ELİF CEMRE SOLMAZ SYSTEMATIC ANALYSIS OF CATEGORY LEARNING Bilkent University 2019

A SYSTEMATIC ANALYSIS OF CATEGORY LEARNING: EFFECTS OF DIFFERENT LEARNING STRATEGIES AND TASK CHARACTERISTICS

A Master's Thesis

by

ELİF CEMRE SOLMAZ

Department of Psychology
İhsan Doğramacı Bilkent University
Ankara

June 2019

To my family who supports me in all circumstances

A SYSTEMATIC ANALYSIS OF CATEGORY LEARNING: EFFECTS OF DIFFERENT LEARNING STRATEGIES AND TASK CHARACTERISTICS

The Graduate School of Economics and Social Sciences

of

İhsan Doğramacı Bilkent University

by

ELİF CEMRE SOLMAZ

In Partial Fulfillment of the Requirements for the Degree of

MASTER OF ARTS IN PSYCHOLOGY

THE DEPARTMENT OF PSYCHOLOGY

İHSAN DOĞRAMACI BİLKENT UNIVERSITY

ANKARA

June 2019

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Psychology.

Asst. Prof. Dr. Miri Besken

Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Psychology.

Asst. Prof. Dr. Gül Günaydın

Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Psychology

----

Assoc. Prof. Dr. Mine Mısırlısoy

Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

----

Prof. Dr. Halime Demirkan

Director

ABSTRACT

SYSTEMATIC ANALYSIS OF CATEGORY LEARNING

Solmaz, Elif Cemre

M.A., Department of Psychology

Supervisor: Asst. Prof. Dr. Miri Besken

June 2019

Many studies showed that organization of study materials has a strong effect on learning performance (Kornell & Bjork, 2008). The current study compared category learning performance through blocked and interleaved learning conditions, using verbal (Experiment 1) and pictorial (Experiment 2) materials. Participants were assigned to one of three conditions at encoding phase: blocked, interleaved, and semi-interleaved learning conditions. In blocked learning, participants studied four exemplars of the same category within the same trial, and categories were blocked across trials. In interleaved learning, participants studied one exemplar from four different categories within the same trial, and categories were interleaved across trials. In semi-interleaved learning, participants studied four exemplars of the same category within the trial, but categories were interleaved across trials. At testing phase, participants were tested on old and novel exemplars of the categories that they studied and were asked to identify the category of each exemplar. Lastly, they were tested on explicit understanding of categories. Both Experiment 1 and 2 revealed that participants produced the highest learning performance in the semi- interleaved learning condition. Learning similarities within the same trial and learning differences across trials might lead to the most optimal learning strategy for category learning, regardless of the type of stimuli used.

Keywords: Categorization, Effective Learning Strategies, Induction, Memory, Spacing Effect

ÖZET

KATEGORİ ÖĞRENİMİNİN SİSTEMATİK ANALİZİ

Solmaz, Elif Cemre

Yüksek Lisans, Psikoloji Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Miri Besken

Haziran 2019

Çalışılan materyallerin nasıl organize edildiğinin öğrenme üzerindeki etkisi birçok çalışma tarafından gösterilmiştir (Kornell & Bjork, 2008). Bu çalışma, dönüşümlü öğrenme (interleaved learning) ve blok öğrenme (blocked learning) yöntemlerini sözsel (Deney 1) ve görsel (Deney 2) materyaller kullanarak karşılaştırmıştır. Çalışmanın ilk aşamasında katılımcılar blok, dönüşümlü ve yarı dönüşümlü olmak üzere üç çalışma koşulundan birine atanmışlardır. Blok öğrenme koşulunda, katılımcılar her çalışma denemesinde bir kategorinin dört örneğini gördüler ve denemeler arasında da kategori örnekleri sırayla sunuldu. Dönüşümlü öğrenme koşulunda, katılımcılar her deneme ekranında dört kategorinin birer örneğini gördüler ve örnekler arasında da kategoriler dönüşümlü olarak gösterildi. Yarı dönüşümlü öğrenme koşulunda ise, katılımcılar örnekler içinde aynı kategorinin örneklerini gördüler, ama denemeler arasında kategoriler

dönüşümlü olarak gösterildi. Test aşamasında katılımcılar çalıştıkları ve çalışmadıkları örnekler için test edildiler. Deneyin son kısmında kategorilerle ilgili açık öğrenme (explicit learning) konusunda test edildiler. Deney 1 ve Deney 2'nin sonuçları yarı dönüşümlü öğrenme yönteminin en yüksek öğrenme performansı sağladığını göstermiştir. Yarı dönüşümlü öğrenme yöntemi, bir kategori içindeki benzerliklerin ve kategoriler arasındaki farklılıkların daha kolay fark edilmesini sağlamaktadır. Bu nedenle, yarı dönüşümlü öğrenme yöntemi blok ve dönüşümlü öğrenme yöntemlerine göre daha başarılı bir kategori öğrenimi sağlamaktadır.

Anahtar Kelimeler: Bariz Öğrenme, Bellek, Etkili Öğrenme, Kategori Öğrenimi ,Tümevarım

ACKNOWLEDGMENTS

First thing first, I would like to express my gratitude with most sincere honesty to Dr. Miri Besken. I have learnt everything I know from her in terms of academic life for four years. She will always be more than an advisor for me. In addition, I express my appreciation to Dr. Gül Günaydın for her support who teaches courses that I will never forget during my life.

I would also like to express gratitude to my mother, father, my lovely sisters Ceren and Ceyda for always believing in me and encouraging me.

I owe to special thanks to Öykü Çiftçi for being my biggest supporter, best roommate and best colleague during my master life. I am also grateful to Dilay Seda Özgen for being my roommate and being there whenever I need. I would like to thank my VIP team who are Hanzade Tepeoğlu, Doğukan Kaya, Durukan Güven, Berka Tanyeli, Beyza Özgöde, Mehmet Günçavdı and Eren Bilaloğlu. They made my Bilkent life unforgettable.

I express my appreciation to my colleagues Irmak Tuğcu and Cansu Gökçe who we started this psychology journey together.

I owe to special thanks for all their support to Burcu Seydioğlu, Okan Ok and Abdülkerim Satuk Buğra İsmailefendioğlu who have been with me since the first day of my Bilkent life.

I express my appreciation to my childhood friends; Dilara Yılmaz, Elif Kürüm, Havva Zeynep Koç, Emine Koç who are the people who made me the person that I am.

I would like to thank to all, Ecem Eylül Ardıç, Gamze Nur Eroğlu & Ezgi Melisa Yüksel for their emotional support. I also would like to thank to all of the undergraduate students of Cognitive Psychology Lab for helping me with my thesis.

Lastly, I also would like to thank myself for bringing this process to the end with my patience and effort.

TABLE OF CONTENTS

LIST OF TABLES

LISTS OF FIGURES

CHAPTER I

INTRODUCTION

Which learning strategy produces better learning performance is a debated issue for a long time. One of the controversial issues is interleaving vs. blocking of study materials. Blocked learning strategy refers to a study method, in which one studies the same type of materials consecutively before moving onto the next type of material. For example, when there are three items (A, B and C) to study, blocking learning strategy would require one to study the materials in the following format: A_1 A_2 A_3 B_1 B_2 B_3 C_1 C_2 C_3. On the other hand, interleaved learning strategy refers to studying materials in a mixed order. An example of this format is as follows: A_1 B_1 C_1 A_2 B_2 C_2 A_3 B_3 C_3 (Kurtz & Hovland,1956). Previous research generally shows that students do not choose effective learning strategies during their study sessions. For instance, although most of the research showed that interleaving study materials produces higher memory performance than blocking them, most of the students believe that blocking is a better learning strategy than interleaving (Cohen, Yan, Halamish & Bjork, 2013). On the other hand, students may not be choosing a specific learning strategy all the time, because there have been contradictory findings in literature. Some studies find that interleaved studying produces higher learning (Birnbaum, Kornell, Bjork & Bjork, 2012; Kornell & Bjork, 2008), whereas some others find that blocked learning leads to better memory performance (Carpenter & Mueller, 2013; Sorensen & Woltz, 2016).

The first aim of the current study is to contribute to this literature through manipulating study order, using blocked, interleaved and semi-interleaved learning strategies. Previous research also reveals that interleaved learning strategy generally produces superior learning for pictorial materials, whereas blocked learning strategy is associated with superior learning of verbal materials. Therefore, the second aim of the current study is to investigate how nature of the study materials (words vs. pictures) affect learning performance.

1.1 Effective Learning Strategies and Category Learning

How to study effectively has been a very controversial issue among students for many years. A branch of educational psychology investigates effectiveness of different learning strategies (Rohrer & Taylor, 2007; Mitchell, Nash, & Hall, 2008; Olina, Reiser, Huang, Lim & Park, 2006). Studies in this area have shown that almost every student uses a different strategy to study. Because different study strategies provide different levels of learning performance, some students are more successful than others. Thus, some students use more effective learning strategies naturally, whereas a majority of them are not aware of their ineffective learning strategies (Hartwig & Dunlosky, 2012).

One of the most investigated learning strategies is *spacing* of study materials. Spacing is a widely investigated issue, because the organization of study materials like blocking or interleaving have a strong effect on learning performance (Kornell & Bjork, 2008). In blocked learning strategy, stimuli that belong to the same concepts are studied consecutively. On the other hand, in interleaved learning strategy stimuli from different concepts are studied together that they are intermixed together (Kurtz & Hovland, 1956). Interesting point in spacing (vs blocking) study strategy is that generally students cannot accurately evaluate which learning strategy is more effective. Thus, there is a discrepancy between people's metacognitive judgments for learning strategies and their actual memory performance. Previous research showed that people's beliefs about effective learning strategies are against some scientific findings. For instance, Cohen et al. (2013) showed that participants choose to block more often than interleaving although interleaved learning strategy provides better learning performance in most cases. In addition, Kornell & Bjork (2008) and Hartwig & Dunlosky (2012) replicated this finding. Therefore, finding which learning strategy produces the most optimal learning performance is important in terms of raising students' awareness.

The effectiveness of blocked and interleaved learning strategies has been measured through different materials, such as mathematical problems (Rohrer et al, 2009), words (Carpenter & Mueller, 2013; Sorensen & Woltz, 2016) and visual materials (Kornell & Bjork, 2008; Birnbaum, Kornell, Bjork & Bjork, 2012). Most of those studies use study materials through creating categories, because when participants see categories, they have to identify differences between categories and shared characteristics within categories. Sana, Yan & Kim (2017) also argues that boundaries of category and features that differ one category from another category should be learnt successfully when separate categories are being learned. Understanding commonalities that are shared by all exemplars of a category is also an important point in category learning. Therefore, while studying categories, people should make inferences about given categories. Likewise, with category learning, participants' explicit understanding about stimuli can be measured because they should apply the rules to novel exemplars (exemplars belong to same categories that they do not study during encoding phase) that they inferred from studied exemplars. In addition, Carvalho & Goldstone (2014) argues that in category learning, how information is organized has huge impact on what is successfully learnt. Therefore, category learning can provide useful information about which learning strategy provides better memory performance than other learning strategies.

