary Learning.

American Journal of Psychology, 128(2), 241-252.

Bloom K. C., & Shuell, T. J. (1981). Effects of Massed and Distributed Practice on the Learning and Retention of Second-Language Vocabulary. *The Journal of Educational Research, 74*(4), 245-248.

Bourne, L. E. & Healy, A. F. (2012). 1. Introduction: Training and Its Cognitive Underpinnings. In A. F. Healy & L. E. Bourne (Eds.), *Training Cognition: Optimizing Efficiency, Durability, and Generalizability.* New York: Psychology Press.

Bower, G. (1977). Commentary on “The Multicomponent Theory of the Memory Trace.” In G. Bower (Ed.), *Human Memory: Basic Processes*. pp.437-443. New York: Academic Press.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some General Constraints on Learning and Memory Research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory.* pp.331-354. New Jersey: Lawrence Erlbaum Associates, Inc.

Brooks, L. R. (1967). The Suppression of Visualization by Reading. *The Quarterly Journal of Experimental Psychology, 19*, 289-299.

Bull, L. B., & Wittrock, M. C. (1973). Imagery in the Learning of Verbal Definitions. *The British Journal of Education Psychology, 43*(3), 289-293.

Carlson, R. A., & Yaure, R. G. (1990). Practice Schedules and the Use of Component Skills in Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 484-496.

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., and Pashler, H. (2009). Optimizing Distributed Practice: Theoretical Analysis and Practical Implications. *Experimental Psychology, 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), 354-380.

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing Effects in Learning: A Temporal Ridgeline of Optimal Retention. *Psychological Science, 19*(11), 1095-1102.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.

Chase, W. G., & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology, 4*, 55-81.

Chujo, K., & Oghigian, K. (2009). How Many Words Do You Need to Know to Understand TOEIC, TOEFL & EIKEN?: An Examination of Text Coverage and High Frequency Vocabulary. *The Journal of Asia TEFL, 6*(2), 121-148.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*(1), 155-159.

Conard, R. (1964). Acoustic Confusions in Immediate memory. *British Journal of Psychology, 55*, 75-84.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of Processing: A Framework for Memory Research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of Processing and the Retention of Words in Episodic Memory. *Journal of Experimental Psychology: General, 104*(3), 268-294.

Crossman, E. R. F. W. (1959). A Theory of the Acquisition of Speed-Skill. *Ergonomics, 2*(2), 153-166.

Crutcher, R. J., & Healy, A. F. (1989). Cognitive Operations and the Generation Effect.

Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(4), 669-675.

Cull, W. L., Shaughnessy, J. J., & Zechmeister, E. B. (1996). Expanding Understanding of the Expanding-Pattern-of-Retrieval Mnemonic: Toward Confidence in Applicability. *Journal of Experimental Psychology: Applied, 2*(4), 365-378.

de Groot, A. D. (1965). *Thought and Choice in Chess*. The Hague, Netherlands: Mouton.

DeKeyser, R. M. (2001). Automaticity and Automatization. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. pp.125-151. Cambridge: Cambridge University Press.

Diamond, R., & Rozin, P. (1984). Activation of Existing Memories in Anterograde Amnesia. *Journal of Abnormal Psychology, 93*(1), 98-105.

Dolan, T. P. (1999). *A Dictionary of Hiberno-English: The Irish Use of English*. Dublin: Gill & Macmillan.

Donnelly, C. (1994). *Linguistics for Writers*. New York: State University of New York Press.

Dougherty, K. M., & Johnston, J. M. (1996). Overlearning, Fluency, and Automaticity. *The Behavior Analyst, 19*(2), 289-292.

Driskell, J. E., Willis, R. P., & Copper, C. (1992). Effect of Overlearning on Retention. *Journal of Applied Psychology, 77*(5), 615-622.

Ebbinghaus, H. (1885). *Memory: A Contribution to Experimental Psychology*. Ruger, H. A., & Bussenius, C. E. (Trans. in 1913). New York: Teachers College Columbia University.

Ethnologue. Retrieved on January 30, 2020 at [https://web.archive.org./web/20020202042447/http://www.ethnologue.com/show_language.asp?code=ENG]

Ericsson, K. A. (2006). The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance*. pp.683-703. New York: Cambridge University Press.

Ericsson, K. A., Krampe, R., & Tesch-Romer, C. (1993). The Role of Deliberate Practice in the Acquisition of Expert Performance. *Psychological Review, 100*(3), 363-406.

Fay, S., Isingrini, M., & Clarys, D. (2005). Effects of Depth-of-Processing and Aging on Word-Stem and Word-Fragment Implicit Memory Tasks: Test of the Lexical Processing Hypothesis. *European Journal of Cognitive Psychology, 17*(6), 785-802.

Fitts, P. M. (1964). Perceptual-Motor Skill Learning. In A. W. Melton (Ed.), *Categories of Human Learning*. pp.243-285. New York: Academic Press.

Fitts, P. M., & Posner, M. I. (1967). *Human Performance*. California: Wadsworth Publishing.

Folse, K. S. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. TESOL Quarterly, 40(2), 273-293.

Foster, J. K. (2009). *Memory: A Very Short Introduction*. New York: Oxford University Press.

Franklin, S., Turner, J., Lambon Ralph, M. A., Morris, J., & Bailey, P. J. (1996). A Distinctive Case of Word Meaning Deafness? *Cognitive Neuropsychology, 13*(8), 1139-1162.

Fujita, T. (1994). Word Fragment Completion Task in the Implicit Memory Research. *Japanese Psychological Review, 37*(1), 72-91.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working Memory and Language*. Hove, England: Psychology Press.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. New Jersey: Lawrence Erlbaum Associates.

Graf, P., & Schacter, D. L. (1985). Implicit and Explicit Memory for New Associations in Normal and Amnesic Subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501-518.

Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming Across Modalities and Priming Across Category Levels: Extending the Domain of Preserved Function in Amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(2), 386-396.

Graf, P., Squire, L. R., & Mandler, G. (1984). The Information That Amnesic Patients Do Not Forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(1), 164-178.

Healy, A. F., Fendrich, D. W., Crutcher, R. J., Wittman, W. T., Gesi, A. T., Ericsson, K. A., & Bourne, L. E. (1992). The Long-Term Retention of Skills. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From Learning Process to Cognitive Process: Essays in Honor of William K. Estes Volume 2*. pp.87-118. New Jersey: Lawrence Erlbaum Associates.

Healy, A. F., Schneider, V. I., & Bourne, L. E. (2012). Empirically Valid Principles of Training. In A. F. Healy & L. E. Bourne Jr. (Eds.,) *Training Cognition: Optimizing Efficiency, Durability, and Generalizability.* pp.13-39. New York: Psychology Press.

Hintzman, D. L. (1974). Theoretical Implications of the Spacing Effect. In R. L. Solso (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium.* pp.77-99. Maryland: Lawrence Erlbaum Associates, Inc.

Hintzman, D. L., Block, R. A., & Summers, J. J. (1973). Modality Tags and Memory for Repetitions: Locus of the Spacing Effect. *Journal of Verbal Learning and Verbal Behavior, 12*(3), 229-238.

Hu, H. M., & Nassaji, H. (2016). Effective Vocabulary Learning Tasks: Involvement Load Hypothesis Versus Technique Feature Analysis. *System, 56*, 28-39.

Hulstijn, J., & Laufer, B. (2001). Some Empirical Evidence for the Involvement Load Hypothesis in Vocabulary Acquisition. *Language Learning, 51*(3), 539-558.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for Words as a Function of Semantic, Graphic, and Syntactic Orienting Tasks. *Journal of Verbal Learning and Verbal Behavior, 12*, 471-480.

Jackson, D.O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A Synthesis and Meta-Analysis of Research on Second Language Task Complexity. *Language Learning, 63*(2), 330-367.

Jacoby, L. L. (1978). On Interpreting the Effects of Repetition: Solving a Problem Versus Remembering a Solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649-667.

Jacoby, L. L. (1983). Remembering the Data: Analyzing Interactive Processes in Reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 485-508.

Jacoby, L.L., & Craik, F. I. M. (1979). Effects of Elaboration of Processing at Encoding and Retrieval: Trace Distinctiveness and Recovery of Initial Context. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory.* pp.1-21. New Jersey: Lawrence Erlbaum Associates, Inc.

Jing, L., & Jianbin, H. (2009). An Emprical Study of the Involvement Load Hypothesis in Incidental Vocabulary Acquisition in EFL Listening. *Polyglossia, 16*, 1-11.

Kane, J. H., & Anderson, R. C. (1978). Depth of Processing and Interference Effects in the Learning and Remembering of Sentences. *Journal of Educational Psychology, 70*(4), 626-635.

Karpicke, J. D., & Roediger, H. L. (2007). Expanding Retrieval Practice Promotes Short-Term Retention, but Equally Spaced Retrieval Enhances Long-Term Retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 704-719.

Kasprowicz, R. E., Marsden, E., & Sephton, N. (2019). Investigating Distribution of Practice Effects for the Learning of Foreign Language Verb Morphology in the Young Learner Classroom. *The Modern Language Journal, 103*(3), 580-606.

Keating, G. D. (2008). Task Effectiveness and Word Learning in a Second Language: The Involvement Load Hypothesis on Trial. *Language Teaching Research, 12*(3), 365-386.

Kim, J. W., Ritter, F. E, & Koubek, R. J. (2013). An Integrated Theory for Improved Skill Acquisition and Retention in the Three Stages of Learning. *Theoretical Issues in Ergonomics Science, 14*(1), 22-37.

Kim, K. M., & Godfroid, A. (2019). Should We Listen or Read? Modality Effects in Implicit and Explicit Knowledge. *The Modern Language Journal, 103*(3), 648-664.

Kim, Y. (2008). The Role of Task-Induced Involvement and Learner Proficiency in L2 Vocabulary Acquisition. *Language Learning, 58*(2), 285-325.

Kohn, S. E., & Friedman, R. B. (1985). Word-Meaning Deafness: A Phonological-Semantic Dissociation. *Cognitive Neuropsychology, 3*(3), 291-308.

Kole, J. A., & Healy, A. F. (2007). Using Prior Knowledge to Minimize Interference When Learning Large Amounts of Information. *Memory & Cognition, 35*(1), 124-137.

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science, 19*(6), 585-592.

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language, 33*(2), 149-174.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. (1970). Short-term Memory While Shadowing: Recall of Visually and Aurally Presented Letters. *Journal of Experimental Psychology, 85*, 220-224.

Kuppens, A. H. (2010). Incidental Foreign Language Acquisition from Media Exposure. Learning, Media and Technology, 35(1), 65-85.

Laufer, B, & Hulstijn, J. (2001). Incidental Vocabulary Acquisition in a Second Language: The Construct of Task-Induced Involvement. *Applied Linguistics, 22*(1), 1-26.

Leow, R. P. (2015). Implicit Learning in SLA: Of Processes and Products. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages. Studies in Bilingualism 48*. pp.47-67. Amsterdam: John Benjamins.

Li, M., & Dekeyser, R. (2019). Distribution of Practice Effects in the Acquisition and Retention of L2 Mandarin Tonal Word Production. *The Modern Language Journal, 103*(3), 607-628.

Lightbown, P. M. (2013). Transfer appropriate processing. In P. Robinson (Ed.), *The Routledge Encyclopedia of Second Language Acquisition.* pp.652-655. New York: Routledge.

Lwin, M. O., Morrin, M., & Krishna, A. (2010). Exploring the Superadditive Effects of Scent and Pictures on Verbal Recall: An Extension of Dual Coding Theory. *Journal of Consumer Psychology, 20*, 317-326.

MacLeod, C. M. (1998). Directed Forgetting. In J. M. Golding, & MacLeod, C. M. (Eds.,) *Intentional Forgetting: Interdisciplinary Approaches.* pp.1-57. New Jersey: Lawrence Erlbaum Associates.

MacLeod, C. M., & Daniels, K. A. (2000). Direct Versus Indirect Tests of Memory: Directed Forgetting Meets the Generation Effect. *Psychonomic Bulletin & Review, 7*(2), 354-359.

McDaniel, M. (2012). Put the SPRINT in Knowledge Training. In A. F. Healy & L. E. Bourne Jr. (Eds.,) *Training Cognition: Optimizing Efficiency, Durability, and Generalizability.* pp.267-286. New York: Psychology Press.

McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The Read-Recite-Review Study Strategy: Effective and Portable. *Psychological Science, 20*(4), 516-522.

Meara, P. (2009). *Connected Words: Word associations and second language vocabulary acquisition.* Amsterdam: John Benjamins.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review, 63*, 81-97.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of Processing Versus Transfer Appropriate Processing. *Journal of Verbal Learning and Behavior, 16*, 519-533.

Morris, P. E., Gruneberg, M. M., Sykes, R. N., & Merrick, A. (1981). Football Knowledge and the Acquisition of New Results. *British Journal of Psychology, 72*, 479-483.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle & Heinle.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (online). *Vocabulary Size Test (bilingual Japanese version)*. Retrieved from https://www.victoria.ac.nz/lals/about/staff/Publications/paul-nation/Vocab_Size_Test_Japanese.pdf on October 18, 2018.

Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Heinle, Cengage Learning.

Norman, D. A. (1988). *The Psychology of Everyday Things.* New York: Basic Books, Inc.

Paivio, A. (1971). *Imagery and Verbal Processes*. New York: Holt, Rinehart and Winston, Inc.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. New York: Oxford University Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, Imagery, and Meaningfulness Values for 925 Nouns. *Journal of Experimental Psychology Monograph Supplement, 76, No.1, Part 2,* 1-25.

Rabley, S. (1998a). *Tinkers Island*. Penguin Readers Easystarts. Harlow: Addison Wesley Longman.

Rabley, S. (1998b). *Flying Home.* Penguin Readers Easystarts. Harlow: Addison Wesley Longman.

Read, J. (1988). Measuring the Vocabulary Knowledge of Second Language Learners. *RELC Journal, 19*(2), 12-25.

Revesz, A. (2009). Task Complexity, Focus on Form, and Second Language Development. *Studies of Second Language Acquisition, 31*, 437-470.

Reitman, J. S. (1976). Skilled Perception in Go: Deducing Memory Structures from Inter-Response Times. *Cognitive Psychology, 8*, 336-356.

Robinson, P. (1995). Task Complexity and Second Language Narrative Discourse. *Language Learning, 45*(1), 99-140.

Robinson, P. (2001a). Task Complexity, Cognitive Resource, and Syllabus Design: A Triadic Framework for Examining Task Influences on SLA. In P. Robinson (Ed.,) *Cognition and Second Language Instruction*. pp.287-318. Cambridge: Cambridge University Press.

Robinson, P. (2001b). Task Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework. *Applied Linguistics, 22*(1), 27-57.

Robinson, P. (2003). The Cognition Hypothesis, Task Design, and Adult Task-based Language Learning. *Second Language Studies, 21*(3), 45-105.

Robinson, P. (2007). Criteria for Classifying and Sequencing Pedagogic Tasks. In M. P. Garcia Mayo (Ed.,) *Investigating Tasks in Formal Language Learning*. pp.7-27. Clevedon: Multilingual Matters.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Time and Motion: Measuring the Effects of the Conceptual Demands of Tasks on Second Language Speech Production. *Applied Linguistics, 30*(4), 533-554.