One of the initial studies that investigated effectiveness of interleaving vs. blocking was conducted by Kurtz & Hovland (1956). They used geometric shapes as category materials that differed in color, shape, size and position. During encoding phase, experimenters showed four stimuli to participants and they were asked to study different type of exemplars either in interleaved or blocked learning strategies. After encoding phase, participants were asked to report common features of each category verbally. If participants gave wrong answers, experimenters repeated the trials until participants reported the true categorization features. Then, participants were tested with an immediate recognition test, in which they were presented with four different exemplars and asked to categorize exemplars. Results revealed that blocked learning produced better category learning than interleaved learning condition. In recent studies, procedure is slightly different than past studies. The procedure of more recent studies consists of three main phases: encoding, distraction and testing phases. During the encoding phase, participants are presented with a stimulus from one category and they are asked to classify presented stimulus. It is followed by feedback for each response they give. The learning strategy is usually manipulated by the experimenter. Participants may be exposed to blocked or interleaved studying strategy in a within or between subject design. In the second part, participants are asked to do some trivia task to distract their attention for two or three minutes. The last phase is the testing phase. During the testing phase, participants are tested on implicit and explicit understanding about categories. Implicit understanding is generally tested by testing studied exemplars. Explicit understanding is tested through using novel exemplars and also free recall test questions about the characteristics of the previously studied categories.

1.2 Research that Found Evidence for Superiority of Interleaved Learning Strategy

Like students, experimenters that investigate effectiveness of interleaved and blocked learning also have not come to a consensus yet. There are contradictory results about effectiveness of blocking and interleaving. There are many studies that found support for the superiority of interleaved learning strategy over blocked learning. For instance, as one of the initials studied that revealed higher learning performance for interleaved learning condition was conducted by Kornell and Bjork (2008). They tested blocked and interleaved learning strategies by using category learning. They used six paintings from twelve different artists as categories. In Experiment 1A, participants studied items both in blocked and interleaved learning conditions in a within subjects design. In Experiment 1B, participants were assigned to either blocked or interleaved learning conditions in a between subjects design. Consequently, all participants were tested on novel exemplars in testing phase. The results of the study revealed that participants' performance was better in interleaved learning condition than blocked condition.

There are a large number of other research that support advantage of interleaved learning strategy. One of the most known studies that was conducted by Kang and Pashler (2012) in which they also used paintings of Kornell & Bjork (2008) as category materials. They conducted two experiments. In Experiment 1, Kang and Pasher (2012) used temporally spaced learning condition in addition to massed and interleaved conditions. In the temporally spaced leaning condition, there were ten seconds delay between exemplars. Moreover, they added another learning condition in which they provided participants four paintings of the same artist at the same time rather than displaying each painting on the screen sequentially, referred to as simultaneous massed condition. Experiment 1 revealed that interleaved condition produced higher learning performance at test than all other three conditions. In Experiment 2, they tested the difference between massed, interleaved and simultaneous different conditions, in which participants were presented with four different paintings from four different artists on the same screen. Experiment 2 showed that simultaneous different condition produced higher learning than interleaved condition. In addition, interleaved condition produced higher learning than massed condition.

In another study, Rohrer, Dedrick & Burgess (2014) used mathematics problems as material in their study, because they claimed that generally in the math problems, same kind of problems and strategies are taught to children and while they solve same kind of problem at the same time which is also known as blocked learning, they do not need to think about the problem and they just apply the rule to question. Thus, they tested 7^{th}-grade students about math problems which were presented in blocked and interleaved condition in a within-subject design. Results of the study showed that interleaved learning provided an average performance twice as high as blocked learning. Similar to Rohrer et al (2014) study, Taylor and Rohrer (2010) investigated spacing effect with using math problems. Results showed superiority effect of interleaved learning condition.

In addition to intervening of different stimuli, Birnbaum et al. (2012) argued that interleaving provides superior memory performance than blocked learning, not because of the temporal spacing, but because of that participants can grasp the different features of categories in interleaved learning more easily than blocked learning condition, because exemplars of different categories intervene with each other. They conducted an experiment, using different bird and butterfly family exemplars. In Experiment 1, there were three conditions which measured the difference between temporal spacing and alternating between stimuli; contiguous (no spacing between exemplars), alternating trivia (irrelevant questions before every photo) and grouped-trivia (eight photos then eight trivia questions). All participants studied exemplars in one of these interleaved learning conditions. Results showed that the accuracy of the participants was higher in the grouped trivia condition as compared to alternating-trivia and contiguous conditions. Thus, interleaving the stimuli provided better memory performance than temporal spacing. In Experiment 2, participants studied items in blocked and interleaved conditions, but they were either in contiguous or temporally- spaced (delayed -10s- then trivia questions) conditions. Results revealed that interleaved learning conditions produced superior category learning than blocked learning conditions. In contrast, in temporally spaced condition, there was no significant difference between blocked and interleaved learning strategies. Therefore, they contended that the effect of interleaving is not related to time. Spacing effect occurs, because exemplars intervene with each other, which allows participants to compare and contrast the exemplars from different categories more efficiently. In a similar line of work, Zulkiply and Burt (2013) conducted a study, in which they used Kornell and Bjork's (2008) paintings. In the experiment, there were four conditions which were blocked-immediate, interleaved-immediate, blocked-temporally spaced and interleaved-temporally spaced learning conditions. During blocked-temporally spaced and interleaved temporally-spaced conditions, presentation of each exemplar were followed by thirty seconds unrelated task. Through these conditions, they tried to find difference between temporally spaced and interleaving study strategies. Results showed that interleaved-immediate and interleaved-temporally spaced conditions enhanced memory more than blocked-immediate and blocked-temporally spaced conditions. Related study, Zulkiply, McLean, Burt & Bath (2012) also used psychopathological disorders. Participants studied different disorder cases and then they were tested on new cases and Zulkiply et al. (2012) found that interleaved learning provided better induction than blocked learning for understanding different psychopathological disorders. This study also showed importance of interleaving for category learning.

Studies which produce more efficient and higher performance for interleaved learning strategy than blocked learning have been explained through *discriminative-contrast-hypothesis,* which is also known as *spacing effect*. *Discriminative-contrast-hypothesis* claims that when studied stimuli intervene with each other, it provides better learning performance than blocked learning (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006). *Discriminative-contrast-hypothesis* explains superiority of interleaved learning strategy with that when participants see one exemplar from different categories, they can distinguish differences between categories easier than blocked learning strategy. Since participants can distinguish between category differences more easily than blocked learning strategy, it provides more induction than blocking (Carvalho & Goldstone, 2014). Sana et al. (2017) claimed that interleaved learning strategy provides higher memory performance than blocked learning strategy, because in interleaved learning participants see different exemplars in mixed order, which leads to discrimination of salient features of categories more easily than blocked learning category. In blocked learning, understanding salient features of categories is harder. For instance, Taylor and Rohrer (2010) conducted a study which explains superiority of interleaved learning strategy. They asked participants solve mathematical questions which they had to make calculation about four dimensions (edge, face, angle and corner) of prisms. Participants were assigned to interleaved or blocked learning conditions and results revealed that interleaved learning strategy provided better problem solving than blocked learning strategy. In addition, Taylor and Rohrer reported that participants in blocked learning strategy made more discrimination errors than interleaved learning strategy while solving problems in testing phase. Thus, *discriminative-contrast-hypothesis* supports the superiority of interleaved learning strategy over blocked learning strategy, because of its ability to provide easy discrimination between different features of different category members leading to more efficient category discrimination.

1.3 Research that Found Evidence for Superiority of Blocked Learning Strategy

However, s*pacing effect* is not applicable for every circumstance. In contrast to *spacing effect*, there are many other studies that found advantage of blocked learning strategy. Foss (1968) conducted a study by using linguistic systems decades ago. Results revealed that blocked learning produced higher learning performance than interleaving of study materials. A recent study conducted by Sorensen and Woltz (2016) also found evidence for superiority of blocked over interleaved learning strategy. They used verbal category materials and generated four conditions which were high blocked, medium blocked, low blocked and all interleaved. In the testing phase, participants were tested for studied and novel exemplars. At the end of the experiment, they were asked to predict the actual category names for the categories with the artificial names that they studied at the beginning of the experiment. Results of the experiment showed that in all-interleaved condition participants made more errors than other conditions. Also, all-interleaved condition produced lower learning performance than other conditions, specifically for novel category exemplars, revealing that participants in the all-interleaved condition had lower general understanding of exemplars and actual category names. Therefore, this study showed that interleaved learning condition may sometimes produce lower performance, compared to blocked learning condition when verbal materials are used.

In a similar line of work, Carpenter and Mueller (2013) tested blocked and interleaved learning strategies, using French vocabulary as their study material. They conducted four experiments that included both within and between subject designs, multiple choice test (recognition) and recall test. In addition, they manipulated level of exposure to stimuli during experiments. Participants were presented four exemplar words for eight different pronunciation rules. Participants were native English speakers and they were asked to study the pronunciation of some French words. Results of the study revealed that in all conditions and all experiments blocked learning strategy provided better word pronunciation learning than interleaved learning strategy.

Studies that found superiority effect for blocked learning strategy usually explain the findings through *massing-aids-induction-hypothesis*. *Massing-aids-induction-hypothesis* claims that blocked learning strategy provides higher category learning than interleaved learning because participants can easily discriminate shared characteristics of exemplars belonging to the same category. Therefore, massing-aids-induction-hypothesis argues that blocked learning condition enables comparison within categories more efficiently than interleaved learning strategy. People can realize common features of the category easily, therefore they can make abstraction about what a category represents more easily than interleaved learning strategy (Sana et al., 2017).

1.4 Research that Found Evidence for Both Interleaved and Blocked Learning Strategies

It is important to emphasize that not all studies produce a definitive finding for the superiority of one learning strategy over the other for interleaved or blocked learning. Some of the studies that used slightly different methodology find evidence for conditions in which either of the learning strategies may be more effective. For instance, level of similarity between and within categories may sometimes change which learning strategy will be more effective for category learning. For instance, Carvalho and Goldstone (2014) conducted a study to test interleaved vs. blocked learning strategies. They hypothesized that superiority of interleaved learning condition may originate from different levels of similarity between the presented categories and also because of the delayed study opportunity of interleaved study. In experiment 1, they used “Blobs” which are some artificial shapes. They manipulated similarity in categories (low vs. high) in a between-subjects design. They also manipulated study condition (interleaved vs. blocked) and test type (immediate vs. delayed) in a within-subjects design. Results revealed that in overall, blocked learning was more beneficial than interleaved study, low similarity better than high similarity. In addition, those participants who were presented with low similarity categories improved more in category learning than participants who were presented with high similarity in interleaved learning strategy better than blocked learning strategy.

Sana et al. (2017) also tested effect of interleaving (vs. blocking) by using statistic chapters. They found that when stimuli were juxtaposed in mixed order, interleaved learning strategy produced higher learning performance than blocked study (Experiment 1), but when stimuli were spaced temporally, spacing effect reduced (Experiment 2). They found that when stimuli were presented three-at-a-time rather than one-at-a-time, also blocked learning strategy provided more induction than interleaved learning. This is consistent with *massing-aids-induction-hypothesis*, because when categories are presented three-at-a-time, shared characteristics of the same category are figured out easier. In a similar line of work, Pan, Tajran, Lovelett, Osuna and Rickard (2018) found that using blocked or interleaved learning strategy may benefit learners at different rates, depending on if the material is repeated or presented only once. They presented Spanish words English speakers to teach them how to conjugate. When stimuli were presented multiple times, interleaving was more beneficial for learning over massed learning. In contrast, when stimuli were presented with no repetition, blocked learning strategy provided superior learning performance.

As mentioned in the foregoing literature review, Zulkiply and Burt (2013) found that interleaving the stimuli helps category induction when the discriminability of categories is low (Experiment 2). They used categories that consisted of different kinds of shapes color and different levels of discriminability as low or high discriminability. Results showed when discriminability was high between categories, blocked condition provided higher learning than interleaved condition. As the massing-aid-induction-hypothesis claims, blocked learning strategy helped participants to abstract the features within the same category more easily. In addition, because within category similarity was high, blocking produced more learning than interleaving.

Table 1. Findings of previous research

Study	Material	Findings
Kornell & Bjork (2008)	Paintings	Interleaved > Blocked
Kang & Pashler (2012)	Paintings	Interleaved > Blocked
Birnbaum et al. (2013)	Birds & Butterflies	Interleaved > Blocked
Carpenter & Mueller (2013)	French Words	Blocked > Interleaved
Sorensen & Woltz (2016)	Words	Blocked > Interleaved
Sana et al. (2017)	Statistic Chapters / 1-at-a-time	Interleaved > Blocked
	Statistic Chapter / 3-at-a-time	Blocked > Interleaved
Pan et al. (2018)	Spanish Words w/ repetition	Interleaved > Blocked
	Spanish Words w/ no repetition	Blocked > Interleaved

Besides blocking and interleaving of studying, people's mental effort is also effective about impact of learning strategies. Carvalho and Goldstone (2015) conducted a study to investigate the attentional mechanisms behind interleaved and blocked learning. To study attentional mechanisms, they used passive versus active learning method in their experiments. In passive learning, participants are provided with all information about stimuli. On the other hand, in active learning, participants are provided with parts of the information about stimuli and they are expected to make inferences about the given stimuli. In Experiment 1, they used different kinds of *Fribbles*, artificial objects which look like animals with different extensions that make them belong to separate categories. Participants were assigned either passive learning condition or active learning condition. Experiment 1 produced no main effect for presentation of stimuli (blocked vs interleaved), type of item (studied vs novel) and study group (passive vs. active), but they found an interaction between presentation of stimuli and study group. In active learning condition, interleaved learning produced higher category learning than blocked learning condition, whereas, in passive learning condition, blocked learning produced superior performance than interleaved condition. In Experiment 2, they used categories which had high similarity between categories but low similarity within categories. Experiment 2 replicated the results of Experiment 1. Thus, the superiority of the learning strategy may also change, depending on the level of activity of participants during encoding.

1.5 Present Study

The first aim of the present study is to investigate effectiveness of spacing (vs blocking). Therefore, participants were presented with materials in three different learning strategies: blocked, interleaved and semi-interleaved. In blocked learning strategy, participants are presented with stimuli consecutively, whereas in interleaved learning strategy are presented stimuli by mixed together. In addition, semi-interleaved learning strategy includes a hybrid combination of both blocked and interleaved learning strategies which are mentioned in detail in the method section. Consistent with previous experiments, current experiment also made use of categories as material. The method used in the current experiments was similar to that of Sorensen and Woltz (2016). Sorensen and Woltz presented participants words in six separate conceptual categories. For example, participants could be presented with exemplars of a category ‘something that is loud’, with eight exemplars in each category. “Something that is loud” category included “trumpet, thunder, cymbals, siren, jackhammer, shout, scream and dynamite”. Additionally, exemplars were presented with non-word category names instead of actual category names name such as “HEECE” for “something that is loud”. Thus, they were expected to make inferences for actual category names.

Experiment of Sorensen and Woltz (2016) consisted of three parts which were encoding, distraction and testing phases. During encoding phase, participants assigned to interleaved learning condition which they were presented with one exemplar from four different categories with non-word category names in each trial. On the other hand, participants assigned to blocked learning condition which they were presented with four exemplars of the same category in one trial. Participants were asked to study for exemplars and try to guess actual category names for the testing phase. Distraction phase consisted of letter sequencing exercises. Testing phase included two parts: exemplar testing and category testing. In exemplar testing, a recognition test for studied and novel exemplars was used. For explicit category testing, participants were asked explicitly what each non-word category name symbolized.

The present study is a modification on Sorenson and Woltz (2016). Consistent with method of Sorensen and Woltz (2016), twelve categories were formed for encoding part of present experiment. Each category consisted of eight exemplars. Categories were formed as abstract categories which means that they were not natural categories like bird families. For instance, in the current study, one category was “something that is white”, including the exemplars “cloud, whiteboard, wedding dress, swan, milk, teeth, cotton, polar bear”. Therefore, categories did not have natural bonding, but exemplars of each category shared some commonality. In addition, non-word category names for each category were formed and categories were presented with those non-word category names rather than actual category names. To test implicit learning performance of participants, they were tested on studied exemplars and novel exemplars during testing phase. Finally, participants were asked to make predictions about actual category names of categories that they studied during encoding phase. Participants’ performance on novel exemplars testing and also predictions show that which learning strategy provides better explicit learning.

With three learning conditions, the present research aims to test *massing-aids-induction-hypothesis* and *discriminative-contrast-hypothesis*. According to massing-aids-induction-hypothesis, blocked learning strategy in both experiments should produce higher learning performance at test than the other two conditions, because participants should understand within category similarities better than others. Therefore, they should have an abstraction for the presented category at the end of studying each category. On the other hand, according to discriminative-contrast-hypothesis, interleaved learning strategy should provide better learning performance than other conditions, because participants should discriminate differences between categories easier because they see four exemplars from different categories in the same trial. The current research hypothesized that semi-interleaved learning strategy which includes both blocked and interleaved learning conditions should produce the best performance compared to other conditions, because semi-interleaved learning conditions can benefit from both within category abstraction of blocked learning and between category discrimination of interleaved learning. Therefore, participants should easily understand what common point of presented category and also what differentiates the presented category from the next category that it is interleaved with.

As can be inferred from the foregoing literature review, generally studies that used pictorial material found superiority effect for interleaved learning strategy over blocked learning strategy (Kornell & Bjork, 2008, Zulkiply et al., 2011; Zulkiply & Burt, 2013; Rohrer et al, 2012; Birnbaum et al., 2012), whereas, studies that used words as material found superiority effect for blocked over interleaved learning strategy (Carpenter & Mueller, 2013; Sorensen & Woltz, 2016; Pan et al., 2018b) (See *Table1*). Like mentioned in research that found evidence for both interleaved learning and blocked learning strategy, when materials or procedure of experiments are changed, benefit of interleaved and blocked learning strategies may also change. Therefore, the present study is conducted for systematic comparison of verbal and pictorial materials in different learning conditions. Although there are many studies used pictorial and verbal material, to our knowledge, there are no studies that directly investigated the difference between words and pictures in the literature. In the current study, I presented participants with words in Experiment 1 and pictures in Experiment 2. I hypothesized that if learning conditions affect learning performance independent from what is studied, there should not be learning performance difference between Experiment 1 and Experiment 2. On the other hand, if Experiment 1 and Experiment 2 will produce significantly different results, it can be explained by nature of study materials

CHAPTER II

EXPERIMENT 1

2.1 Method

2.1.1 Participants

Based on a power analysis, a hundred and two participants were required to detect a large effect size *f* of .42 to have a 95% power when the traditional criterion of .05 for statistical significance is employed. Consistent with power analysis, a hundred and two Bilkent University students, thirty-four for each condition, who are between the ages of 18-30 and native speakers of Turkish participated in the study. Participants were compensated with course credit. All participants were randomly assigned to one of three conditions.

2.1.2 Material and Design

Twelve artificial categories were created for the experiment. Eight exemplars were determined for each category. Artificial categories refer to exemplars which do not have natural bonding. I used these artificially-formed categories instead of natural categories, in which exemplars had natural commonalities, because using naturally bonded categories would make the procedure easier which can create a confound. When participants understood categories easier because of the nature of the study, it can moderate effect of learning strategies. Moreover, since the study is a modification of Sorenson and Woltz (2016), the materials were kept as similar to the original as possible, as they also used artificial categories. Four of the exemplars were presented in the encoding phase and the other four exemplars were used in the testing phase as novel exemplars. Presentation of exemplars was counterbalanced across participants. Exemplars of each category were unique to the determined category, such that they could not be categorized under the description of any of the other categories. For instance, “cloud” from “something that is white” category did not fit to “something that is yellow” or “something that is made of steel” or “something that is made of wooden” (See Appendix A for further details). Also, a non-word category name for each category was formed. Each non-word category name started with a different letter of the alphabet and none of these names had a meaning in Turkish (See Appendix A1 for all non-word category names). Exemplars were presented under those non-word category names during the experiment.

The experiment had a between-subjects design. Participants were assigned to one of the three learning conditions: blocked learning, semi-interleaved learning and interleaved learning. All conditions had three study blocks during the encoding phase. Each block included 4 categories and 4 exemplars from each category. Each exemplar was presented twice during the learning phase. Therefore, each block had 32 trials. In the blocked condition, all categories were presented in order. After each exemplar of one category was presented, participants were asked to study next category and its exemplars. Thus, in blocked learning condition participants saw 4 exemplars of one category within the same display. Then, they moved onto the following category sequentially. In the interleaved learning condition, all categories were distributed randomly within each block. In each trial of interleaved learning condition, participants were presented with four different categories and one of those category exemplars within the same display. Presentation of all categories and their exemplars were randomized across participants. In the semi-interleaved condition, participants were presented one blocked category for the first scene and another blocked category for second scene. Therefore, different categories were presented for each scene. Also, order of categories was randomized across participants.

Two separate lists were created to counterbalance for the study status at test across participants. The first list consisted of 48 words, which were the same words in encoding phase as old exemplars and other lists consists of 48 words belonging to same categories as new exemplars. For example, if the word “cotton”, “polar bear”, “milk", “wedding dress" were presented during the learning phase to half of the participants within category “something that is white”, the rest of the category (teeth, swan, cloud and whiteboard) was only presented during the category inference test. The items that are presented during learning phase are “old” items, whereas the items that have not been presented before at learning phase are “new” items. The study status (old vs. new) was counterbalanced across participants, and these lists were presented to an equal number of participants in each condition.

During the encoding phase, reaction times for first and second presentation for each response to each stimulus were recorded. In the encoding phase, to be consistent with Sorensen & Woltz (2016) that procedure of current research modified by their method, each exemplar were presented for 6 seconds. Similarly, during the testing phase, reaction times for response to old and new exemplars were recorded by a computer program.

Finally, category lists were prepared to measure participants' inferences about relations between exemplars and categories. Those lists included all 12 non-word category names.

2.1.3 Procedure

Participants were tested individually on desktop computers and standard keyboard with an Intel® Core™ processor in the Bilkent University Cognitive Laboratory Area. They were tested through E-prime® software program. The experiment consisted of four parts: encoding, distraction, exemplar testing and category testing. Participants were given different instructions for each learning condition. Participants assigned to blocked condition were instructed that they would be presented with four exemplars of one category sequentially. After all exemplars of one category were presented twice in a row, the program would proceed onto the other category exemplars in the same block. In interleaved learning condition, participants were instructed that they were going to see one exemplar from 4 categories in one trial and exemplars would be changed in each trial. Also, in semi-interleaved condition, participants were told that they would see one category and its four exemplars in one trial and in next trial they would see another category with its four exemplars.