Roediger, H. L. (1990). Implicit Memory: Retention Without Remembering. *American Psychologist, 45*(9), 1043-1056.

Roediger, H. L. (2008). Relativity of Remembering: Why the Laws of Memory Vanished. *Annual Review of Psychology, 59*, 225-254.

Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). Processing Approaches to Cognition: The Impetus from the Levels-of-Processing Framework. *Memory, 10*(5), 319-332.

Rohrer, D., Tylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The Effect of Overlearning on Long-Term Retention. *Applied Cognitive Psychology, 1*9, 361-374.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). Impact of Concreteness on

Comprehensibility, Interest, and Memory for Text: Implications for Dual Coding Theory and Text Design. *Journal of Educational Psychology, 85*(2), 291-304.

Sadoski, M., & Paivio, A. (2001). *Imagery and Text: A Dual Coding Theory of Reading and Writing*. New Jersey: Lawrence Erlbaum Associations.

Sanajou, N, Zohali, L., & Zabihi, F. (2017). Do Task Complexity Demands Influence the Learners' Perception of Task Difficulty? *International Journal of Applied Linguistics & English Literature, 6*(6), 71-77.

Sartori, G., Masterson, J., & Job, R. (1987). Direct-Route Reading and the Locus of Lexical Decision. In M. Coltheart, G. Strtori, & R. Job (Eds.), *The Cognitive Neuropsychology of Language*. pp.59-78. London: Erlbaum.

Schacter, D. L. (1987). Implicit Memory: History and Current Status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 501-518.

Schacter, D. L. (1992). Priming and Multiple Memory Systems: Perceptual Mechanisms of Implicit Memory. *Journal of Cognitive Neuroscience, 4*(3), 244-256.

Schacter, D. L., & Church, B. A. (1992). Auditory Priming: Implicit and Explicit Memory for Words and Voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 915-930.

Schacter, D. L., Cooper, L. A, Delaney, S. M., Peterson, M. A, & Tharan, M. (1991). Implicit Memory for Possible and Impossible Objects: Constraints on the Construction of Structural Descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(1), 3-19.

Schacter, D. L., Wagner, A. D., & Buckner, R, L. (2000). Memory Systems of 1999. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*. pp.627-643. New York: Oxford University Press.

Schmidt, R. A., & Bjork, R. A. (1992). New Conceptualizations of Practice: Paradigms

Suggest New Concepts for Training. *Psychological Science, 3*(4), 207-217.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E. (1995). The Effects of Contextual Interference on the Acquisition and Retention of Logical Rules. In A. F. Healy & L. E. Bourne (Eds.), *Learning and Memory of Knowledge and Skills: Durability and Specificity*. pp.95-131. California: Sage Publications.]

Schneider, V. I., Healy, A. F., & Bourne, L. E. (1998). Contextual Interference Effects in Foreign Vocabulary Acquisition and Retention. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. pp.77-90. New Jersey: Lawrence Erlbaum Associates.

Schneider, V. I., Healy, A. F, & Bourne, L. E. (2002). What Is Learned under Difficult Conditions Is Hard to Forget: Contextual interference Effects in Foreign Vocabulary Acquisition, Retention, and Transfer. *Journal of Memory and Language, 46*, 419-440.

Seamon, J. G, & Murray, P. (1976). Depth of Processing in Recall and Recognition Memory: Differential Effects of Stimulus Meaningfulness and Serial Position. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 680-687.

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

Shimamura, A. P. (1986). Priming Effects in Amnesia: Evidence for a Dissociable Memory Function. *The Quarterly Journal of Experimental Psychology, 38A*, 619-644.

Shimamura, A. P., & Squire, L. R. (1984). Paired-Associate Learning and Priming Effects in Amnesia: A Neuropsychological Study. *Journal of Experimental*

Psychology: General, 113(4), 556-570.

Shimamura, A. P., & Squire, L. R. (1987). A Neuropsychological Study of Fact Memory and Source Amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 464-473.

Shulman, H. G. (1972). Semantic Confusion Errors in Short-term Memory. *Journal of Verbal Learning and Verbal Behavior, 11*(2), 221-227.

Slamecka, N. J., & Graf, P. (1978). The Generation Effect: Delineation of a Phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

Squire, L. R. (1987). *Memory and Brain*. New York: Oxford University Press.

Squire, L. R. (2004). Memory Systems of the Brain: A Brief History and Current Perspective. *Neurobiology of Learning and Memory, 82*, 171-177.

Squire, L. R. & Zola-Morgan, S. (1991). The Medial Temporal Lobe Memory System. *Science, 253*, 1380-1386.

Strong, B., & Bores, F. (2019). Weighing Up Exercises on Phrasal Verbs: Retrieval Versus Trial-and-Error Practices. *The Modern Language Journal, 103*(3), 562-579.

Surprenant, A. M., & Neath, I. (2009). *Principles of Memory: Essays in Cognitive Psychology.* New York: Psychology Press.

Taylor, I. (1976). Similarity Between French and English Words: A Factor to Be Considered in Bilingual Language Behavior. *Journal of Psycholinguistic Research, 5*(1), 85-94.

Thorndike, E. L. (1912). The curve of work. *Psychological Review, 19*, 165-194.

Tulving, E. (1983). *Elements of Episodic Memory.* New York: Oxford University Press.

Tulving, E. (1984). Precis of Elements of Episodic Memory. *The Behavioral and Brain*

Sciences, 7, 223-28.

Tulving, E. (1985). How Many Memory Systems Are There? *American Psychologist, 40*, 385-398.

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology, 53*, 1-25.

Tulving, E., & Thomson, D. M. (1973). Encoding Specificity and Retrieval Process in Episodic Memory. *Psychological Review, 80*(5), 352-373.

van Dijk, T. A., & Kihtsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

van Overschelde, J. P., & Healy, A. F. (2001). Learning of Nondomain Facts in High- and Low-Knowledge Domains. *Journal of Experimental Psychology, 27*(5), 1160-1171.

vanPatten, B. (1996). *Input Processing and Grammar Instruction: Theory and Research*. New Jersey: Ablex.

Waugh, N. C., & Norman, D.A. (1965). Primary Memory. *Psychological Review, 72,* 89-104.

Webb, S. (2005). Receptive and Productive Vocabulary Learning: The Effects of Reading and Writing on Word Knowledge. *Studies in Second Language Acquisition, 27*, 33-52.

Wheeler, M. A. (2000). Episodic Memory and Autonoetic Awareness. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory*. pp.597-608. New York: Oxford University Press.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation Models in Language Comprehension and Memory. *Psychological Bulletin, 123*(2), 162-185.

Appendix A

Consent Forms for Participation

実験参加者の方へ

第二言語習得研究に関する実験説明書

研究の目的

本研究は、異なる学習プロセスにおける第二言語語彙習得の可能性について調査することを目的としています。

実験方法

コンピュータ画面上に提示される問題に回答していただきます。実験セッションは約１週間の間隔をおいて合計３回実施します。各セッションはそれぞれ３０分～１時間程度です。

個人情報とデータの取り扱い

取得したデータおよび個人情報は、研究目的以外には使用しません。データには番号付けを行うことにより匿名化しますので、学会や学術誌などを通じて研究発表をする際も個人情報は守秘されます。データの保存には万全を期し外部へ漏えいすることはありません。

被験者の権利

この実験への参加は、被験者の自由意思によってのみ行われます。また、一度同意した後でいつでも同意を取り消すことができ、それによる不利益はありません。その場合、匿名化番号を破棄するとともに、関連するデータの一切を破棄し、その後の研究には使用いたしません。ただし、取り消し要求があった時点で公表済みの解析結果がある場合は、データを破棄できないことがありますのでご承知おきください。

実験に参加することによる利益および不利益

本研究に参加するための被験者自身の費用負担はありません。また、参加されなくても不利益を受けることは一切ありません。被験者が在学生の場合でも、履修科目の成績評価には一切関係がありません。

謝礼の支払い

実験に参加いただいた方には、謝礼金として１セッションにつき１，０００円を現金にてお支払いします。（土日の場合は１，５００円）

知的財産権

本研究に関する知的財産権の帰属は研究者及び大学の協議によって決定され、データ提供者に帰属することはありません。

（実験責任者）

青山学院大学大学院文学研究科英米文学専攻博士後期課程

田邉　正行 email: d1314002@aoyama.jp

Appendix A (Continued)

参加同意書

（第二言語習得研究に関する実験）

≪説明を受けた項目≫

- ☐ 研究の目的
- ☐ 実験方法
- ☐ 個人情報とデータの取り扱い
- ☐ 被験者の権利
- ☐ 実験に参加することによる利益および不利益
- ☐ 謝礼の支払い
- ☐ 知的財産権
- ☐ 実験者の氏名、所属

説明実施：　　　　年　　月　　日

説明者：田 邉　正 行

私は、以上の説明を理解し、本実験に参加することに同意します。

年　　月　　日

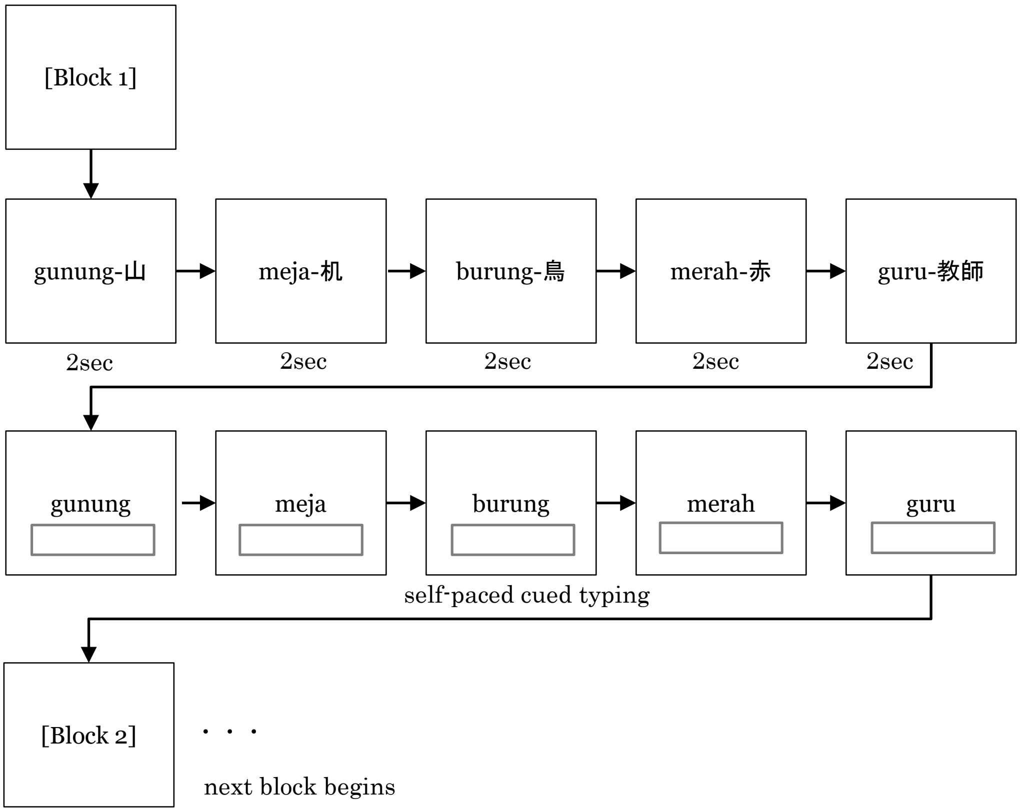
署　名：＿＿＿＿＿＿＿＿＿＿

●本研究へのご理解とご協力に深く感謝いたします●

Appendix B

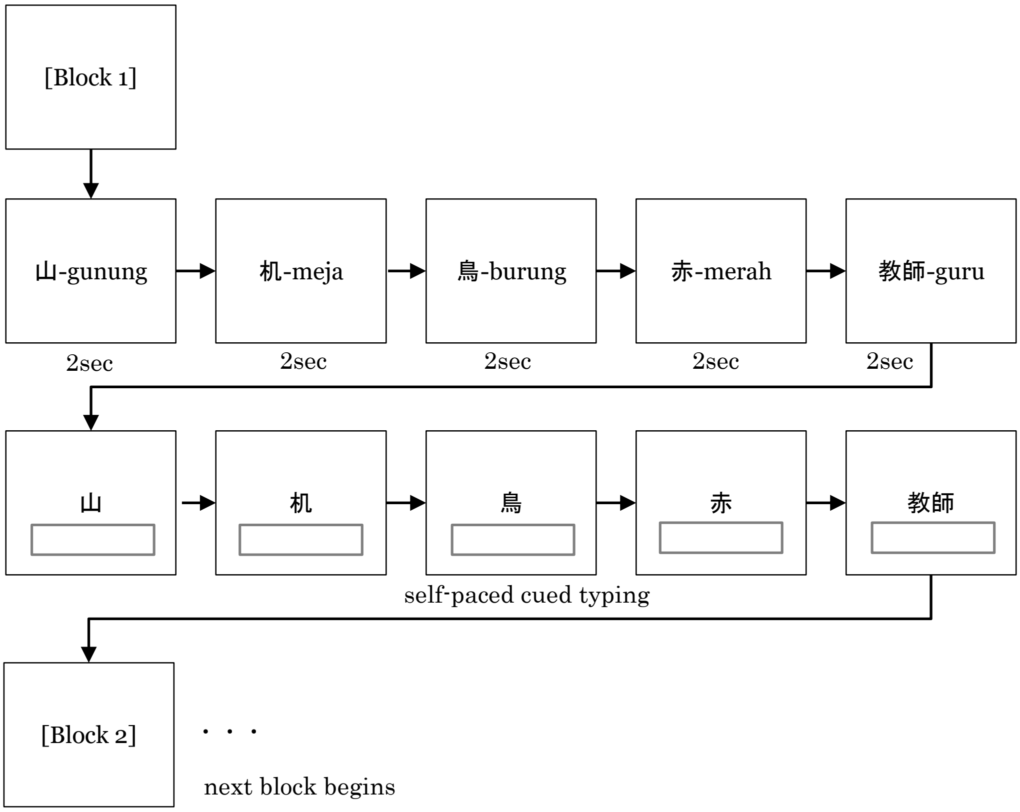
A Flowchart of a Given Learning Block in Study1 Learning Tasks

Easy Learning Condition



Appendix B (Continued)

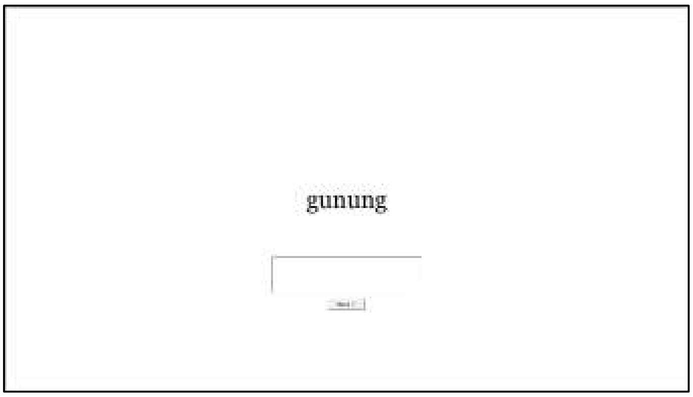
Difficult Learning Condition



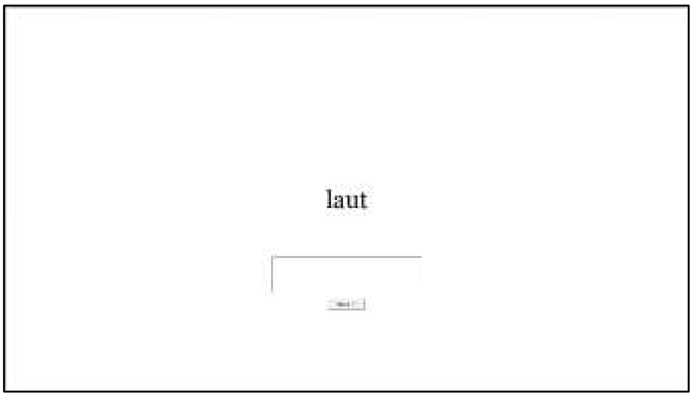
Appendix C

All Question Items on the “L2 cue–L1 response” Retention Posttest in Study1

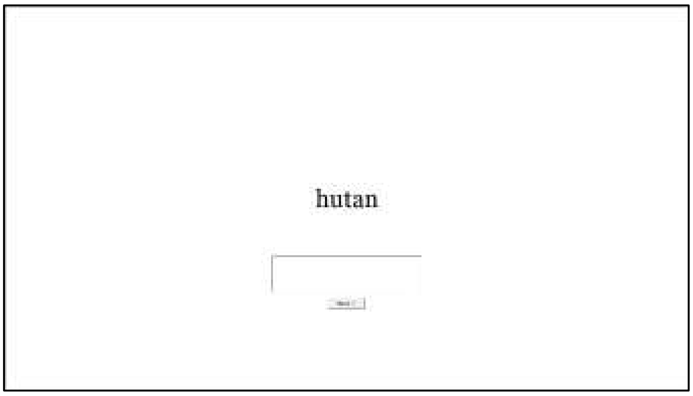
An Indonesian cue for 山 (mountain) *やま(Hiragana form) was also acceptable.