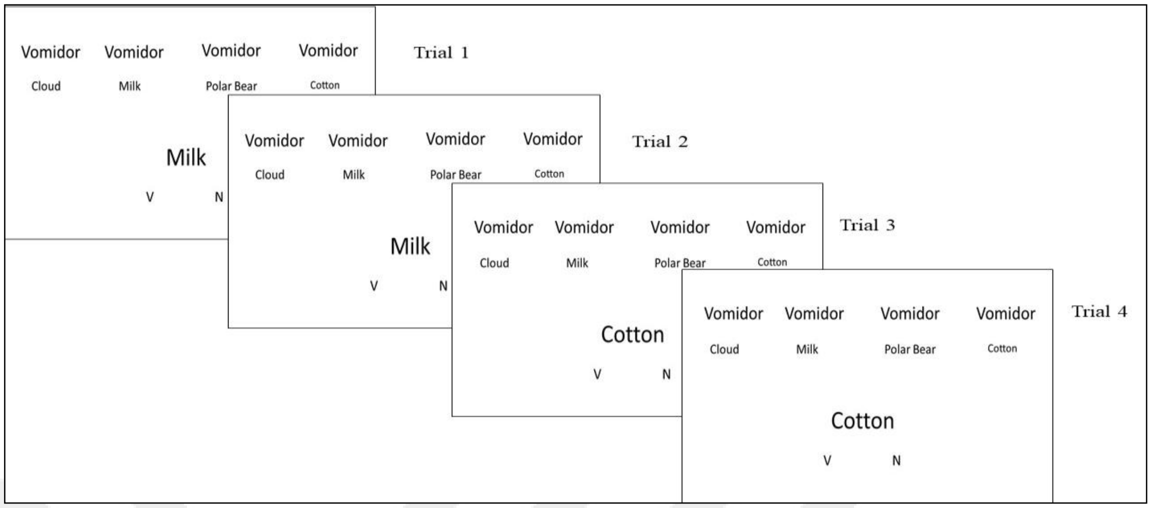

Figure 1. Sample for blocked learning condition

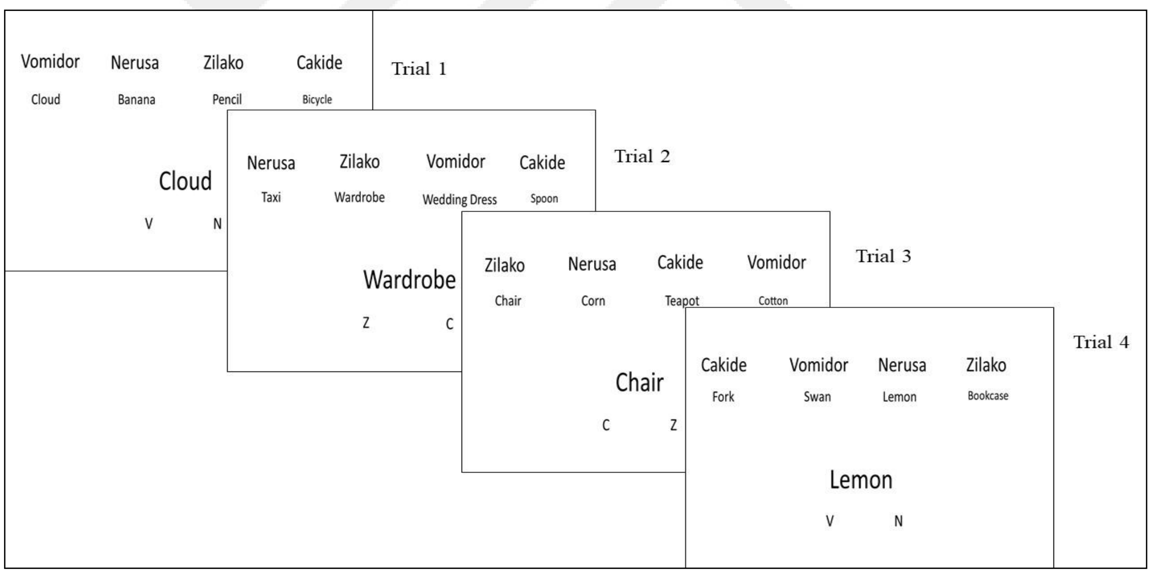

Figure 2. Sample for interleaved learning condition

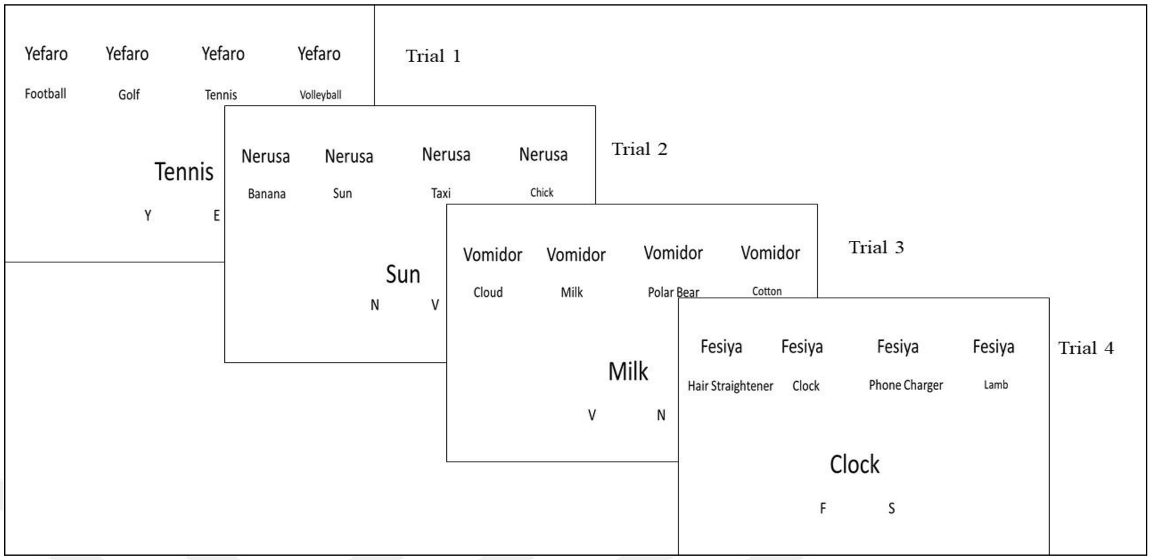

Figure 3. Sample for semi-interleaved learning condition

In the learning phase of the experiment, participants were presented four exemplars with category names at top of the screen. In the middle of the screen, one of exemplars that was presented at the top of the screen was presented as target word. Also, at the bottom of the target word, two letters were presented. One letter was first letter of correct category name and other letter was first letter of one of the other category names. Therefore, participants were asked to press one of the letters from keyboard in 6 seconds. When they pressed one of the keys, feedback was provided about the accuracy of their responses. If they took more than 6 seconds, the program warned them to respond faster. In all conditions, each exemplar was presented two times during encoding phase. Moreover, in the encoding phase of blocked condition, 4 exemplars of same category were presented in one trial. In the blocked condition four same non-word category names were seen at the top of scene (See *Figure1*). Under each category names, exemplars of the category were presented. For instance, in the category of "something is white"; polar bear, milk, cloud and cotton were presented in that trial. In blocked learning condition exemplars of categories were studied one after another. After four exemplars of one category were studied, program proceeded onto another category. In the interleaved condition, 4 different categories and one exemplar from the 4 different categories were presented in one trial. For instance, “something that is white”, “something that is yellow”, “something that is made of wooden” and “something that is made of steel” were presented under with one exemplar for each different category in one trial (See *Figure2*). Other procedure was the same as blocked condition. Exemplars and categories were randomized for each block. In the semi-interleaved condition, trials were the same as blocked conditions, but order of the trials were different (See *Figure3*). Therefore, in the first trial, participants saw “something that is white” category, in second trial, participants saw “something that is yellow category”, in third trial, they saw “something that is made of wood” and in the fourth trial, they saw “something that is made of steel”. Presentation of trials were the same as blocked conditions. Once participants studied all the items within each category, they proceeded onto the distraction phase. Distraction phase was same for all conditions. In the distraction phase, participants were asked to solve arithmetic problems for 3 minutes.

After three minutes, participants proceeded onto the third phase of the experiment, which was the exemplar testing phase. This was identical across all three learning conditions. Exemplar testing phase consists of two parts: testing of the studied items and testing of the novel items. Exemplar testing phase of experiments consisted of recognition test. Therefore, in the testing phase, participants were presented with a word in the middle of the screen, and with the names of four categories from the learning phase and were asked to choose the right category for each word (See *Figure4*). For the first part of the test, participants saw items from the encoding phase which can be classified as “old” items. Participants were asked to categorize them in one of the four categories. Participants were asked to press first letter of correct category name from the keyboard (e.g. V for vomidor). The testing phase was self-paced. The second part of the testing phase included only novel exemplars of the categories (new exemplars) that the participants did not study during encoding phase. Thus, they had to infer the category information.

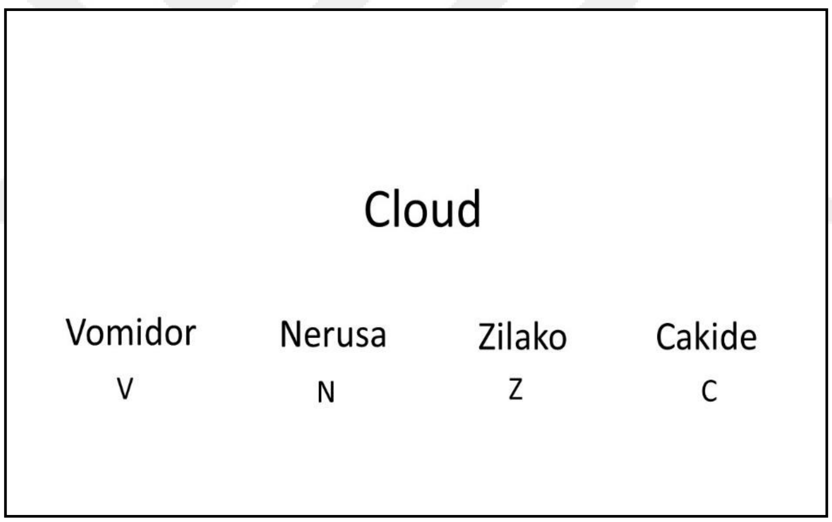

Figure 4. Sample for testing trial

The last part of experiment was explicit testing of category information. Participants were given a paper which had non-word category names on it, and they were asked to think about all exemplars of categories that they studied. They were asked to find actual category names by using common properties of exemplars. The last part of the experiment was self-paced.

2.2 Results

Learning Phase. Alpha level was set to .05 for all analyses. %80 correct answer in encoding phase was determined as a lower limit to include each data in the analyses. All participants successfully chose the correct category in blocked (*M* = .99, *SD* = .12), interleaved (*M* = .99, *SD* =.13) and semi-interleaved (*M* =.98, *SD* = .39) learning phases. Median of each participants' reaction times for encoding phase was calculated, as the median analysis provides more accurate results than the mean analysis for reaction times in terms of finding central tendency. In addition, mean analysis is more affected by outliers and extreme scores (Leys, Ley, Klein, Bernard, & Licata, 2013). For the encoding part of the experiment, mean of medians for each condition was analyzed. In addition, reaction time for the first and the second presentation of target word was analyzed separately. The mean of median RTs was submitted to a mixed analysis of variance (ANOVA), with the presentation time (first vs second) as the repeated measure and the learning condition as the between-subjects variable (blocked, interleaved and semi-interleaved learning conditions). There was a significant main effect of presentation time (first vs second) on reaction time, $F(1,99) = 152.58$, $p < .001$, $MSE = 36666.22$, $\eta_p^2 = .61$. Participants got faster from first presentation (*M* = 2438, *SD* = 606.29) to second presentation (*M* =2107, *SD* = 687.12). There was a main effect of learning condition on reaction times (RT), $F(2,99) = 22.46$, $p < .001$, $MSE = 12580075.59$, $\eta_p^2 = .31$. Therefore, a Bonferroni post-hoc analysis revealed a significant difference between blocked (*M* = 1768, *SD* = 437.84) and interleaved learning conditions (*M* = 2429, *SD* =552.50), $p < .001$, blocked and semi-interleaved (*M* = 2603, *SD* = 585.97), $p < .001$. On the other hand, there was not a significant difference between interleaved and semi-interleaved learning conditions in terms of reaction times to target stimuli, $p = .53$.