An Indonesian cue for 海 (sea) *うみ(Hiragana form) was also acceptable.

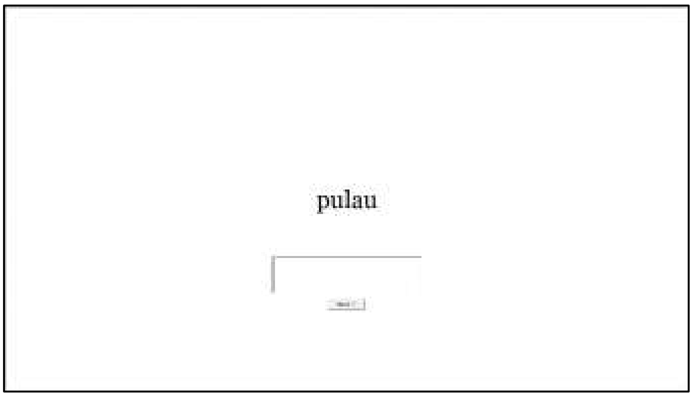


An Indonesian cue for 森 (forest) *もり(Hiragana form) was also acceptable.

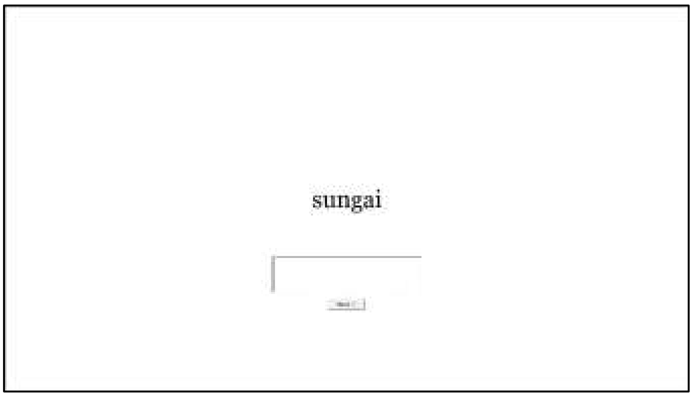


Appendix C (continued)

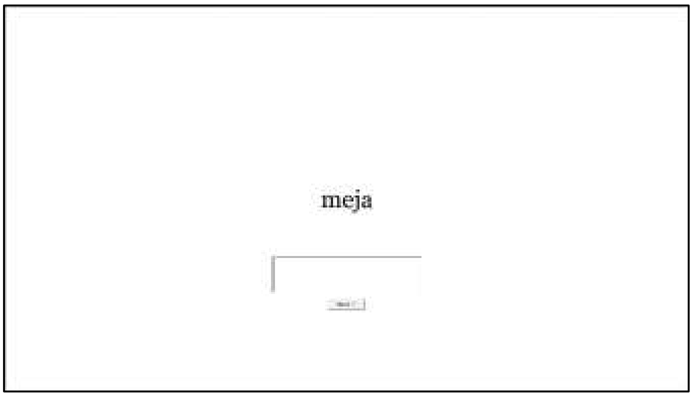
An Indonesian cue for 島 (island) *しま(Hiragana form) was also acceptable.



An Indonesian cue for 川 (river) *かわ(Hiragana form) was also acceptable.

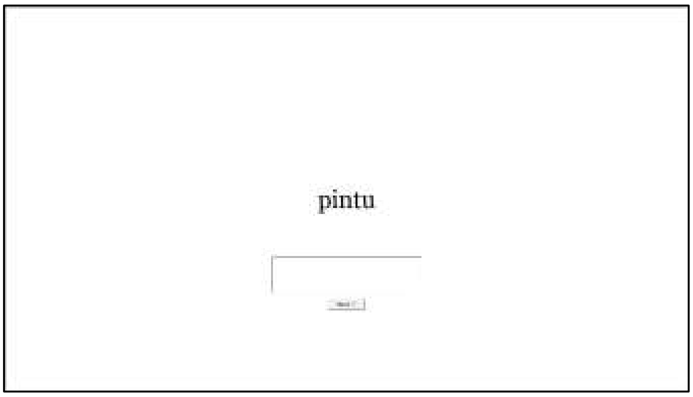


An Indonesian cue for 机 (table) *つくえ(Hiragana form) was also acceptable.

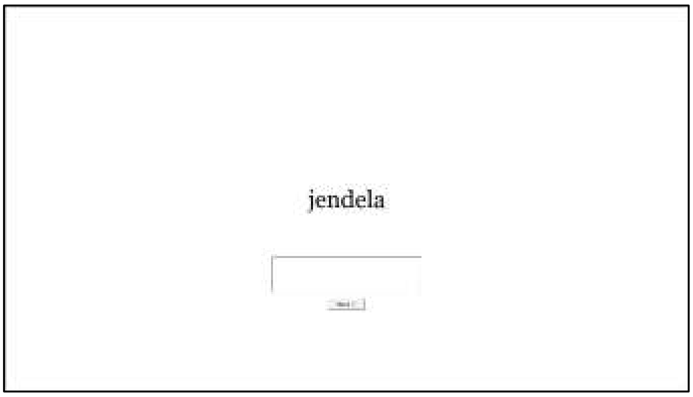


Appendix C (continued)

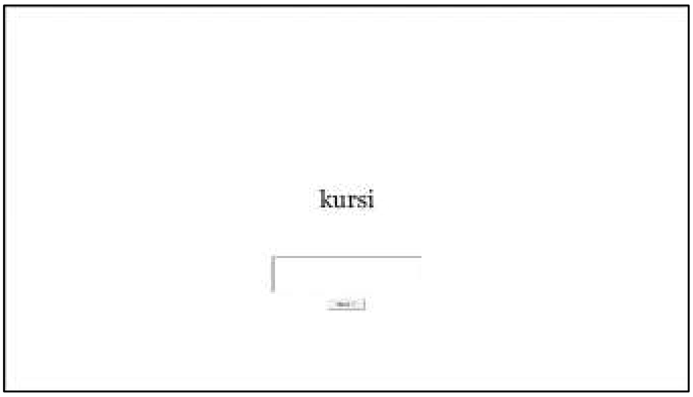
An Indonesian cue for ドア (door)　*どあ(Hiragana form) was also acceptable.



An Indonesian cue for 窓 (window)　*まど(Hiragana form) was also acceptable.

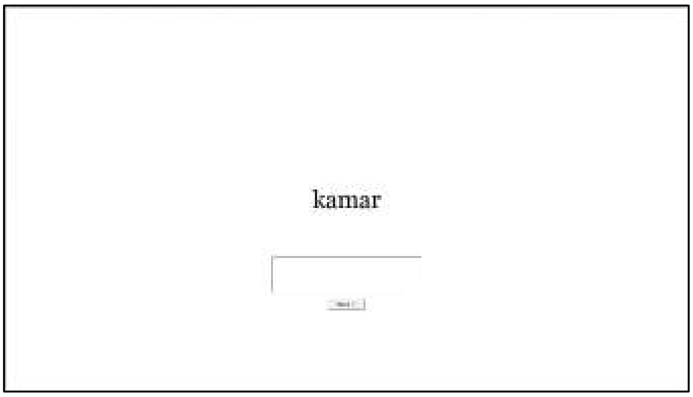


An Indonesian cue for いす (chair)　*椅子(Kanji form) was also acceptable.

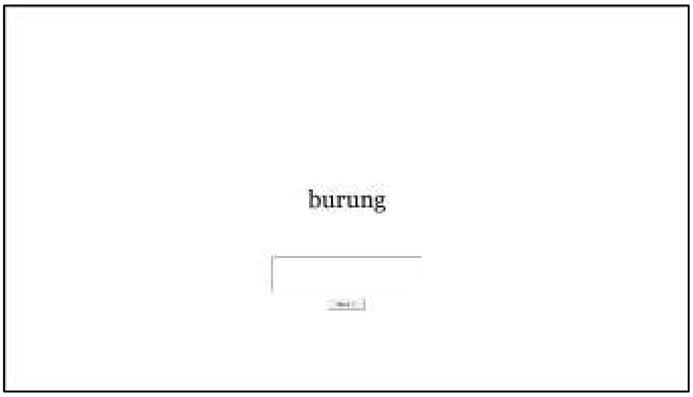


Appendix C (continued)

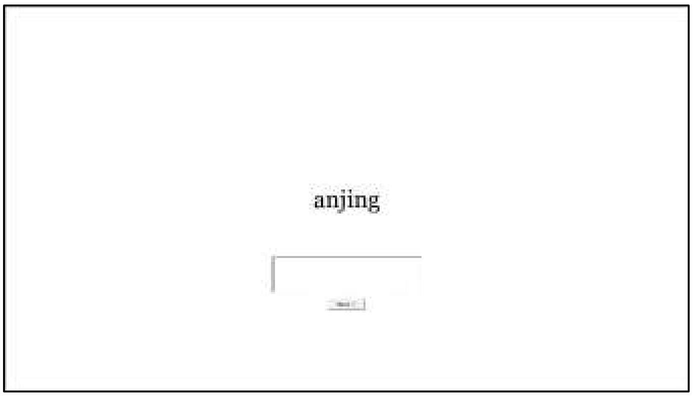
An Indonesian cue for 部屋 (room) *へや(Hiragana form) was also acceptable.



An Indonesian cue for 鳥 (bird) *とり(Hiragana form) was also acceptable.

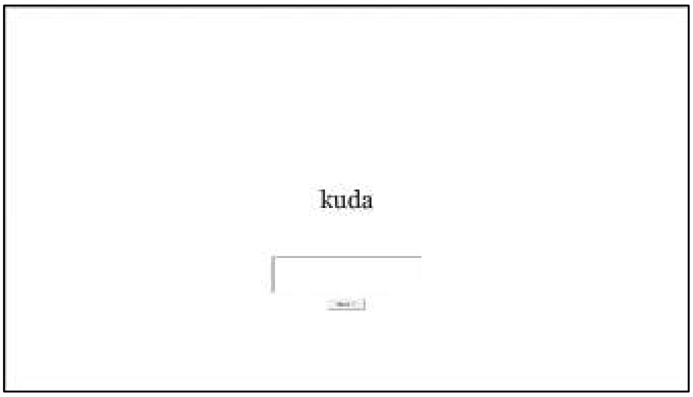


An Indonesian cue for 犬 (dog) *いぬ(Hiragana form) was also acceptable.

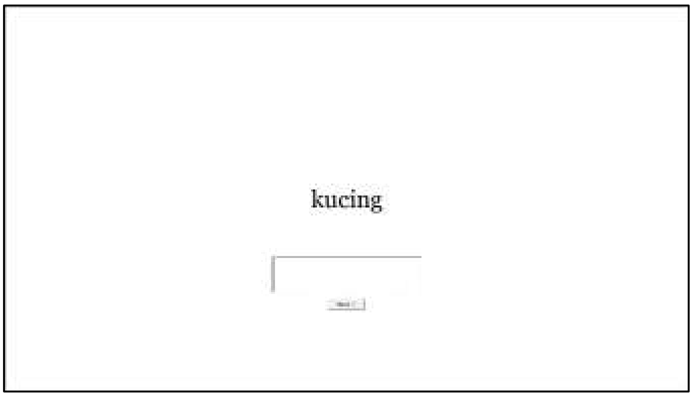


Appendix C (continued)

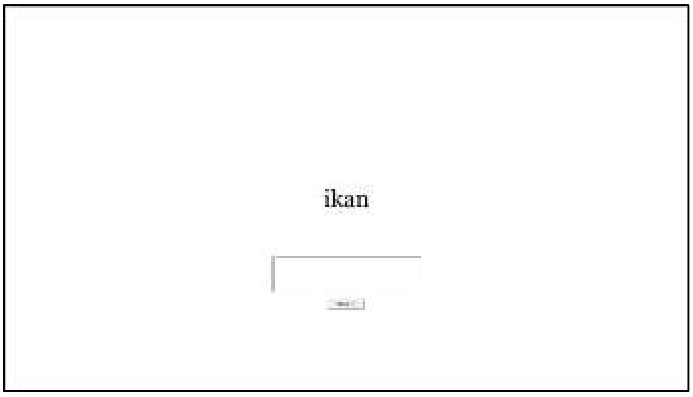
An Indonesian cue for 馬 (horse) * うま(Hiragana form) was also acceptable.



An Indonesian cue for 猫 (cat) * ねこ(Hiragana form) was also acceptable.

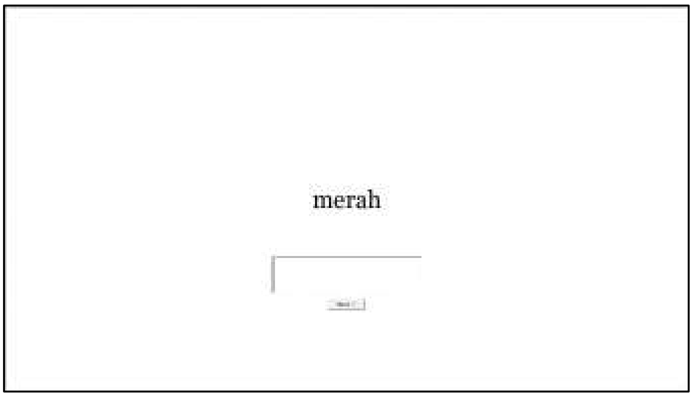


An Indonesian cue for 魚 (fish) * さかな(Hiragana form) was also acceptable.