Table 2.Reaction times for first and second presentations of exemplars in learning phase

Reaction Time	First Presentation	Second Presentation	Total Presentation
	Mean(SD)	*Mean(SD)*	*Mean(SD)*
Blocked Condition	2026(417.18)	1547(510.04)	1786(437.79)
Interleaved Condition	2558(566.50)	2300(561.50)	2437(553.87)
Semi-interleaved Condition	2731(593.21)	2475(606.92)	2601(566.44)

There was a significant interaction between presentation RT and learning conditions, $F(2,99) = 7.64$, $p = .001$, $MSE = 36666.22$, $\eta_p^2 = .13$. Post-hoc tests using pairwise comparisons were conducted to find reaction time (See *Table2*) differences between first and second presentation of category exemplars within participants. First and second RTs for each condition was separately submitted to a repeated measures ANOVA. In blocked condition, there was a significant difference between first presentation ($M = 2026$, $SD = 417.18$) and second presentation ($M = 1547$, $SD = 510.04$) of exemplars, $F(1,33) = 76.96$, $p < .001$, $\eta_p^2 = .70$. In interleaved condition, there was a significant reaction time difference between first ($M = 2558$, $SD = 566.50$) and second presentation ($M = 2300$, $SD = 561.50$), $F(1,33) = 44$, $p < .001$, $\eta_p^2 = .57$. In addition, in semi-interleaved learning condition, again there was a statistically significant difference between first presentation (M = 2731, SD = 101.73), and second (M = 2475, SD = 104.08) presentation, $F(1,33)$ = 33.30, p <.001, $\eta_p^2 = .50$. Shortly, participants became faster in pressing key from the keyboard in the second presentation of exemplars in all learning conditions.

Testing phase. For testing part of the experiment, testing accuracy of participants, response times to target exemplars, study status (old vs. new exemplars) and category inferences were analyzed. First of all, accuracy of exemplars in testing was submitted to mixed analysis of variance (ANOVA) with study status (old vs. new exemplars) as repeated measure variable and the learning conditions (blocked, interleaved and semi-interleaved) as the between-subject variable. First, there was a significant main effect of study status on testing accuracy, $F(2,99) = 35.37$, $p < .001$, $MSE = .01$, $\eta_p^2 = .26$,. Also, there was a significant main effect of learning condition on testing accuracy, $F(2,99)$ = 5.26, $p = .007$, $MSE = .09$, $\eta_p^2 = .10$. A Bonferroni post-hoc test showed that there was no statistically significant accuracy difference between blocked learning condition (M =.49, SD = .23) and interleaved learning conditions (M = .50, SD = .21), p = 1.00. On the other hand, there was significant difference between blocked (M =.49, SD = .23) and semi-interleaved (M = .64, SD = .21), p = .01 and between interleaved (M = .50, SD = .21) and semi-interleaved conditions, p = .03.

For testing accuracy, a two-way ANOVA analysis revealed that there was a significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved and semi-interleaved), $F(2,99) = 4.24$, , $p = .017$, $MSE = .01$, $\eta_p^2 = .08$,. Moreover, post-hoc tests using pairwise comparison were conducted to find accuracy differences between for old and new testing exemplars within each learning condition

(See *Table3*). The accuracy for exemplars was submitted to a repeated measures ANOVA separately for each learning condition, with the exemplar type (old vs. new) as the repeated measure. Results showed there was no significant difference between old blocked exemplars (M = .50, SD = .24) and new blocked exemplars (M = .48, SD =.23), $F(1,33) = 1.32$, $p = .26$, $\eta_p^2 = .04$. There was a significant difference between old (M = .54, SD = .22) and new exemplars (M = .47, SD = .21) for interleaved learning condition, $F(1,33) = 14.74$, $p < .001$, $\eta_p^2 = .31$. Lastly, there was significant difference between old (M = .69, SD = .22) and new exemplars (M = .59, SD = .22) for semi-interleaved condition, $F(1,33) = 31.93$, $p < .001$, $\eta_p^2 = .49$.

Table 3. Mean accuracy rates for old and new exemplars in exemplar testing phase and the standard deviations in parentheses

Accuracy Rates	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	.50(.24)	.48(.23)
Interleaved Condition	.54(.22)	.47(.21)
Semi-interleaved Condition	.69(.22)	.59(.22)

Two different types of response times were analyzed separately: unconditional response times and response times conditionalized on accuracy. Unconditional analysis included all response times for all exemplars, regardless of whether participants identified the category correctly and incorrectly. For conditional analyses, response times only for trials in which participants identified the categories correctly were included in the analyses. For testing unconditional response times, the mean of response times was submitted to a mixed ANOVA with study status (old vs. new exemplars) as within-subject variable and learning conditions (blocked, interleaved and semi-interleaved conditions) as between-subjects variable. Analysis showed that there was a main effect of study status (old vs new) on response times. Therefore, there was a statistically significant difference between old exemplars (*M* = 3291, *SD* = 1281.35) and new exemplars (*M* = 2716, *SD* = 987.43) for unconditional response times in testing phase $F(1,99) = 47.07$, $p < .001$, $MSE = 358600.54$,$\eta_p^2 = .32$, (See *Table 4* for response times). There was not significant main effect of learning condition on response times, $F(2,99) = 2.33$, $p = .10$, $MSE = 5114730.38$, $\eta_p^2 = .04$,. Likewise, there was no significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved &semi-interleaved) on response times, $F(2,99) = 2.45$, $p = .09$, $MSE = 358600.54$, $\eta_p^2 = .05$.

Table 4.Unconditonal response times for new and old exemplars in exemplar testing part

Response Time	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	3551(1415.64)	2797(1106.97)
Interleaved Condition	3476(1272.09)	2824(1044.08)
Semi-interleaved Condition	2847(1050.14)	2527(783.16)

In conditional analysis, the analyses were conducted in the same way, only analyzing response times for correct answers: the analyses revealed similar results. There was significant main effect of study status on conditional response times, $F(1,99) = 31.67$, $p < .001$, $MSE = 433322.85$, $\eta_p^2 = .24$,. Therefore, there was a statistically significant difference between old exemplars ($M = 3041$, $SD = 1159.62$) and new exemplars ($M = 2523.13$, $SD = 900.89$) for conditional response times in testing phase. On the other hand, there was no significant main effect of learning conditions on conditional response times (See *Table5*), $F(2,99) = 2.74$, $p = .07$, $MSE = 1660674.10$, $\eta_p^2 = .05$, . There was no significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved & semi-interleaved), $F(2,99) = 1.61$, $p = .20$, $MSE = 433322.85$, $\eta_p^2 = .03$.

Table 5.Conditonal response times for new and old exemplars in exemplar testing part

Reaction Time	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	3320 (1336.24)	2579 (936.04)
Interleaved Condition	3149 (1090.75)	2676 (999.37)
Semi-interleaved Condition	2656 (946.41)	2313 (732.19)

Last analysis was conducted to find participants' inferences about categories and exemplars. Each participants' inference for each category was coded by three coders. Participants' inferences that were participants' guesses about actual category names for studied categories were coded as "1" if they could infer the category name correctly and

“0” if the response was incorrect or the response was empty. In addition, response was coded as “.5” participant could understand the concept partially. For example, URTESA was non-word category name for “something that has round shape” and responses like “something that does not have an edge” were considered as “.5” point. After all categories were coded, participants’ scores were summed, and their proportion were calculated by dividing the sum to 12 (number of categories used in the study).

Table 6.Inference rate for actual category names

Inference rate	*Mean(SD)*
Blocked Condition	.32(.26)
Interleaved Condition	.23(.27)
Semi-interleaved Condition	.38(.27)

A one-way ANOVA analysis conducted to find category inferences difference between learning conditions. Analysis revealed that there was not main effect of learning conditions on category inferences $F(1,99) = 2.90$, $p = .06$, $\eta_p^2 = .05$. Therefore, there was not significant differences between blocked, interleaved and semi-interleaved study conditions in terms of finding actual category names for exemplars (See *Table6*).

CHAPTER III

EXPERIMENT 2

3.1 Method

3.1.1 Participants

A hundred and two Bilkent University students same as in Experiment 1, who are between the ages of 18-30 and native speakers of Turkish participated in the study. Participants were compensated with course credit.

3.1.2 Material, Design and Procedure

Material, design and procedure were same as in Experiment 1 with one major difference. In Experiment 1, words were used as experimental material. Differently from Experiment 1, Experiment 2 presented participants with pictures of words. The design and the procedure were identical to Experiment 1 except for the nature of the material.

3.2 Results

Learning Phase. All participants successfully chose the correct category in blocked (M = .98, SD = .04), interleaved (M = .97, SD =.03) and semi-interleaved (M =.99, SD = .01) learning phase. The mean of median RTs was submitted to a mixed analysis of variance (ANOVA), with the presentation time (first vs second) as the within-subject measure and the learning condition as the between-subjects variable (blocked, interleaved and semi-interleaved learning conditions). There was a significant main effect of presentation time (first vs second) on reaction time, $F(1,99) = 208.088$, $p < .001$, MSE =

36904.96, η_p^2 =.68,. Participants got faster from first presentation (*M* = 2063.43, *SD* = 649.94) to second presentation (*M* =1675, *SD* = 635.99). There was a main effect of earning condition on RT, $F(2,99) = 13.38$, $p < .001$, *MSE* = 607482.12, η_p^2 = .22,. A Bonferroni post-hoc analysis was conducted that there was a significant difference between blocked (*M* = 1480, *SD* = 94.52) and interleaved learning conditions (*M* = 2182.07, *SD* = 513.53), $p < .001$, blocked (*M* = 1480, *SD* = 628.10) and semi-interleaved (*M* = 1945, *SD* = 501.86), $p = .002$. On the other hand, there were no significant reaction time differences between interleaved and semi-interleaved learning conditions, $p = .24$.

Table 7. Reaction times for first and second presentations of exemplars in learning phase

Reaction Time	First Presentation *Mean(SD)*	Second Presentation *Mean(SD)*	Total Presentation *Mean(SD)*
Blocked Condition	1654 (628.82)	1306 (659.12)	1480 (629.00)
Interleaved Condition	2254 (538.56)	2109 (509.72)	2182 (513.53)
Semi-interleaved Condition	2281 (590.15)	1609 (452.85)	1945 (501.86)

There was a significant interaction between presentation reaction times and learning conditions, $F(2,99) = 32.48$, $p < .001$, *MSE* =36904.96, η_p^2 = .40,. Post-hoc pairwise comparisons were conducted to find reaction times difference between first and second presentation of category exemplars separately for each learning condition to gain insight about the interaction (See *Table7*). RTs for first and second presentations were entered into a repeated measures ANOVA as the repeated factor. Separate analyses for each learning condition were conducted. In blocked condition, there was a significant difference between first presentation (*M* =1654, *SD* = 628.82) and second presentation (*M* = 1306, *SD* = 659.12) of exemplars, $F(1,33) = 53.20$, $p < .001$, $\eta_p^2 = .62$. In interleaved condition, there was a significant difference between first (*M* = 2254, *SD* = 538.56) and second presentation (*M* = 2109, *SD* = 509.72) for reaction times, $F(1,33) = 15.97$, $p < .001$, $\eta_p^2 = .37$. In addition, in semi-interleaved learning condition, there was a significant difference between first presentation (*M* = 2281, *SD* = 590.15), and second presentation (*M* = 1609, *SD* = 452.85), $F(1,33) = 154.23$, $p < .001$, $\eta_p^2 = .83$.

Testing phase. For testing part of the experiment, testing accuracy of participants, response times to target exemplars, study status (old vs. new exemplars) and category inferences were analyzed. Accuracy of exemplars in testing was submitted to mixed analysis of variance (ANOVA) with study status (old vs. new) as within-subject variable and the learning condition (blocked, interleaved and semi-interleaved) as the between-subjects variable. (See *Table8*). First, there was a significant main effect of study status on testing accuracy, $F(2,99) = 95.10$, $p < .001$, $MSE = .01$, $\eta_p^2 = .50$,. Also, there was a significant main effect of learning condition on testing accuracy, $F(2,99) = 11.01$, $p < .001$, $MSE = .05$, $\eta_p^2 = .18$. A Bonferroni post-hoc test revealed that there was not a significant difference between blocked (*M* = .66, *SD* = .16) and interleaved learning condition (*M* = .63, *SD* = .19), $p = 1.00$. On the other hand, there was a significant accuracy difference between blocked and semi-interleaved (*M* = .80, *SD* = .14) learning conditions, $p = .01$ and between interleaved and semi-interleaved learning conditions, $p <$

.001. For testing accuracy a mixed ANOVA analysis revealed that there was not a significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved and semi-interleaved), $F(2,99) = 2.42$, , $p = .09$, $MSE = .01$, $\eta_p^2 = .05$.