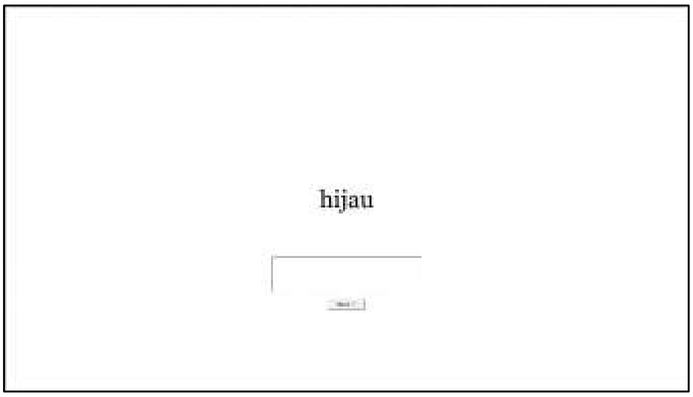


Appendix C (continued)

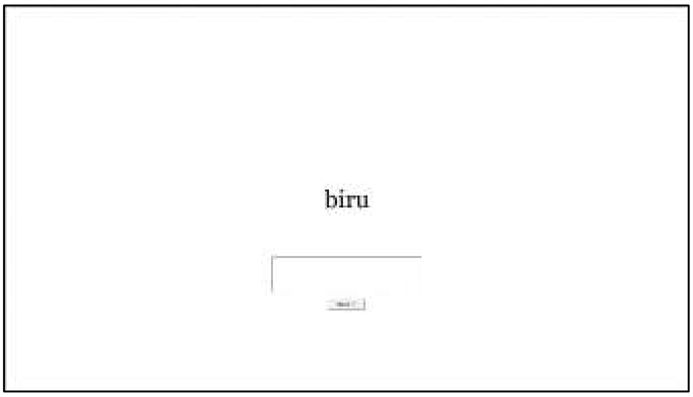
An Indonesian cue for 赤 (red) *あか(Hiragana form) was also acceptable.



An Indonesian cue for 緑 (green) *みどり(Hiragana form) was also acceptable.

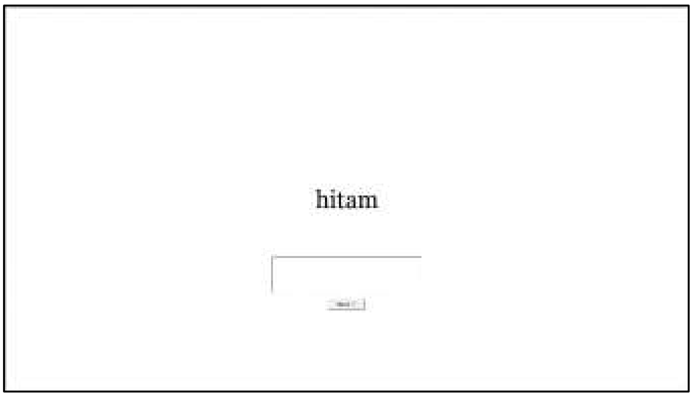


An Indonesian cue for 青 (blue) *あお(Hiragana form) was also acceptable.

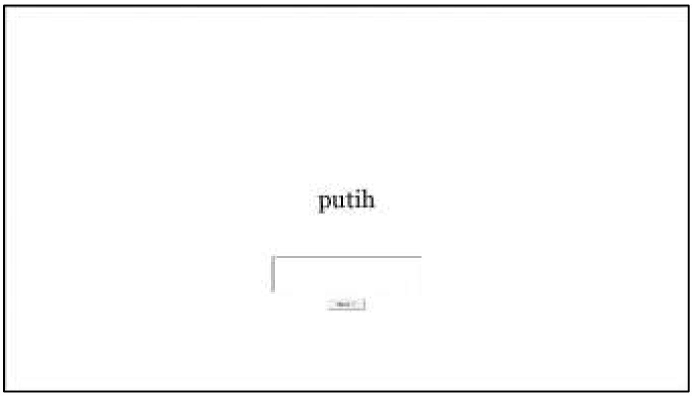


Appendix C (continued)

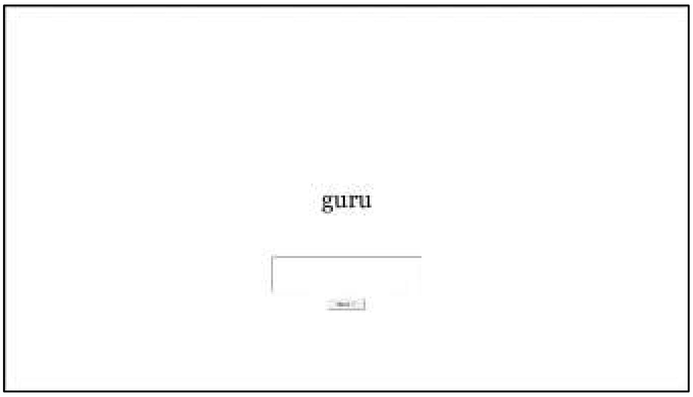
An Indonesian cue for 黒 (black) *くろ(Hiragana form) was also acceptable.



An Indonesian cue for 白 (white) *しろ(Hiragana form) was also acceptable.

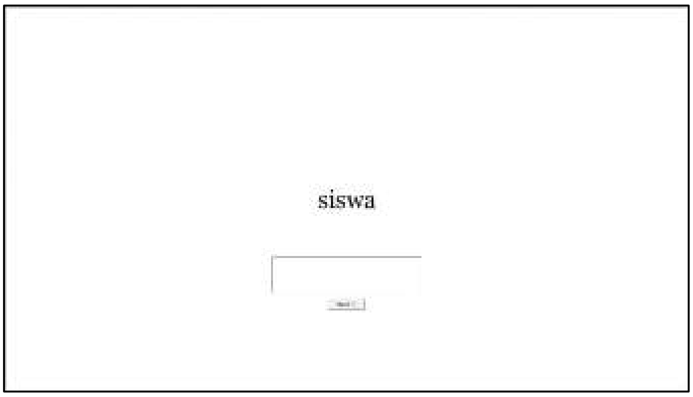


An Indonesian cue for 教師 (teacher) *きょうし(Hiragana form) was also acceptable.

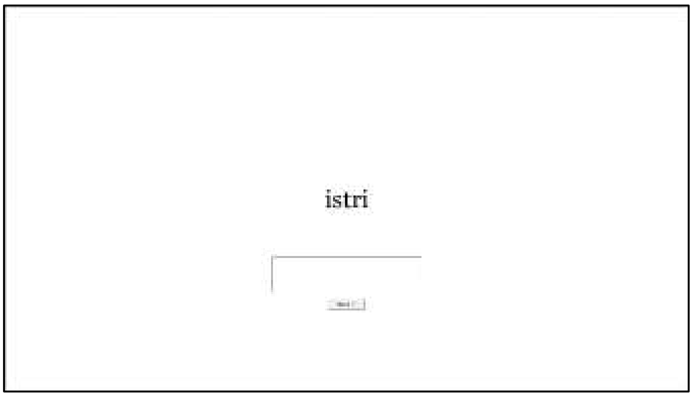


Appendix C (continued)

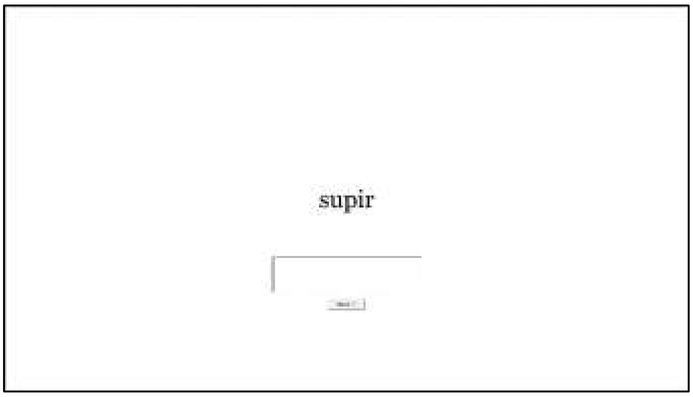
An Indonesian cue for 学生 (student) *がくせい(Hiragana form) was also acceptable.



An Indonesian cue for 妻 (wife) *つま(Hiragana form) was also acceptable.

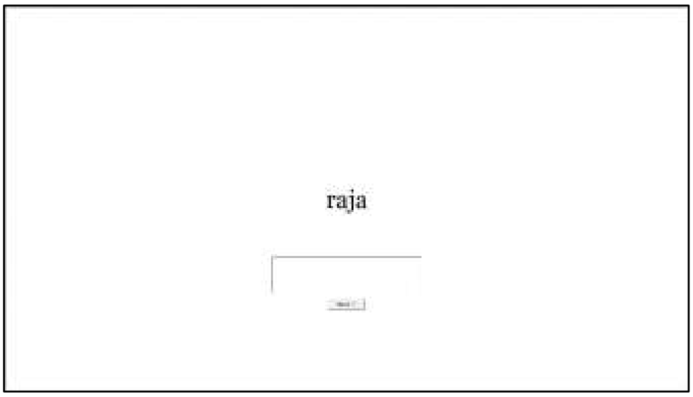


An Indonesian cue for 運転手 (driver) *うんてんしゅ(Hiragana form) was also acceptable.



Appendix C (continued)

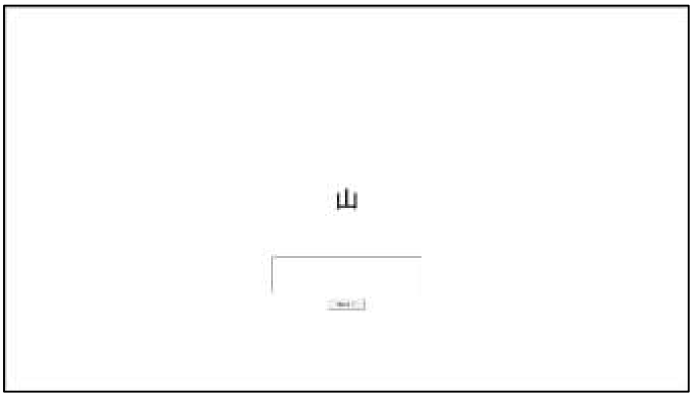
An Indonesian cue for 王 (king) *おう(Hiragana form) was also acceptable.



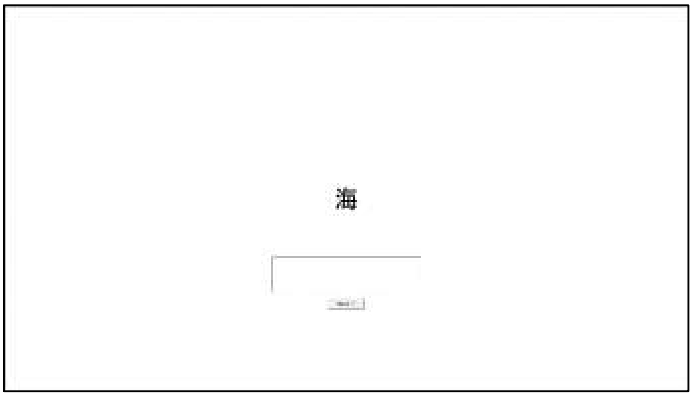
Appendix D

All Question Items on the "L1 cue–L2 response" Retention Posttest in Study1

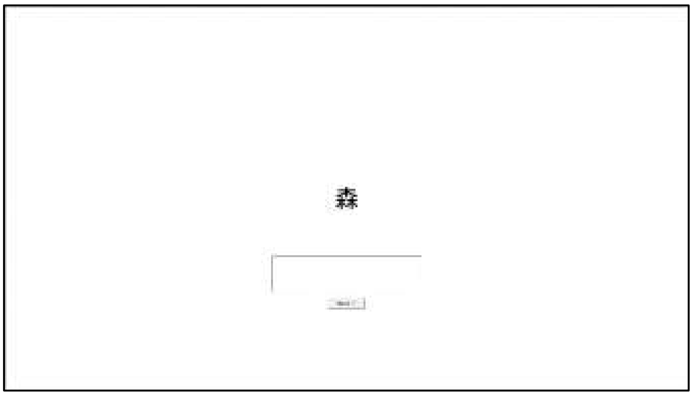
A Japanese cue for *gunung* (mountain)



A Japanese cue for *laut* (sea)

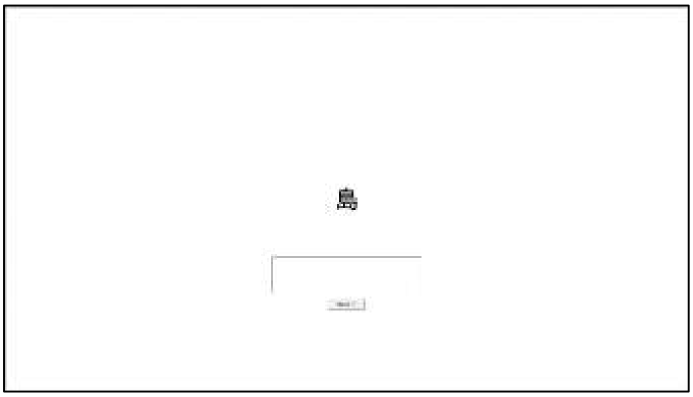


A Japanese cue for *hutan* (forest)

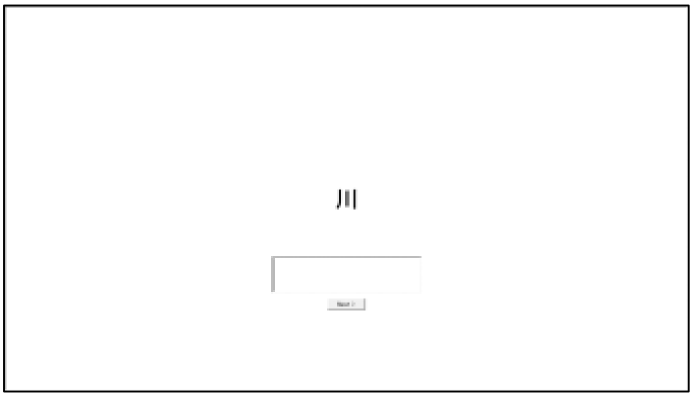


Appendix D (continued)

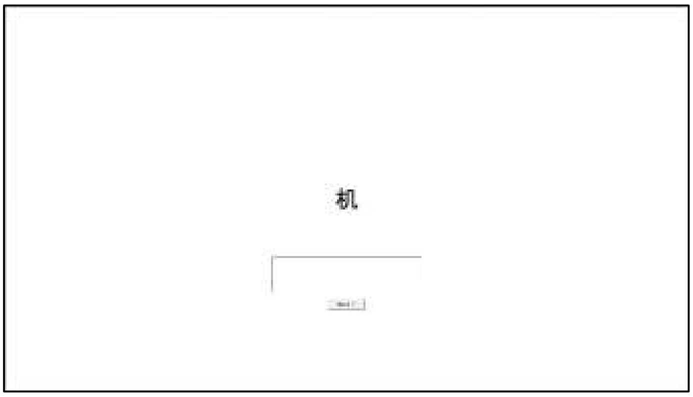
A Japanese cue for *pulau* (island)



A Japanese cue for s*ungai* (river)

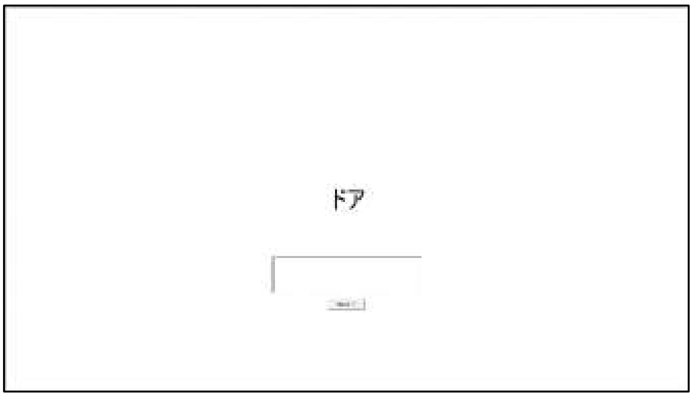


A Japanese cue for *meja* (table)

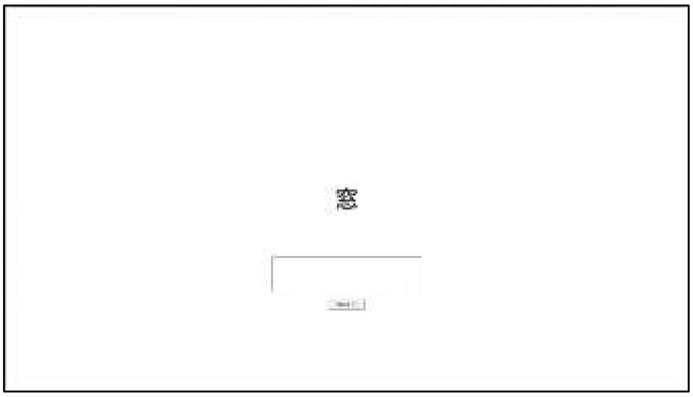


Appendix D (continued)

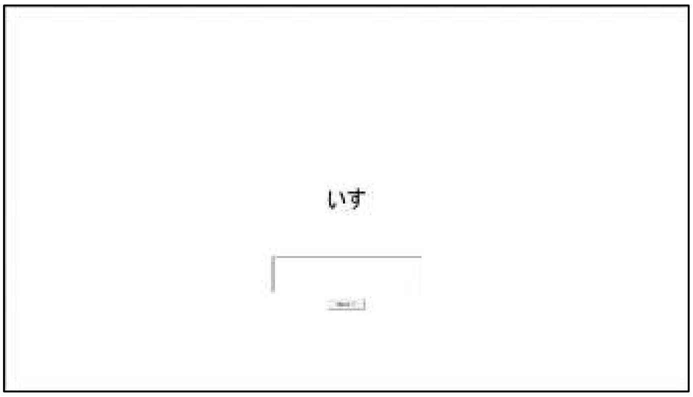
A Japanese cue for *pintu* (door)



A Japanese cue for *jendela* (window)



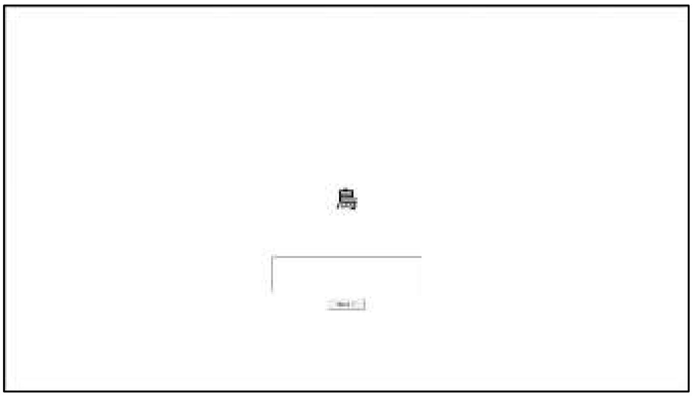
A Japanese cue for *kursi* (chair)



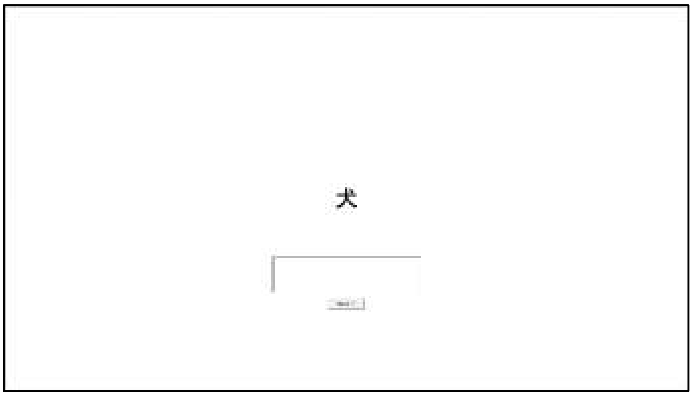
Appendix D (continued)

A Japanese cue for *kamar* (room)

A Japanese cue for *burung* (bird)

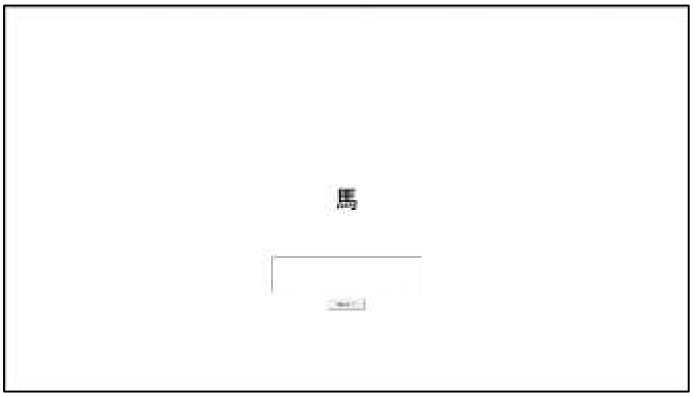


A Japanese cue for *anjing* (dog)

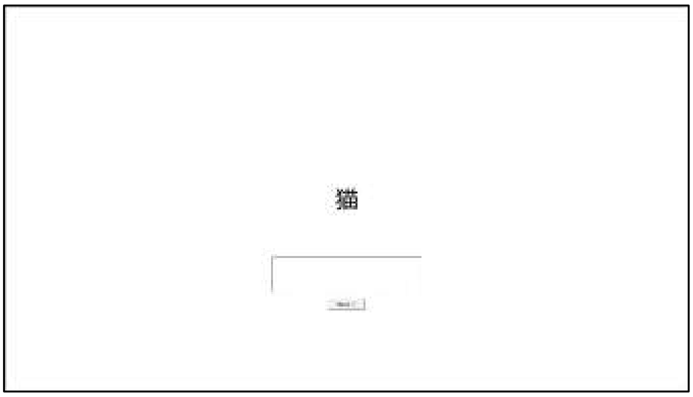


Appendix D (continued)

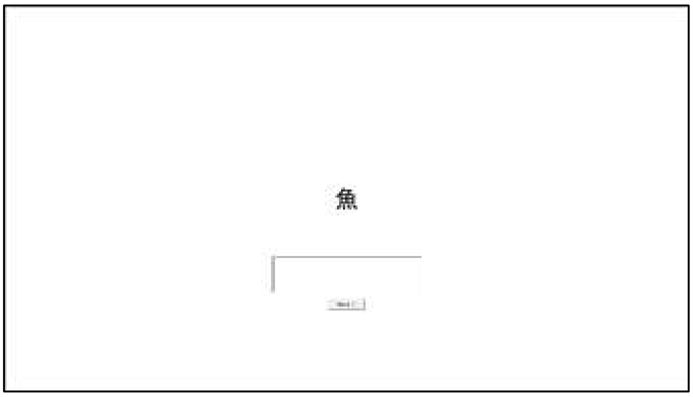
A Japanese cue for *kuda* (horse)



A Japanese cue for *kucing* (cat)

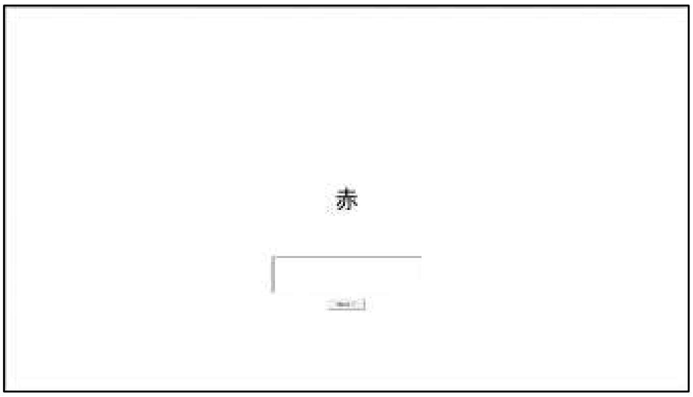


A Japanese cue for *ikan* (fish)

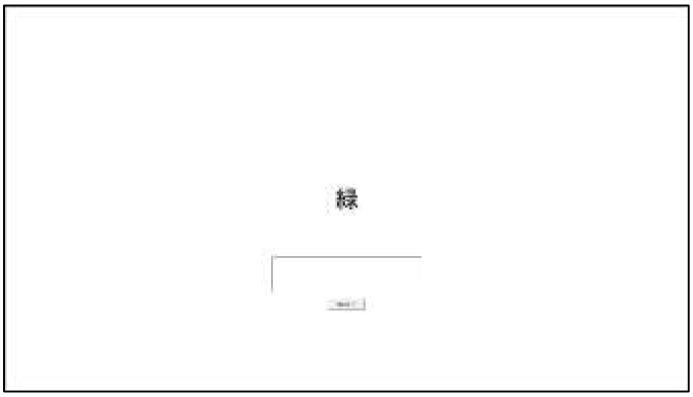


Appendix D (continued)

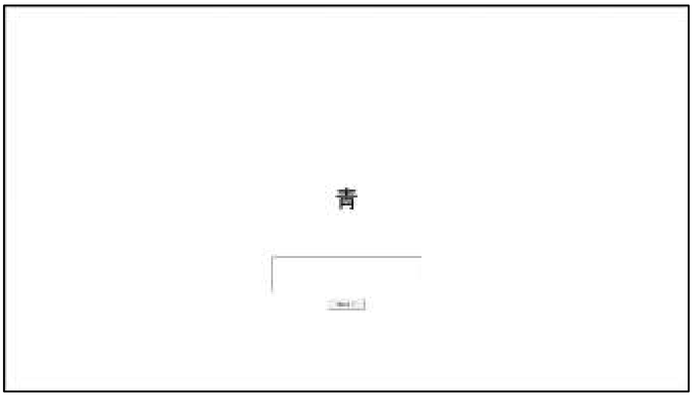
A Japanese cue for *merah* (red)



A Japanese cue for *hijiau* (green)

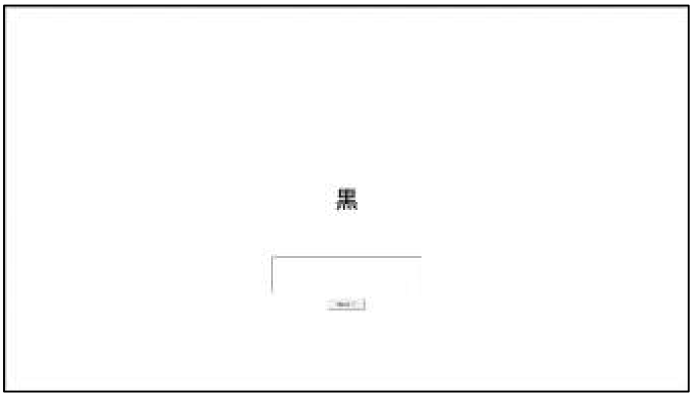


A Japanese cue for *biru* (blue)

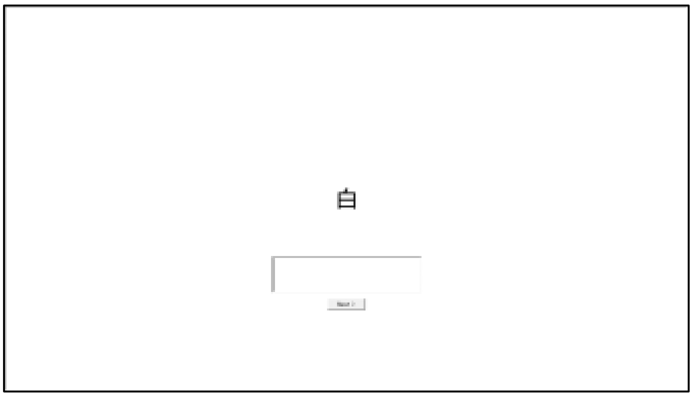


Appendix D (continued)

A Japanese cue for *hitam* (black)



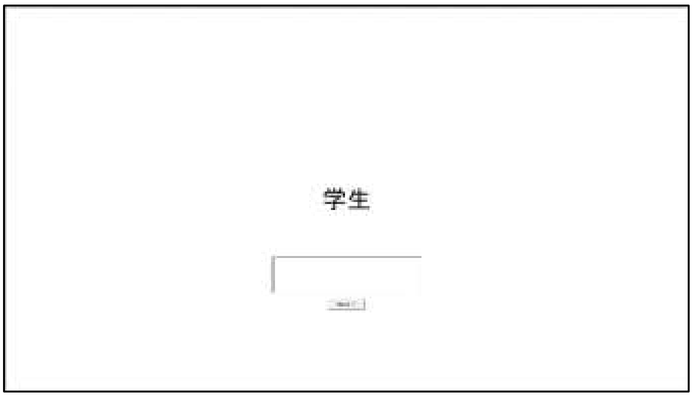
A Japanese cue for *putih* (white)



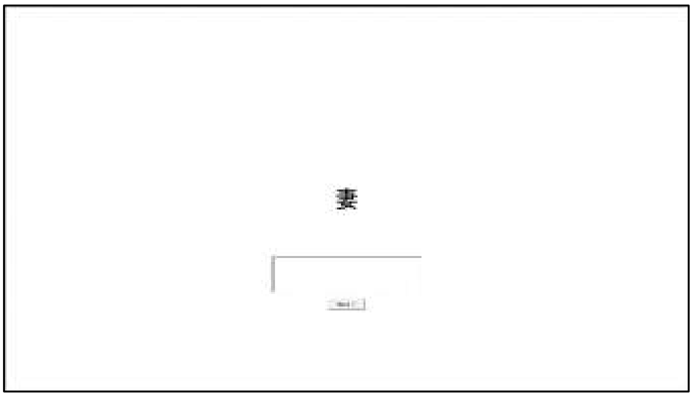
A Japanese cue for *guru* (teacher)

Appendix D (continued)

A Japanese cue for *siswa* (student)