Table 8. Mean accuracy rates for old and new exemplars in exemplar testing phase

Accuracy Rates	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	.69(.19)	.62(.15)
Interleaved Condition	.68(.21)	.57(.18)
Semi-interleaved Condition	.86(.14)	.75(.14)

For response times to the exemplars, conditional and unconditional analyses were conducted like in Experiment 1. For testing unconditional reaction times, a two-way ANOVA analysis was conducted. Analysis failed to show main effect of study status (old vs new), there was not a statistically significant difference between old exemplars and new exemplars for unconditional response times in testing phase $F(1,99) = .001$, $p = .98$, $MSE = 240601.85$, $\eta_p^2 = .00$, . There was a significant main effect of learning condition on response times, $F(2,99) = 9.60$, $p < .001$, $MSE = 1027433.75$, $\eta_p^2 = .16$. A Bonferroni post-hoc analysis was conducted that, there was not a significant difference between blocked ($M = 2586$, $SD = 580.24$) and interleaved learning conditions (M

=2744, *SD* =960.99), p = .10, but there was a significant difference between blocked and semi-interleaved strategy (*M* =2148, *SD* = 530.06) for first and second presentation reaction times, $p < .001$. There was a significant difference between interleaved and semi-interleaved learning conditions in terms of reaction times to target stimuli, p = 003. There was significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved & semi-interleaved) on response times, $F(2,99) = 5.07$, $p = .01$, *MSE* = 240601.85, $\eta_p^2 = .10$,. Also, post-hoc test using pairwise comparison was conducted to find was conducted to find response time (See *Table9*) difference between old and new exemplars through a repeated measures ANOVA, separately for each learning condition. In blocked condition, there was a significant difference between old exemplars (*M* = 2983, *SD* = 651.51) and new exemplars (*M* = 2729, *SD* = 683.49) of exemplars, $F(1,33) = 5.02$, $p = .03$, $\eta_p^2 = .13$. In interleaved condition, there was not a significant response time difference between old (*M* = 2755, *SD* = 1072.74) and new exemplars (*M* = 2734, *SD* = 994.08), $F(1,33) = .03$, $\eta_p^2 = .001$, $p = .88$. In addition, in semi-interleaved learning condition, there was a statistically significant difference between old exemplars (*M* = 2007, *SD* = 600.90), and new exemplars (*M* = 2288, *SD* = 642.21), $F(1,33) = 6.33$, $p = .02$, η_p^2= .16.

Table 9.Unconditional response times for new and old exemplars

Response Time	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	2983 (651.51)	2729 (683.49)
Interleaved Condition	2755 (1072.74)	2734 (994.08)

Table 9 (cont'd)		
Semi-interleaved Condition	2007 (600.90)	2288 (642.21)

In conditional analysis, the analyses were conducted in the same way, only analyzing response times for correct answers. There was not significant main effect of study status on conditional response times, $F(1,99) = 2.80$, $p = .10$, $MSE = 215516.22$, $\eta_p^2 = .03$, but there was significant main effect of learning conditions on conditional response times, $F(2,99) = 9.32$, $p < .001$, $MSE = 768637.06$, $\eta_p^2 = .16$. A Bonferroni post-hoc analysis revealed that there was not a significant difference between blocked ($M = 2585$, $SD = 515.68$) and interleaved learning conditions ($M = 2459$, $SD = 842.23$), $p = .10$, but there was a significant difference between blocked and semi-interleaved learning ($M = 1971$, $SD = 421.52$), $p < .001$. Likewise, there was a significant difference between interleaved and semi-interleaved learning conditions in terms of reaction times to target stimuli, $p = 005$.

There was also a significant interaction between study status (old vs. new) and learning conditions (blocked, interleaved & semi-interleaved) on response times, $F(2,99) = 7.24$, $p = .01$, $MSE = 215516.22$, $\eta_p^2 = .13$,. Also, post-hoc tests using pairwise comparisons were conducted to find response time (See *Table10*) difference between old and new exemplars, using repeated measures ANOVA, with study status as the repeated factor, separately for each learning condition . In blocked condition, there was a significant difference between old exemplars ($M = 2798$, $SD = 612.98$) and new exemplars ($M = 2372$, $SD = 630.92$) of exemplars, $F(1,33) = 12.70$, $p = .001$, $\eta_p^2= .28$. In the interleaved condition, there was not a significant response time difference between old (*M* = 2499, *SD* = 874.22) and new exemplars (*M* = 2419, *SD* = 969.61), $F(1,33) = .37$, $p = .55$, $\eta_p^2 = .01$. In semi-interleaved learning condition, there was a statistically significant difference between old exemplars (*M* = 1881, *SD* = 464.31), and new exemplars (*M* = 2060, *SD* = 508.59), $F(1,33) = 4.55$, $p = .04$, $\eta_p^2 = .12$.

Table 10.Conditional response times for new and old exemplars in exemplar testing part

Reaction Time	Old Exemplar *Mean(SD)*	New Exemplar *Mean(SD)*
Blocked Condition	2798 (612.98)	2372 (630.92)
Interleaved Condition	2499 (874.22)	2419 (969.61)
Semi-interleaved Condition	1881 (464.31)	2060 (508.59)

A one-way ANOVA analysis conducted to find category inferences difference between learning conditions. Analysis revealed that there was a main effect of learning conditions $F(2,99) = 12.72$, $p < .001$. (See *Table11*). A Bonferroni post-hoc analysis was conducted. There was a significant difference between blocked (*M* =.46, *SD* = .19) and interleaved learning conditions (*M* =.34, *SD* =.20), $p = .02$, but there was no significant difference between blocked and semi-interleaved(*M* = .56, *SD* = .15) for category inferences, p = .08. On the other hand, there was a significant difference between interleaved and semi-interleaved learning conditions for category inferences, $p < .001$.

Table 11.Inference rate for actual category names

Inference rate	*Mean(SD)*
Blocked Condition	.46(.19)
Interleaved Condition	.34(.20)
Semi-interleaved Condition	.56(.15)

CHAPTER IV

COMPARISON OF EXPERIMENT 1 & EXPERIMENT 2

As a result of finding how conceptual and visual study conditions provide different learning performances, reaction times of encoding phase, exemplar testing accuracy, response times and category inferences in testing phase data of Experiment 1 and Experiment 2 were analyzed together.

4.1 Results

Encoding Phase. First of all, the mean of RTs was submitted to a mixed analysis of variance (ANOVA), with the presentation time (first vs second) as the repeated measure, learning condition (blocked, interleaved and semi-interleaved learning conditions) and experiment group (first vs second experiment) as the between-subjects variables. Mixed ANOVA results revealed that presentation time x learning condition x experiment group interaction was significant, $F(2,198) = 22.28$, $p < .01$, $MSE = 36785.59$. Means for all conditions can be found in *Figure 5*.

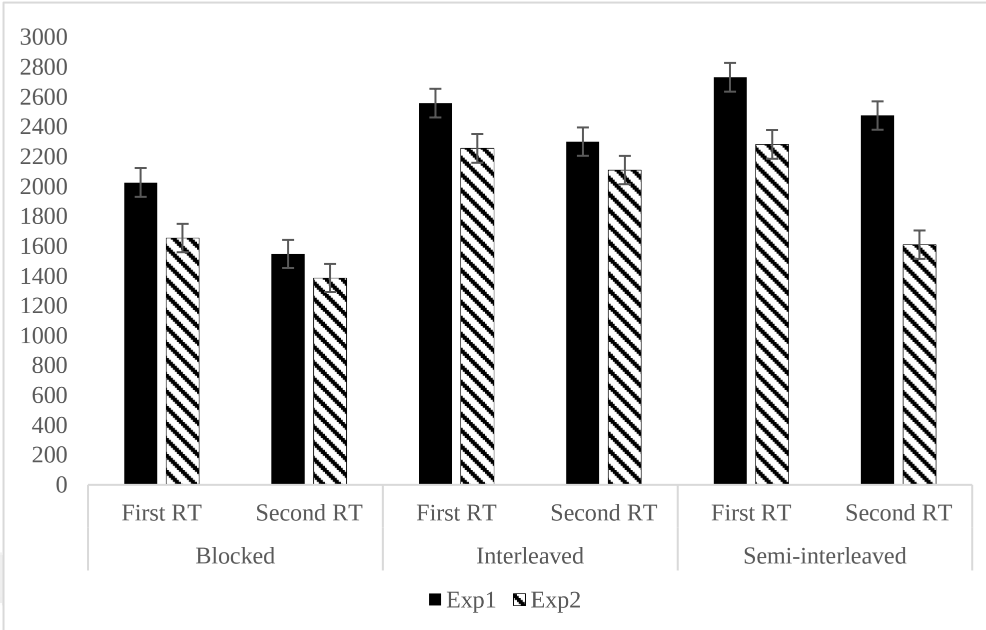

Figure 5. Reaction times of Experiment 1 and Experiment 2 for all learning conditions

Testing Phase. For exemplar testing part, testing accuracy for study status (old vs new) was submitted to a mixed analysis of variance (ANOVA) with learning condition (blocked, interleaved and semi-interleaved learning conditions) and experiment group (first vs second experiment) as the between-subjects variable. Results did not reveal a significant three-way interaction, $F(2,198) = .74$, $p = .48$, $MSE = .01$. Thus, testing accuracy study status (old vs new) was submitted to mixed ANOVA as within-subject variable with experiment group (first vs second experiment) as between-subject variable. Results revealed a significant two-way interaction, $F(1.202) = 4.37$, $p = .04$, $MSE = .01$, $\eta_p^2 = .02$, (See *Figure 6* for accuracy rates).

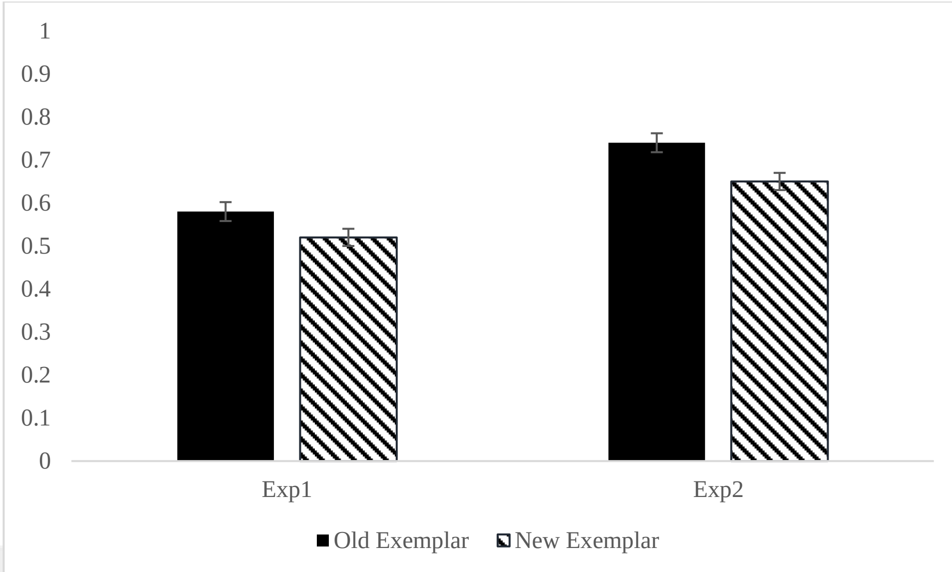

Figure 6. Testing accuracy of old and new exemplars for Experiment 1 and Experiment 2

Moreover, for unconditional and conditional response times for testing phase were analyzed seperately. For unconditional response times, there was significant three-way interaction between study status (old vs new), learning condition (blocked, interleaved and semi-interleaved) and experiment group (first experiment vs second experiment), $F(2,198) = 6.87$, $p = .001$, $MSE = 29960.20$. All response times for experiment group and learning condition can be found in *Figure 7*.

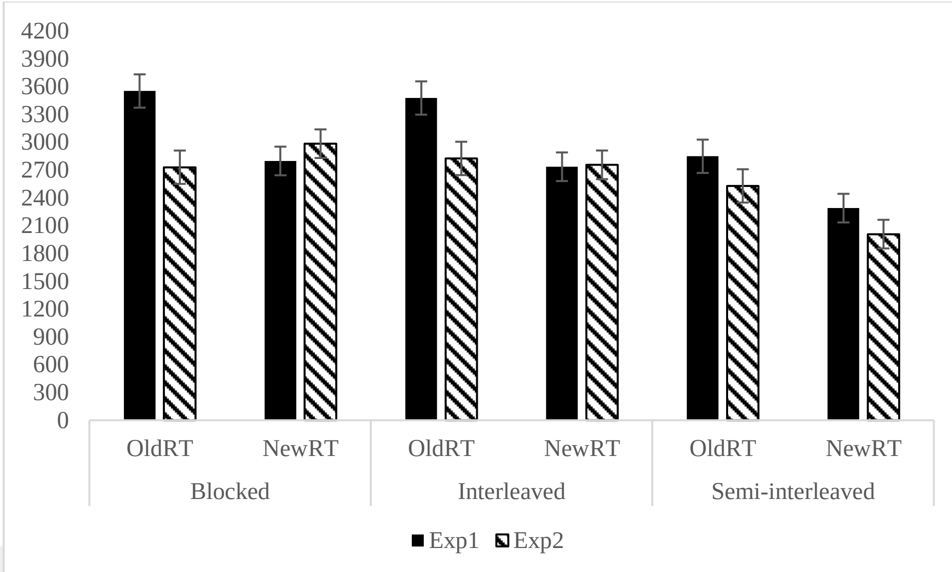

Figure 7.Unconditional response times of old and new exemplars of Experiment 1 and Experiment 2

Conditional analysis for response times revealed that there was not significant interaction between study status (old vs new), learning condition (blocked, interleaved and semi-interleaved) and experiment group (first experiment vs second experiment), $F(2,198) = .28$, $p = .75$, $MSE = 324419.54$. Therefore, all response times was submitted to mixed ANOVA as within subject variable and experiment group as between-subject variable. Results revealed significant interaction between study status and experiment group, $F(1,202) = 12.60$, $p < .001$, $MSE = 340344.30$, $\eta_p^2 = .06$ (See *Figure 8* for means).

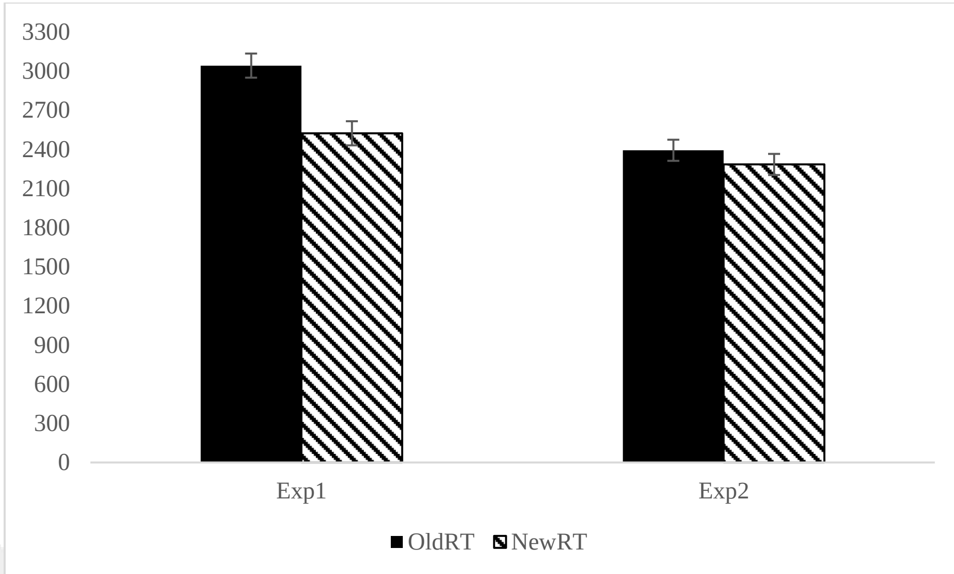

Figure 8.Conditional response times of old and new exemplars of Experiment 1 and Experiment 2

For category testing, category accuracy was submitted to univariate analysis of variance with experiment group and learning conditions as between subject variables. A univariate analysis showed that category accuracy, learning condition and experiment group interaction was not significant, $F(2,198) = .36$, $p = .70$, $MSE = .05$. Thus, a one-way ANOVA analysis was conducted finding for category accuracy difference between two experiments. Results showed that there was a significant main effect of experiment groups on category accuracy, $F(1.202) = 19.31$, $p < .001$, $\eta_p^2 = .09$ (See *Figure 9* for means).

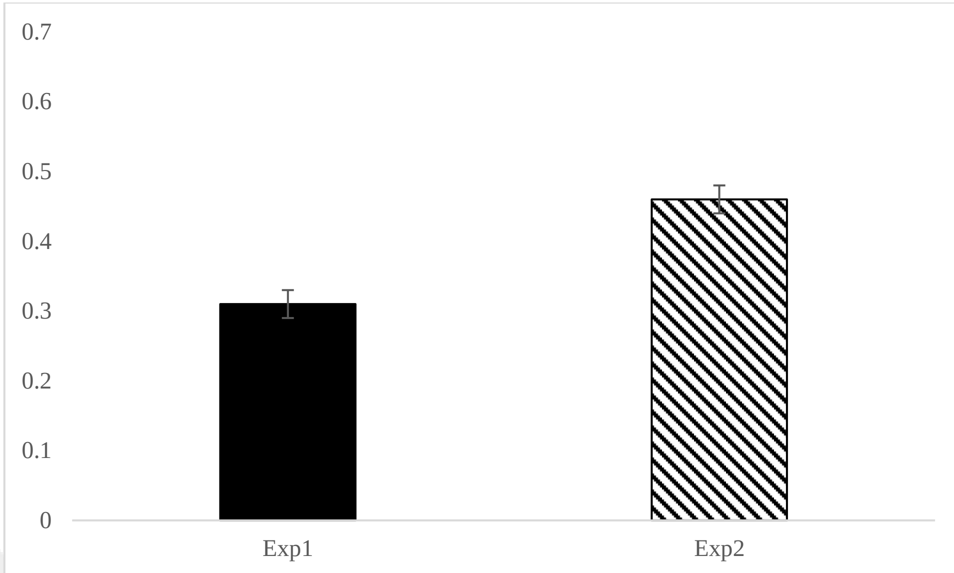

Figure 9.Category accuracy difference between experiment groups

Besides, an additional analysis was conducted to find participants' general understanding differences in terms of actual categories between first and second experiments. First, "category inference accuracy" which consisted of how many category names participants could correctly guess out of 12 categories and "exemplar testing accuracy" for each exemplar were calculated. Then, correlation between "category inference accuracy" and "exemplar testing accuracy" for each category and each participant were analyzed. Because some of the participants could not correctly guess none of actual category names during category testing part, their category inference data were "zero" for all twelve categories. Therefore, correlation for those constant data was not calculated by the statistic program and were not added to analyses. In Experiment 1, 78 sample (blocked = 27, interleaved = 22 and semi-interleaved = 29) and in Experiment 2, 97 sample (blocked = 33, interleaved = 31 and semi-interleaved = 33) could be included in the analyses. One-way ANOVA analyses revealed that there were not significant effect of learning conditions on correlations both for Experiment 1, $F(2,75) = 2.33$, $p = .11$, and Experiment 2, $F(2,94) = 2.71$, $p = .07$. In addition, the correlation values were submitted to a univariate analysis of variance as dependent variable with learning conditions and experiment groups (first vs second experiment) as between-subject variables. Results could not show significant interaction, $F(2, 169) = 2.71$, $p = .07$, $MSE = .08$, $\eta_p^2 = .03$.

CHAPTER V

GENERAL DISCUSSION

The effects of blocked and interleaved learning on category learning have been a long-debated issue. Some studies find evidence for superiority of interleaved learning over blocked learning strategy (Birnbaum et al., 2012; Kang & Pashler, 2012; Kornell & Bjork, 2008), whereas other studies support that blocked learning at encoding leads to higher category induction over interleaved learning (Carpenter & Mueller, 2013; Sorensen & Woltz, 2016). The first aim of the present study was to test effective learning strategies through using three different learning designs at study: blocked, interleaved and semi-interleaved. Participants were assigned to one of three conditions at learning phase: blocked, interleaved, and semi-interleaved learning. In blocked learning, participants studied four exemplars of the same category within the same trial, and categories were blocked across trials. In interleaved learning, participants studied one exemplar from four different categories within the same trial, and categories were interleaved across trials. In semi-interleaved learning, participants studied four exemplars of the same category within the trial, but categories were interleaved across trials. At test, participants were presented with old and novel exemplars of the categories they studied and asked to identify the categories. Experiment 1 used verbal materials and Experiment 2 used pictorial materials. Results of Experiment 1 & 2 revealed that during studying categories, participants got faster from first presentation to second presentation in all learning strategies. It means that in all learning conditions, participants became familiar with exemplars. In addition, learning conditions also influenced reaction times. Blocked learning strategy produced the fastest reaction time in encoding phase, whereas participants in semi-interleaved learning strategy were the slowest.

Although blocked learning condition revealed the fastest reaction time, being quicker did not produce higher performance than other learning conditions. In exemplar testing, semi-interleaved condition showed better performance than blocked and interleaved learning conditions. In addition, blocked and interleaved showed the same amount of category learning. Likewise, in old and new exemplar testing, interleaved and semi-interleaved learning conditions provided better old exemplar learning than new exemplars in both experiments. On the other hand, old and new exemplars provided same results in blocked learning in Experiment 1, but old exemplars were remembered better than new exemplars in blocked learning condition in Experiment 2.

In both experiments, semi-interleaved learning condition produced the highest implicit learning than all other conditions, which confirmed our hypothesis. In addition, results of semi-interleaved learning condition were also consistent with both *massing-aids-induction-hypothesis* and *discriminative-contrast-hypothesis*. Because in semi-interleaved learning condition, all exemplars from one category were presented in one trial, participants could easily abstract the features of a given category, as the *massing-aids-hypothesis* argues. In semi-interleaved condition, different category for each trial was randomly presented to participants. Thus, this method provided easier discrimination between different categories, as the *discrimination-contrast-hypothesis* contends.

Results of Experiment 1 and 2 both showed that blocked and interleaved learning strategies produced similar levels of learning in exemplar testing part, but the difference between blocked and interleaved learning almost became undistinguishable in Experiment 2. Although it did not reach significance, blocked learning condition produced higher exemplar memory performance than interleaved learning condition. Therefore, it showed that blocked learning showed more improvement than interleaved learning strategy from Experiment 1 to Experiment 2. This can be explained by the procedure of the experiment. Because participants could see all exemplars of one category all at once with pictures in blocked condition of Experiment 2, it might have become much easier to perceive the shared characteristics within the same category, specifically when it was a visual characteristic. For example, in “something that is yellow” category participants saw four pictures of objects that have yellow color; therefore, the similarity might have been perceived more readily than through the verbal label presented in Experiment 1.