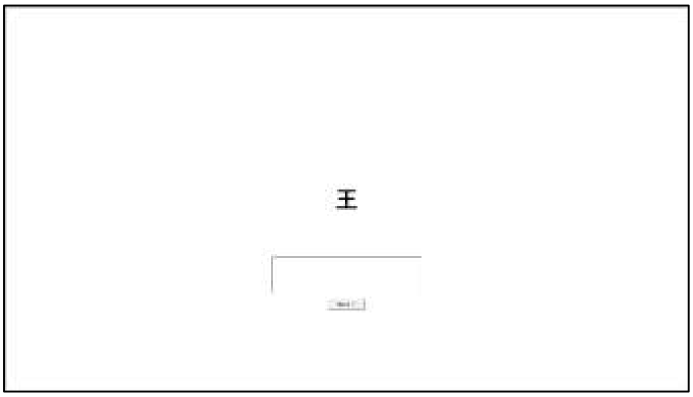
A Japanese cue for *istri* (wife)



A Japanese cue for *supir* (driver)

Appendix D (continued)

A Japanese cue for *raja* (king)



Appendix E

All Question Items on the English Vocabulary Size Test

英語語彙クイズ

○以下英文の太字で表される部分はどのような意味になりますか。最も相応しいと思う選択肢(a-d)に○をつけて下さい。

11. Can they **maintain** it?
a.維持する b.拡大する c.改良する d.入手する

12. He sat on a **stone**.
a.石　b.腰かけ　c.敷物(しきもの)　d.枝

13. I am **upset**.
a.疲れた　b.有名な　c.金持ちの　d.うろたえた

14. The **drawer** was empty.
a.引き出し　b.車庫　c.冷蔵庫　d.貯金箱

15. He has no **patience**.
a.忍耐力がない　b.自由な時間が無い
c.信念がない　d.公平さが無い

16. His mark for that question was **nil**.
a.とても悪い　b.ゼロ　c.とても良い　d.中間の

17. They went to the **cliff**.
a.崖　b.浜辺　c.商店街　d.市民プール

18. Make a **circle**.
a.下書き　b.空白　c.円　d.大きな穴

19. Please use the **microphone**.
a.電子レンジ　b.マイク　c.顕微鏡
d.携帯電話

20. He's an **idiot**.
a.スパイ　b.ばか者　c.記者　d.プロ選手

21. He is a **soldier**.
a.実業家　b.学生　c.金属細工人　d.兵士

22. It has been **restored**.
a.繰り返された　b.移転された　c.値引きされた
d.復元された

23. He was holding a **jug**.
a.水入れ　b.おしゃぶり　c.ベレー帽　d.拳銃

24. He is **scrubbing** it.
a.ひっかく　b.修理する　c.ごしごし磨く
d.下絵を描く

25. The children were pretending to be **dinosaurs**.
a.海賊　b.妖精　c.竜　d.恐竜

26. He broke the **strap**.
a.約束　b.ふた　c.皿　d.ひも

27. It was **paved**.
a.ふさがれた　b.分割された
c.金の縁取りがされた　d.舗装(ほそう)された

28. They **dashed** over it.
a.駆けた　b.ぶらぶら歩いた　c.戦った
d.ちらりと見た

29. He couldn't stop **roving**.
a.酔うこと　b.放浪すること　c.鼻歌を歌うこと
d.一生懸命働くこと

30. He felt **lonesome**.
a.不愉快な　b.とても疲れた　c.孤独な
d.活発な

Appendix E (continued)

英語語彙クイズ

31. They made a new **compound**.
    a.契約　b.複合物　c.会社　d.予想

32. I agree with the **latter**.
    a.牧師　b.理由　c.後者　d.答え

33. Please be **candid**.
    a.注意深い　b.思いやりのある　c.公平な
    d.率直な

34. Look at my **tummy**.
    a.襟(えり)巻き　b.腹　c.リス　d.親指

35. We found a **fossil**.
    a.遺跡　b.洞窟　c.化石　d.わな

36. We need more **acid** to make it.
    a.酸　b.水素　c.炭素　d.水銀

37. Do you like **crabs**?
    a.カニ　b.クラッカー　c.かたい襟(えり)
    d.コオロギ

38. You will need more **nutrition**.
    a.栄養　b.技術　c.知恵　d.支援

39. We found a good **remedy**.
    a.改善法　b.飲食店　c.調理法　d.方程式

40. They **alleged** it.
    a.それを主張した　b.それを盗作した
    c.それを証明した　d.それに反論した

41. The company **had a large deficit**.
    a.多額の赤字を抱えた　b.価値が大いに下がった
    c.多額の支出計画を立てた
    d.多額の銀行預金があった

42. He **wept**.
    a.卒業した　b.泣いた　c.死んだ　d.心配した

43. We saw a **nun**.
    a.ミミズ　b.大事故　c.修道女　d.火の玉

44. The house is **haunted**.
    a.装飾品にあふれた　b.賃借りされた
    c.空き家の　d.幽霊の出没する

45. We need some **compost**.
    a.強力な援助　b.精神的支え　c.コンクリート
    d.堆肥(たいひ)

46. I need one more **cube**.
    a.ピン　b.立方体　c.マグカップ　d.はがき

47. It is a **mimic**.
    a.ミニチュア　b.宝箱　c.微生物　d.模造品

48. Shall I **peel** it?
    a.水に浸す　b.皮をむく　c.漂泊する
    d.薄く切る

49. They found a **fracture**.
    a.破損　b.小片　c.上着　d.珍しい宝石

50. They didn't find a single **bacterium**.
    a.細菌　b.植物の一種　c.ラクダ　d.盗品

Appendix E (continued)

英語語彙クイズ

51. Your plans are **devious**.
    a.ごまかしの多い　b.十分に練られた
    c.十分に考えられていない
    d.必要以上に費用のかかる

52. The **premier** spoke for an hour.
    a.弁護士　b.大学講師　c.冒険家　d.首相

53. They have a **butler**.
    a.執事(しつじ)　b.のこぎり　c.家庭教師
    d.地下貯蔵庫

54. They gave us some **antibodies**.
    a.処方箋　b.遺伝子　c.試薬　d.抗体

55. They raised the **threshold**.
    a.旗　b.境界点　c.天井　d.利子

56. She has completed her **thesis**.
    a.論文　b.要約　c.見習い期間
    d.長期入院治療

57. He **strangled** her.
    a.絞(し)め殺した　b.甘やかした　c.誘拐した
    d.賞賛した

58. He treated her in a **cavalier** manner.
    a.思慮に欠ける　b.礼儀正しい　c.ぎこちない
    d.子供にふさわしい

59. His **malign** influence is still felt.
    a.悪い b.良い c.非常に重要な d.秘密の

60. The car **veered**.
    a.急に進路を変えた b.揺れた c.大きな音をたてた
    d.横滑りした

61. We bought **olives**.
    a.油を含んだ果実　b.香りの強い赤い花
    c.男性用水着　d.草刈り機

62. They made a **quilt**.
    a.遺言　b.契約　c.織物の一種　d.羽ペン

63. They did it by **stealth**.
    a.大金を使うこと　b.無理強いすること
    c.ひそかなやり方　d.直面している問題を気に留めない

64. The boy **shuddered**.
    a.ささやいた　b.落ちかけた　c.身震いした　d.叫んだ

65. The **bristles** are too hard.
    a.質問　b.剛毛　c.折りたたみベッド　d.靴底

66. They have joined this **bloc**.
    a.楽団　b.窃盗団　c.偵察隊　d.連合

67. This book is about **demography**.
    a.地理学　b.幾何学　c.水力学　d.人口統計学

68. That's a good **gimmick**.
    a.作業用足場　b.財布　c.注意をひくもの
    d.からくり

69. This **azalea** is very pretty.
    a.植物の一種　b.天然糸から作られた軽い素材
    c.インドの女性が身に着ける衣装　d.貝の一種

70. This **yoghurt** is disgusting.
    a.沈殿物　b.傷口　c.乳製品の一種
    d.植物の一種

Appendix E (continued)

英語語彙クイズ

71. He was **erratic**.
  a.落ち度がない　b.非常に悪い
  c.非常に礼儀正しい　d.気まぐれな

72. He lost his **palette**.
  a.釣りかご　b.食欲　c.若い女性の友人
  d.塗料を混ぜる皿

73. His influence **was null**.
  a.良い結果を生んだ　b.役に立たなかった
  c.効果がなかった　d.長引いた

74. This is a good **kindergarten**.
  a.気晴らし　b.幼稚園　c.リュックサック
  d.市営の図書館

75. There was an **eclipse**.
  a.暴風雨　b.水しぶきの音　c.大虐殺
  d.日食

76. This is the **marrow**.
  a.マスコット　b.骨髄(こつずい)　c.操縦かん
  d.昇給分の差額

77. There were hundreds of **locusts**.
  a.バッタ　b.志願者　c.菜食主義者
  d.山菜の一種

78. It is **authentic**.
  a.本当の　b.とても騒がしい　c.古い
  d.乾燥した

79. We saw the **cabaret**.
  a.壁画　b.ナイトクラブ　c.虫の一種
  d.想像上の生物

80. He started to **mumble**.
  a.集中する　b.身震いする　c.居座る
  d.つぶやく

81. Does it have a **hallmark**.
  a.消費期限　b.品質証明
  c.王室の承認印　d.偽造防止の印

82. He is a **puritan**.
  a.注目を浴びたがる人　b.厳格な道徳心を持つ人
  c.放浪する人　d.ケチな人

83. Now he has a **monologue**.
  a.片めがね　b.一人芝居　c.独占　d.絵文字

84. We looked at the **weir**.
  a.奇人　b.紅葉　c.旧式金管楽器　d.ダム

85. He had lots of **whims**.
  a.古い金貨　b.雌(めす)馬　c.思いつき　d.こぶ

86. I was **perturbed.**
  a.同意させられた　b.不安な　c.困り果てた
  d.ずぶぬれの

87. They choose a **regent**.
  a.無責任な人　b.臨時司会者　c.摂政(せっしょう)
  d.代表者

88. They saw an **octopus**.
  a.フクロウ　b.潜水艦　c.プロペラ機　d.タコ

89. The story is set in the **fens**.
  a.沼地　b.岩山　c.貧困街　d.大昔

90. He painted the **lintel**.
  a.ドアや窓の枠の上部　b.渡し舟　c.柑橘系の植物
  d.垂れ幕

Appendix F

"Reading-A" Passages and While-reading Questions in Study 2

Screen – 1: Reading Passage

Jenny Tinker is fourteen. She and her father, Sam, work in Covent Garden market. One day after work Jenny sees a man. He is running and he has got a watch in his ponger*.

"Here – take this," he says, and gives Sam the watch. Five seconds later a lot of angry people run down the street after Sam. "There's the thief! He's got my watch – stop him!" says an old man. "You're coming with us," a dobron* says to Sam. They do not listen to Jenny's story.

Who can she talk to? She goes to Peter Stone's house. "What's going to happen?" asks Jenny. "They send thieves to Australia," he says. Week later Jenny watches a big ship – the Black Star – on the River Thames. One of the men on it is her father.

"How can I help him?" thinks Jenny. Five minutes later she talks to the ship's captain. "I want a job," she says. "My name's Ned Bell." "Well ... why not," says the captain. "OK, Ned – you can carry my gistra*, here."

For six weeks Jenny works very hard on the Black Star. >> [Press SPACE]

*ponger = hand

*dobron = policeman

*gistra = bag

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix F (continued)

Screen – 1: While-reading Questions

Probably True or False? (Press T/F key)

1. The running man has a watch in his hand.

2. The doctor asked Sam to come with him.

3. Jenny takes the captain's gistra because she is hungry.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. T, 2. F, 3. F

Appendix F (continued)

Screen – 2: Reading Passage

One night, she goes to the captain's room. Usually his keys are next to the bed. Tonight they are on the suiste*. Jenny takes them and goes down to the bottom of the ship. There she finds her father. "Ssshh – don't ask questions," says Jenny. "Come with me."

Fifteen minutes later they are in a small boat. The Black Star is behind them. It is a beautiful night and there is a big voteen*. It shines on Sam Tinker's thin, tired murdal*. "Oh, Dad," she says. "You're free!"

After three days Sam and Jenny are very hungry and thirsty. Then, Jenny sees a big white bird. It is carrying some grass. "Look Dad," she says. "Do you think there's an island near hear?" Sam stands up in the small boat. "Yes, Jenny, you're right! There is an island!!"

The island is big and has large woods. First Sam and Jenny look for food and water. They find a small river. They drink and drink and drink. There are coconuts on the island, too. >> [Press SPACE]

*suiste = table

*voteen = moon

*murdal = face

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix F (continued)

Screen – 2: While-reading Questions

Probably True or False?   (Press T/F key)

1. Jenny has to reach over her head to take the keys off the suiste.

2. The voteen was not green or purple.

3. Sam's murdal could be seen more clearly by looking at him from the back.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. F, 2. T, 3. F

Appendix F (continued)

Screen – 3: Reading Passage

That night Sam and Jenny sleep on the fuadar*. They are very tired. In the morning they start to make a small house with wood. After four days they finish. Then Sam makes a liober* with his shirt and puts it on the roof. “This is ‘Tinkers Island’ now,” Sam says.

Sam and Jenny are very happy for the first year. They like their new home. One day, something happens. It is early in the morning. Jenny is fishing. Sam is looking for coconuts. He does not see a long, thick snake on the ground. He puts his foot on the snake and it bites him.

Two hours later Jenny finds her father on the ground. He cannot walk. She pulls him to the house. After that, there is nothing she can do. Sam is very, very weak. He cannot eat or drink. Jenny sits next to him for three days and nights. Then Sam opens his arnauns*. “Jenny?” he asks.

After ten days Sam is well again, but not very strong. He sleeps a lot and Jenny does all the work. >> [Press SPACE]

*fuadar = beach

*liober = flag

*arnaun = eye

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix F (continued)

Screen – 3: While-reading Questions

Probably True or False?   (Press T/F key)

1. They slept on the fuadar so that they could avoid getting wet in the rain.

2. Sam made a liober to show they are living there.

3. Sam said "Jenny?" after opening his arnauns because he could see her.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. F, 2. T, 3. T

Appendix F (continued)

Screen – 4: Reading Passage

Then, one morning, she sees a ship and it is coming to the island. She runs to the house. “Dad, Dad!” she says. “Come and look!” Sam walks to the brosna* of the small house. “Do you think it’s the Black Star?” he asks. He can see three men. They are putting a boat in the water. “I don’t like this. Come on,” he says.

Twenty minutes later the three men are on Tinkers Island. Sam and Jenny stand behind a big ruadog* and look at them. “It’s OK, Dad,” say Jenny. “They’re not from the Black Star.” At that moment, one of the men turns and says, “Who’s there?” He has got a gun. Jenny says “Please don’t kill us”. The man looks at her and his targer* opens. “You’re English!!” he says.

That evening one of the men says, “Our ship, the Red Rose, is going to America. Come with us. You can start a new life there.” Next day, Sam and Jenny are on the Red Rose. They are happy, but sad, too. The island is very small now. “Goodbye, Tinkers Island,” they say. >> [Press SPACE]

*brosna = door

*ruadog = tree

*targer = mouth

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix F (continued)

Screen – 4: While-reading Questions

Probably True or False?   (Press T/F key)

1. Sam went through the brosna into the house.

2. There was a large ruadog in the house.

3. When one of the men saw Jenny, he opened his targer to speak.

Notes: The questions were shown one at a time on the same location in an overridden manner.
Answer Key: 1. T, 2. F, 3. T

Appendix G

"Reading-B" Passages and While-reading Questions in Study 2

Screen – 1: Reading Passage

Felix is a blue and yellow bird from Brazil. He lives with the Baxter family in New York. His home is a big cage. Felix likes the Baxters, and the Baxters like him. But Felix is not happy. He wants to go home to Brazil.

Every night Felix looks up at the sky. He remembers his last day in the jungle. The man with a white coisin*. The big box. The long plane journey. The shop – "Beautiful Birds!" – in New York. "I want to be with my family again," he thinks. "I want to fly home to the jungle. He puts his head under one moiley*. "One day," he says.

"One day" comes two weeks later. Mr. Baxter opens Felix's cage to give him some food. Then he hears the pingin*. "Can you answer that, George? I'm in the bath." asks Mrs. Baxter. "OK," says Mr. Baxter. He leaves the room, but forgets to shut Felix's cage. Felix can see an open window. "This is it!" he thinks.

He flies out of his cage and out of the window. >> [Press SPACE]

*coisin = hat

*moiley = wing

*pingin = telephone

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix G (continued)

Screen – 1: While-reading Questions

Probably True or False? (Press T/F key)

1. Felix remembered the white coisin because it looked dangerous.

2. Felix put his head under one moiley to help him go to sleep.

3. Mr. Baxter heard the sound of a pingin that came from outside the house.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. F, 2. T, 3. F

Appendix G (continued)

Screen – 2: Reading Passage

He looks down at the streets and buildings and flies through the city. A girl in a moneen* with her mother sees him. "Look!" she says. Her mother does not hear. She is reading her dallog*.

After an hour Felix stops. He stands on the head of the Statue of Liberty. "Where are you from?" asks a small, grey and white bird beside him. "I'm from Brazil," answers Felix. Felix flies up into the sky again. "I'm going home," he says. "Goodbye."

Felix flies south. Soon he cannot see New York. All he can see is the Atlantic Ocean. In the evening the tamall* starts to go down. Felix is hungry, but he is happy, too. For the first time in two years he is free. He wants to fly all night.

Two hours later it starts to rain. Soon, Felix cannot see the stars. "Where am I?" he thinks. He looks down at the sea. Then he sees something. Is it a star in the black water? He looks again. It is a ship. Felix flies down to the ship.

>> [Press SPACE]

*moneen = café

*dallog = newspaper

*tamall = sun

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix G (continued)

Screen – 2: While-reading Questions

Probably True or False? (Press T/F key)

1. People in a moneen usually eat or drink something.

2. The mother's dallog was over ten years old.

3. The tamall always goes down in the evening.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. T, 2. F, 3. T

Appendix G (continued)

Screen – 3: Reading Passage

There are hundreds of fish on it. He eats twenty in five minutes, then he sleeps. In the morning a man on the ship sees him. “Oh no!” thinks Felix, but it is all right. The man only wants to take his gogail*. Felix flies into the sky again.

Two days later Felix flies across Peru. He looks down and sees the Inca city of Machu Picchu. “Those old stone buildings look interesting,” he thinks. Felix starts to fly down to the buildings. Then he sees a big black bird with a white olagon*.

“Hello, you are not from Peru, are you?” says the bird. Felix tells his story. “Now I’m going home. Can you tell me how to get there?” “I have a friend - Aca – in Rio,” says the bird. “He knows the jungle. Go there and ask him.”

Felix flies from Peru to Rio. There he sees a lot of people with colorful baitin* in the street. He can hear music and see some birds on a building. Is one of them Aca? >> [Press SPACE]

*gogail = photograph

*olagon = neck

*baitin = clothes

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix G (continued)

Screen – 3: While-reading Questions

Probably True or False? (Press T/F key)

1. The man on the ship tried to take Felix's gogail off him.

2. The black bird had a white neck.

3. Felix saw people carrying colorful baitin in the street.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. F, 2. T, 3. F

Appendix G (continued)

Screen – 4: Reading Passage

“Aca! Aca!” he says. “Yes, that’s me,” answers one of the birds. Felix cannot see him. Then a fecker* sees Felix. “Look at the bird,” he says to his father. “Can I have him? Can I? Please?” he asks.

Felix is still looking at all the birds on the building. Then the father catches him. Felix fights with his feet. “Help! Help!” Aca helps him. “Aiiiieee!!!” says the man. The two birds fly into the sky. Then he tells Aca his story.

Felix and Aca fly from Rio to the jungle. Felix says, “Yes, I remember that village. My home’s very near here. It’s only...” Then he stops. In front of them there are a lot of men and machines. “Oh no, they are making a new mullan*!” says Aca.

They stop on a scolog*. Felix is tired and sad. Then Felix sees a small blue and yellow feather in the air. He looks up. Above him, there are four birds flying. “It’s them, Aca!” he says. Then he flies into the sky. The birds see him. “Felix, is it you?” they ask. “Yes,” answers Felix. “I’m home!!!” >> [Press SPACE]

*fecker = boy

*mullan = town

*scolog = tower

Note: Glosses were provided only in the here-and-now reading condition. Also, an asterisk after each target nonword in the passage was shown only in the here-and-now reading condition.

Appendix G (continued)

Screen – 4: While-reading Questions

Probably True or False? (Press T/F key)

1. The fecker could not catch Felix by herself because she was too young.

2. Aca thought that the men were making a mullan probably because he saw a lot of new roads there.

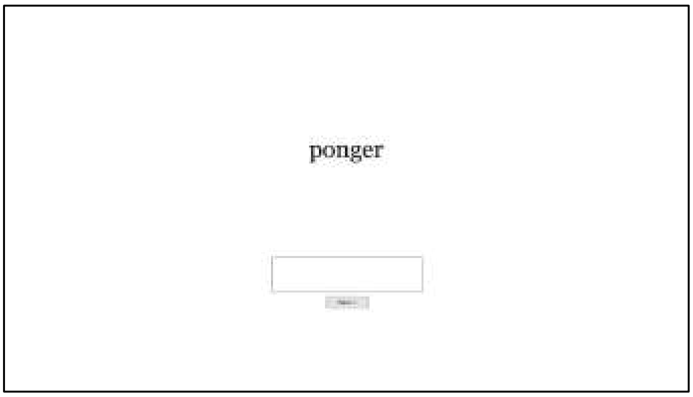
3. Felix and Aca stop on a scolog which is high above the ground.

Notes: The questions were shown one at a time on the same location in an overridden manner. Answer Key: 1. F, 2. T, 3. T

Appendix H

All Question Items on the Retention Posttest in Study 2

A nonword cue for *hand*

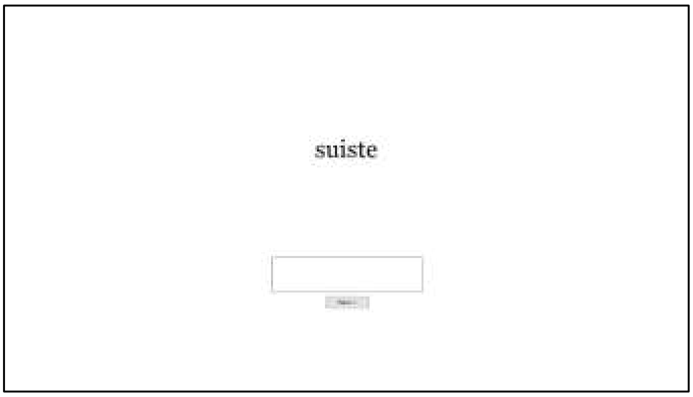


A nonword cue for *policeman*

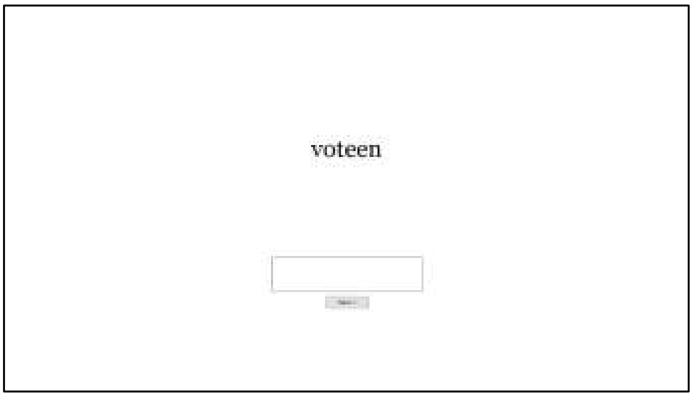
A nonword cue for *bag*

Appendix H (continued)

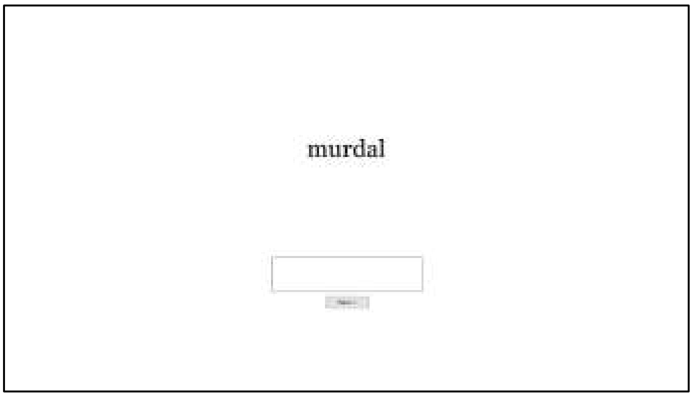
A nonword cue for *table*



A nonword cue for *moon*



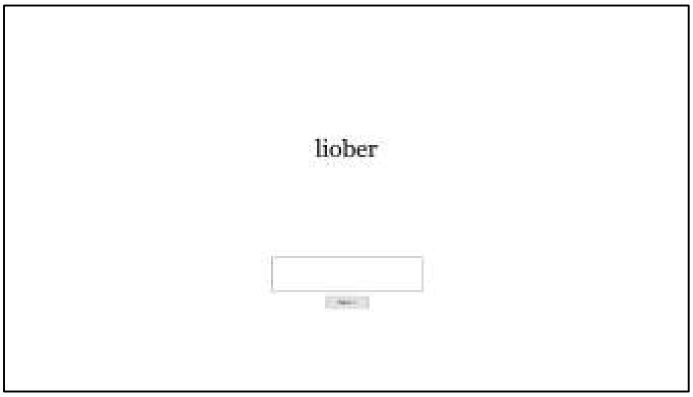
A nonword cue for *face*



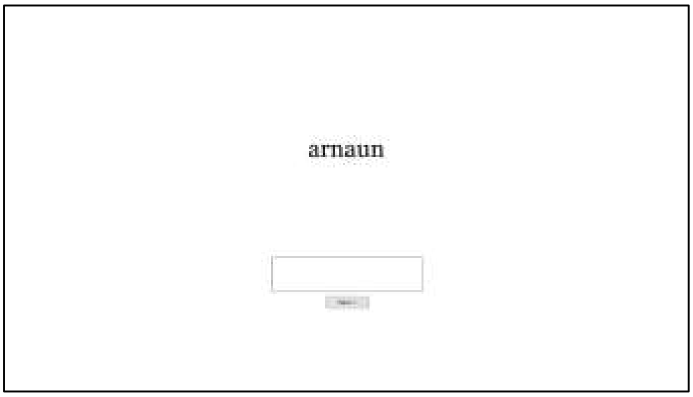
Appendix H (continued)

A nonword cue for *beach*

A nonword cue for *flag*



A nonword cue for *eye*



Appendix H (continued)

A nonword cue for *door*

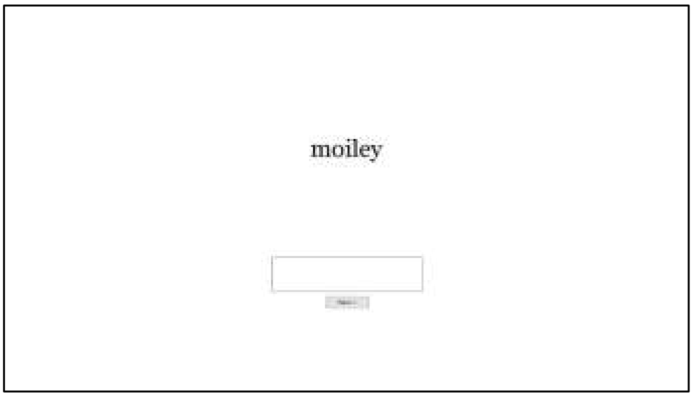
A nonword cue for *tree*

A nonword cue for *mouth*

Appendix H (continued)

A nonword cue for *hat*

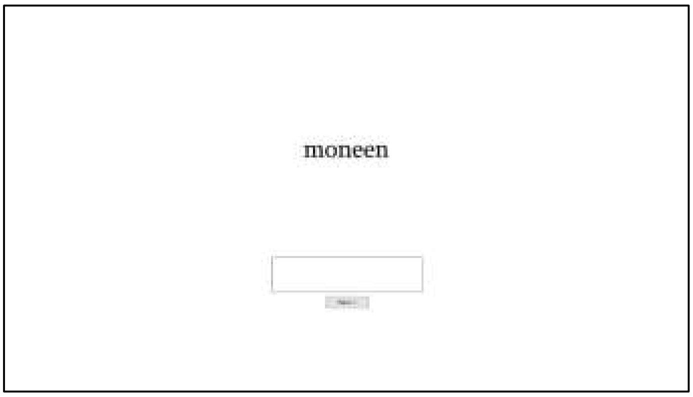
A nonword cue for *wing*



A nonword cue for *telephone*

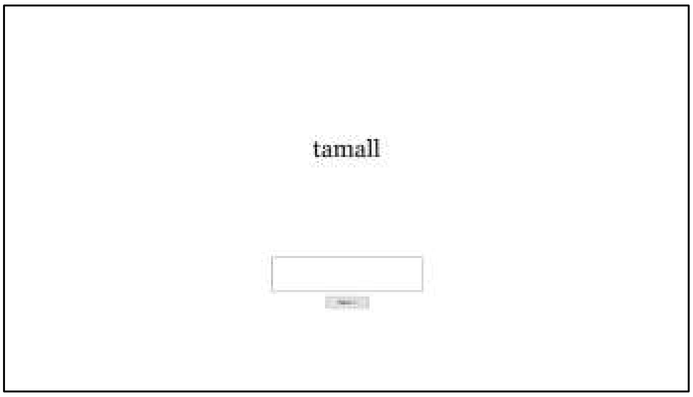
Appendix H (continued)

A nonword cue for *cafe*



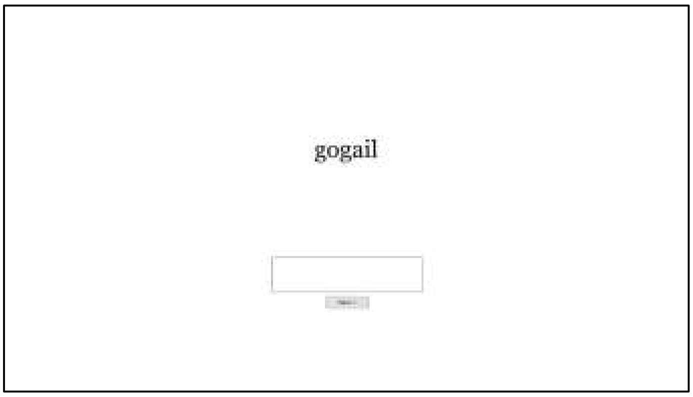
A nonword cue for *newspaper*

A nonword cue for *sun*



Appendix H (continued)

A nonword cue for *photograph*



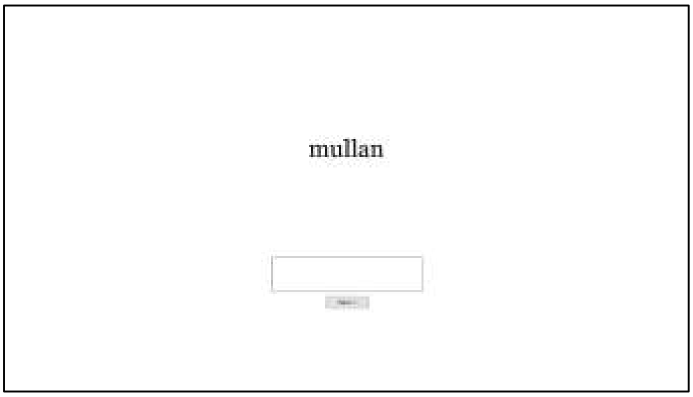
A nonword cue for *neck*

A nonword cue for *clothes*

Appendix H (continued)

A nonword cue for *boy*

A nonword cue for *town*

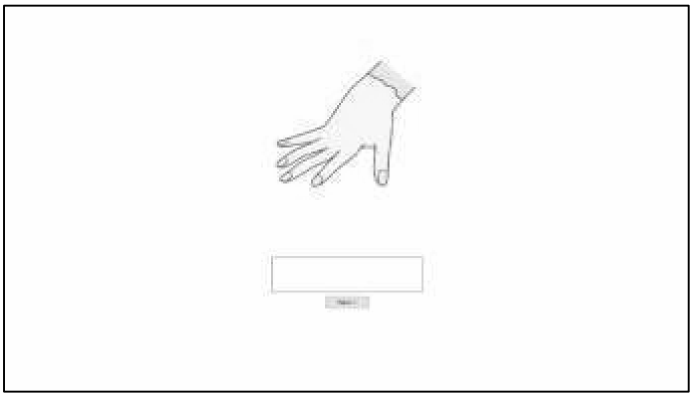


A nonword cue for *tower*

Appendix I

All Question Items on the Transfer Posttest in Study 2

A pictorial cue for *ponger* (hand)



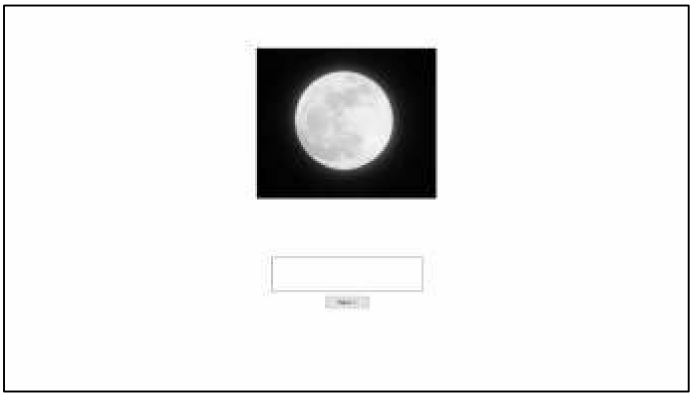
A pictorial cue for *dobron* (policeman)

A pictorial cue for *gistra* (bag)

Appendix I (continued)

A pictorial cue for *suiste* (table)

A pictorial cue for *voteen* (moon)

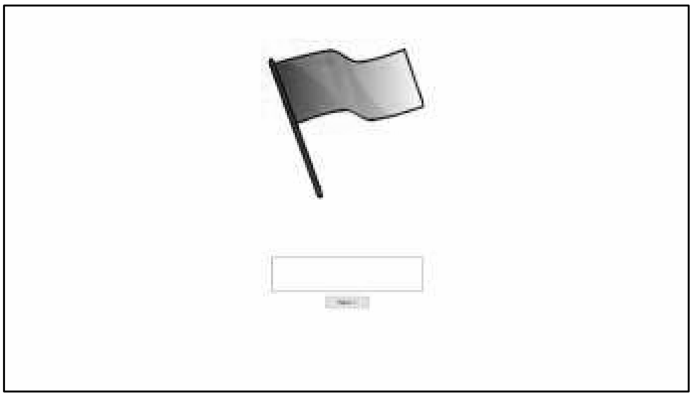


A pictorial cue for *murdal* (face)

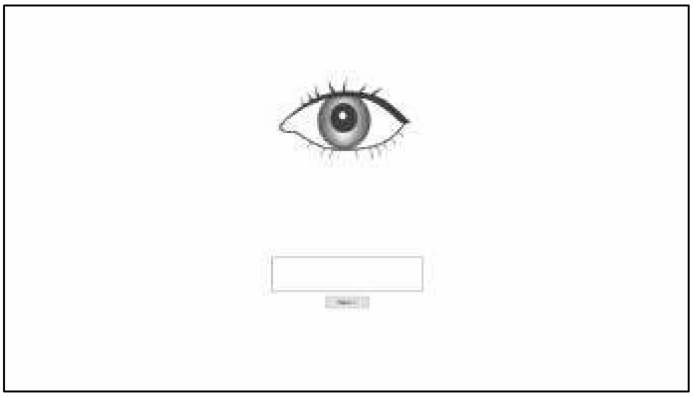
Appendix I (continued)

A pictorial cue for *fuadar* (beach)

A pictorial cue for *liober* (flag)



A pictorial cue for *arnaun* (eye)



Appendix I (continued)

A pictorial cue for *brosna* (door)

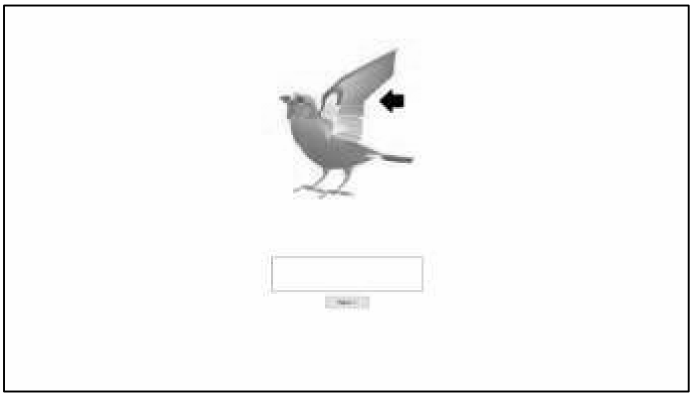
A pictorial cue for *ruadog* (tree)

A pictorial cue for *targer* (mouth)

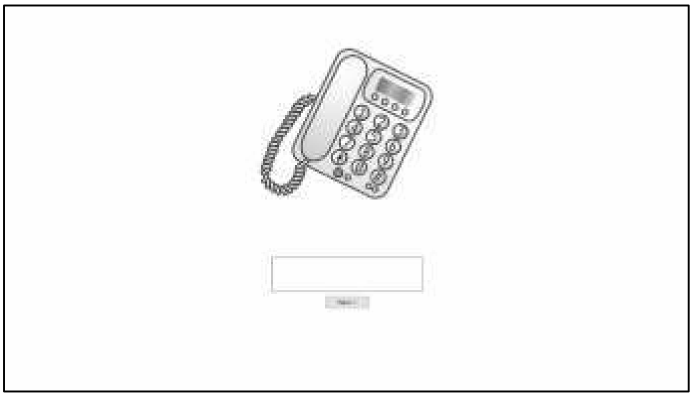
Appendix I (continued)

A pictorial cue for *coisin* (hat)

A pictorial cue for *moiley* (wing)



A pictorial cue for *pingin* (telephone)

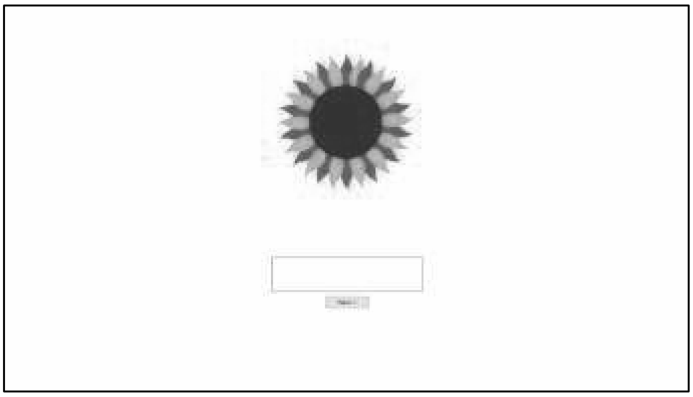


Appendix I (continued)

A pictorial cue for *moneen* (cafe)

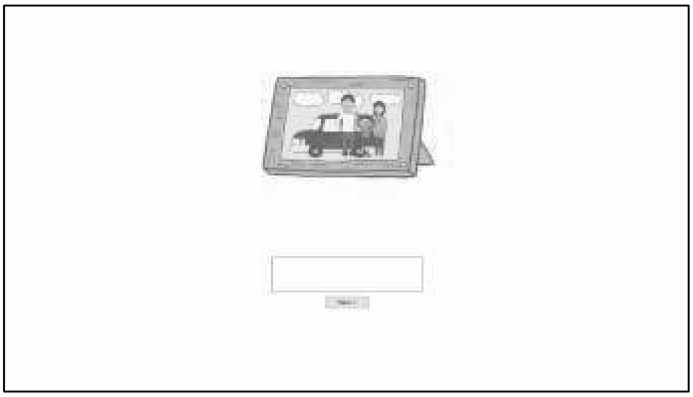
A pictorial cue for *dallog* (newspaper)

A pictorial cue for *tamall* (sun)

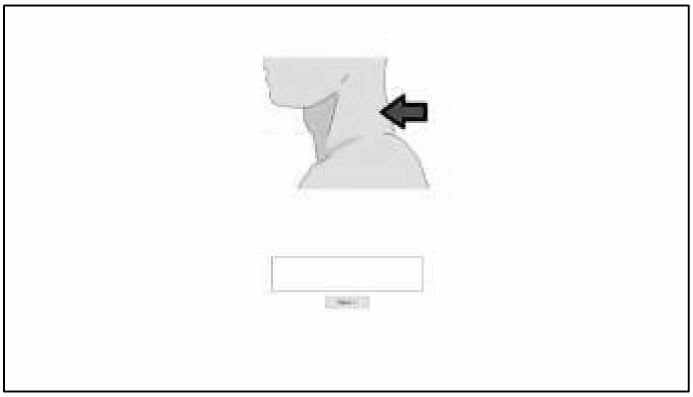


Appendix I (continued)

A pictorial cue for *gogail* (photograph)



A pictorial cue for *olagon* (neck)

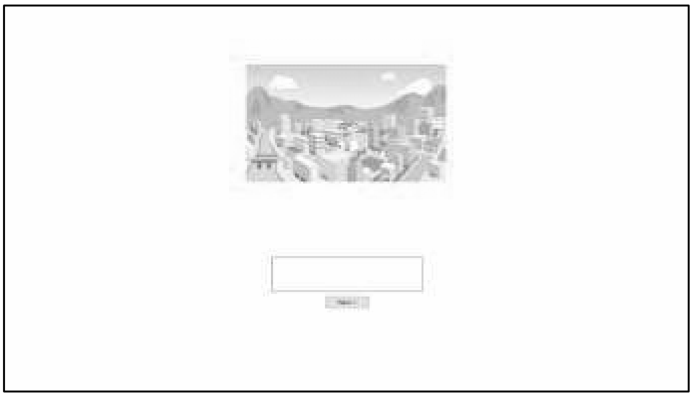


A pictorial cue for *baitin* (clothes)

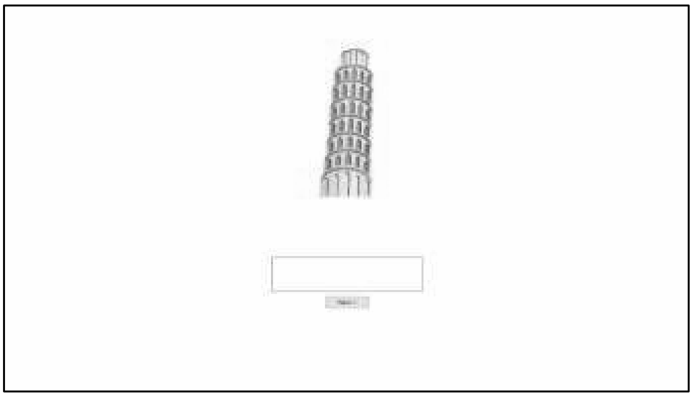
Appendix I (continued)

A pictorial cue for *fecker* (boy)

A pictorial cue for *mullan* (town)



A pictorial cue for *scolog* (tower)



Appendix J

Intercorrelations of the Present Tasks

Intercorrelations of Performance Based on Easy Conditions of Study1 and Study2 Tasks (Easy – Easy), and Vocabulary Size

			Study2 Simple condition				English Vocabulary Size
			Reading Comprehension	Immediate Retention	Delayed Retention	Delayed Transfer	
Study1 Easy condition	Learning		-.16	$.55^{\dagger 1}$	-.05	-.69	-.11
	Immediate	Retention	-.25	.18	-.39	-.07	-.08
	Delayed	Retention	.33	-.15	.06	.00	.22
		Transfer	.51	.63	.87	.87	.86*
English Vocabulary Size			$.55^{\dagger 2}$	.11	.15	.29	–

* $p < .05$, $^{\dagger 1}$ $p = .051$, $^{\dagger 2}$ $p = .052$

Intercorrelations of Performance Based on Difficult Conditions of Study1 and Study2 Tasks (Difficult – Difficult), and Vocabulary Size

			Study2 Complex condition				English Vocabulary Size
			Reading Comprehension	Immediate Retention	Delayed Retention	Delayed Transfer	
Study1 Difficult condition	Learning		-.14	.03	.63	-.54	.17
	Immediate	Retention	-.05	-.02	-.21	-.54	-.06
	Delayed	Retention	.50	-.87	.00	.00	.50
		Transfer	-.64	-.11	.00	-.50	.25
English Vocabulary Size			.31	$.64^{\dagger}$	.80	.73	–

† $p = .063$

Appendix K

Types of Processes Involved in the Present Tasks

Types of Processes Involved in the Present Tasks

	Study1 Task		Study2 Task	
	Easy	Difficult	Simple	Complex
Learning				
Pre-learning	–			
Type of Processing	–		Data-driven	
Production	–		(-)Implicit*	
Learning / Training Task				
Type of Processing	Data-driven	Conceptually-driven	Conceptually-driven	
Production	Implicit	Explicit	-Generation	+Generation
Immediate Retention Posttest				
Type of Processing	Data-driven	Conceptually-driven	Data-driven	
Production	Implicit	Explicit	(-)Implicit*	
Delayed Retention Posttest				
Type of Processing	Data-driven	Conceptually-driven	Data-driven	
Production	Implicit	Explicit	(-)Implicit*	
Delayed Transfer Posttest				
Type of Processing	Conceptually-driven	Data-driven	Data-driven	Conceptually-driven
Production	Explicit	Implicit	(-)Implicit*	Explicit

* *(-)Implicit* means that production relies largely on implicit knowledge but in a lesser degree than those when responding in L1.