Results of the interleaved learning condition did not produce superior category learning than blocked learning condition in Experiment 1 & 2. Results did not support *discriminative-contrast-hypothesis*. In addition, results of blocked condition in both experiments also did not produce better memory performance in exemplar testing than interleaved learning. Those results also did not support *massing-aids-induction-hypothesis*.

As one of the previous studies that supported superiority of interleaving, Kang and Pashler (2012) found that “simultaneous different learning condition” produced best inductive learning than blocked and temporally spaced learning conditions. “Simultaneous different” learning had same procedure with interleaved learning condition of current research, but findings of Kang & Pashler and current research were inconsistent. Blocked and interleaved learning conditions also had inconsistent results with previous research that found superiority for blocked learning strategy (Carpenter & Mueller, 2013; Sorensen & Woltz, 2016) and for interleaved learning strategies (Kang & Pashler, 2012). Those inconsistent results may be related to desirable difficulty. Desirable difficulty is a term refers to making to-be-learned task more challenging. Bjork (1994) argued that making to-be-learned materials desirably difficult enhances memory performance. For instance, Winstanley, Bjork, & Bjork, 1996 conducted a study to show effects of desirable difficult. They presented word pairs to the participants. Pairs were presented either in with missing letters such as "cold-h__" or in intact condition such as "cold-hot". They found that pairs that presented with missing letters were remembered better than just read pairs. In missing paired condition, participants had to think about the possible word that was suitable for the pair which was more effortful than just reading the paired words. Thinking about the words and filling the blanks provided better encoding and retrieval of stimuli (Bjork, 1994). Therefore, it produced better memory performance which was a good example of desirable difficulty. Bjork (1994) also mentioned that not every difficulty is desirable. When people do not have enough experience and prior knowledge for responding to stimuli, desirable difficulty becomes "undesirable difficulty".

Because all exemplars of the same category were presented consecutively in blocked learning, participants were asked to press the same key for eight times. Therefore, blocked learning procedure does not require much attention. In addition, in blocked learning strategy, participants knew what they would study next. Accordingly, they may not pay much attention to upcoming stimuli. Blocked learning condition can be under the level of desirable difficulty.

In contrast, in interleaved learning, because exemplars from different categories are intervening, it may prove to be a more challenging procedure than the desirable difficulty, because participants may need to pay more attention beyond the limits of their working memory span. Therefore, the difficulty may not be desirable when so many category-item associations need to be learned within a single trial. Difficulty level could be below the desirable level for blocked learning and above the desirable level for interleaved learning which prevented successful encoding of stimuli. The semi-interleaved learning strategy proved to be the desirable challenge that leads to increased category learning over all other learning strategies.

Another factor that affects results of interleaved learning condition can be, participants' potential prior associations for exemplars from different categories. If participants have prior associations for exemplars, it is harder to construct a categorization process with determined exemplars of one category. For example, when participants saw "wardrobe" and "chair" from "something that made of wood" category and "spoon" and "teapot" from "something that made of wood" category within same trials, they could form a new category such as "household items". As an anecdotal evidence "household items" category also was formed by some of the participants in interleaved learning condition. This situation can be problematic only in interleaved learning condition because participants saw exemplars from different categories together. Participants from interleaved learning had difficulties to come up with general understanding for categories more than other conditions. Underwood (1952) also argued that understanding relationship between different stimuli requires adjacent perception of those stimuli. Since participants assigned in blocked and semi-interleaved learning strategy have never seen different exemplars from different categories in one trial, they could benefit from adjacent perception of related stimuli.

On the other hand, in interleaved learning condition it can lead to confusion in terms of discrimination of different category exemplars.

Sana et al. (2017) found that participants' working memory capacity matters in category learning performances. Interestingly, they found that working memory capacity predicted people's learning performance in blocked learning condition more than interleaved learning condition. They found that participants who had higher working memory capacity had ability to make easier classification than participants who had lower working memory capacity. It can be because of that participants in blocked learning condition had to keep all exemplars of categories to be able to compare other categories whereas, in interleaved learning strategy, participants did not have to keep other exemplars, because exemplars were presented in mixed order. Participants in blocked learning who had lower working memory capacity may have been affected from low category understanding disproportionately.

Another factor that mitigates the effects of blocked and interleaved learning conditions on category induction may be task characteristics. Previous research that supported superiority of interleaved learning condition used different task to measure category learning than current research. Current research used categories that required rule induction. It means that participants had to come up with a description that can easily verbalize for each category like "something that made of steel". On the other hand, generally previous research used information-integration task method for categories such as paintings of different artists. Therefore, it was hard to come up with a concrete category understanding of artists' painting styles. This may explain why interleaved learning provided support for implicit understanding in previous research, because they could compare different exemplars and it helped to understand which characteristics differentiate one category form another. On the other hand, in the current study understanding shared commonalities within category exemplars can be more important than what an exemplar has a characteristic different from another category exemplars. The current study required participants to understand category rules and remember non-word category names with correct rules. When participants understand a category rule, it immediately differentiates the category from other categories. Understanding shared characteristics of the category may help category learning for current research. Therefore, interleaved learning could not provide superior category learning performance.

The last analysis from Experiment 2 showed that learning conditions did not have an effect on category inferences which tested participants' explicit understanding of the twelve categories. Thus, even though semi-interleaved learning condition had superiority in exemplar testing, it did not provide superiority for explicit knowledge of categories. Those results showed that while working with words, there was no difference between different category learning strategies in terms of explicit learning. On the other hand, in Experiment 2, performances in category inferences changed according to learning strategies. Semi-interleaved produced the best explicit learning performance. Interestingly, blocked learning provided better category inferences than interleaved learning. Therefore, it can be concluded that Experiment 2, which consisted of pictorial materials rather than verbal materials produced more category learning than Experiment 1. Even though there was not significant difference in exemplar testing between blocked and interleaved learning conditions, blocked learning provided better category learning than interleaved in Experiment 2. Therefore, benefit of semi-interleaved learning strategy may come from blocking part of procedure.

The second aim of the current study was to compare encoding of verbal and pictorial materials. Semi-interleaved learning strategy showed the most improvement for reaction times for both Experiment 1 and Experiment 2 from the repeated presentation of exemplars. Thus, participants from semi-interleaved condition became faster in second presentation than blocked and interleaved conditions. In addition, semi-interleaved learning strategy produced the highest learning performance in both experiments. Since semi-interleaved learning condition produced higher encoding than other strategies, it produced more successful category learning.

Testing exemplar accuracy analyses showed that both in old and new exemplars, learning performance raised in all learning conditions in Experiment 2 compared to Experiment 1. Superiority of Experiment 2 can be related to better encoding of visual materials. Shepard (1967) found that pictures are more memorable than words in general, which is also known as *picture superiority effect*. Clark & Paivio (1991) argued that pictures can be encoded both visually and verbally, but words can be encoded only verbally. Therefore, pictures provided better memory performance than words. When categories were presented as pictures, participants' memory performance was significantly higher than categories were presented as words. On the other hand, although Experiment 2 provided higher memory performance than Experiment 1, studying with verbal or pictorial materials could not moderate effect of learning conditions. Those results revealed that organization of study materials have importance on category learning.

Unlike exemplar testing result, nature of study materials moderated the effect of learning conditions on explicit category learning. For category accuracy, Experiment 2 showed more explicit understanding of categories than Experiment 1. In addition, pictures can lead to easier abstractions about commonalities of exemplars, especially in categories that have perceptual commonalities like “something that is white”, “something that is yellow”, “something that has triangular shape” and etc. In contrast to those findings, correlation between exemplar testing and category accuracy analysis showed that there was no difference between Experiment 1 and Experiment 2. Correlation analysis was formed by conducting correlation between participants’ exemplar testing performance (how many exemplars they could correctly remember out of eight exemplars during testing phase) and category inference accuracy (how many correct category names participants know out of twelve categories). Those results mean that although studying with pictures provided more explicit memory performance than words, understanding concepts did not necessarily require to remember each exemplar of the category. Participants who successfully understand what a category represents, showed better performance in explicit learning task even though they did not remember what exemplars of the category were.

Current research has also some drawbacks as well as its theoretical implications. Differently from previous research, current research used rule-based categories. During encoding phase participants were asked to find common points of a category and in category inference test, they also were expected to write actual category names based on their guesses. Participants had never been presented with the correct category names, and the whole experiment was based on their inferences. On the other hand, previous research used information integrative categories, which participants were presented with correct category name and exemplar pairs. In current research, participant could not show their category understanding during category inference test because they did not have any cues about the categories. This problem could be ameliorated by a matching task which participants were presented with non-word category names and actual category names. They were expected to match two categories which can show participants' category understanding better.

Overall, with results of Experiment 1 and 2, it can be said that although effects of nature of study materials (verbal vs. pictorial) can be observed in some levels of categorization performance, organization of study materials still have huge effect on learning performance. It is clear with results of semi-interleaved learning strategy that includes both blocking and interleaving learning strategies help category learning, but it is still unclear that how much blocking and interleaving contribute to category learning. Further studies should investigate level of blocking or interleaving during category learning rather than pure blocking and interleaving which is not effective at all. For further studies, level of blocking and interleaving can be manipulated within one trial. In one trial, exemplars can vary from blocked learning to interleaved learning. It refers that in one trial, participants can see four exemplars from one category, in second trial, three exemplars from same category and one exemplar from different category. In third trial, they see two exemplars from two different categories, and it can vary until a trial interleaved learning condition of current research. This procedure can reveal that effects different level of blocking and interleaving on category learning.

REFERENCES

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition*, *41*, 392-402.

Bjork, R. A. (1994). Memory and meta-memory considerations in the training of human beings. In J. Metcalfe, & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185–205). Cambridge, MA: MIT Press.

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*, 671-682.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*, 481-495.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, *22*, 281-288.

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*, 354–380.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3, 149-210.

Cohen, M. S., Yan, V. X., Halamish, V., & Bjork, R. A. (2013). Do students think that difficult or valuable materials should be restudied sooner rather than later?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1682-1696.

de Winstanley, P. A., Bjork, E. L., & Bjork, R. A. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, 4, 31–48.

Foss, D. J. (1968). Learning and discovery in the acquisition of structured material: Effects of number of items and their sequence. *Journal of Experimental Psychology*, 77, 341-344.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement?. *Psychonomic Bulletin & Review*, *19*, 126-134.

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97-103.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"?. *Psychological science, 19*, 585-592.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of Experimental Psychology*, 51, 239– 243.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764-766.

Mitchell, C., Nash, S., & Hall, G. (2008). The intermixed-blocked effect in human perceptual learning is not the consequence of trial spacing. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34*, 237–242.

Olina, Z., Reiser, R., Huang, X., Lim, J., & Park, S. (2006). Problem format and presentation sequence: Effects on learning and mental effort among US high school students. *Applied Cognitive Psychology*, *20*, 299-309.

Pan, S. C., Tajran, J., Lovelett, J., Osuna, J., & Rickard, T. C. (2019). Does interleaved practice enhance foreign language learning? The effects of training schedule on Spanish verb conjugation skills. *Journal of Educational Psychology*.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review*, *24*, 355-367.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498.

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review*, *21*, 1323-1330.

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology*, *109*, 84-98.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156-163.

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory & Cognition*, *44*, 1000-1013.

Taylor, K., & Rohrer, D. (2010). The effect of interleaving practice. *Applied Cognitive Psychology*, *24,* 837–848.

Underwood, B. J. (1952). An orientation for research on thinking. *Psychological Review*, *59*, 209–220.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16-27.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215-221.

APPENDICES

APPENDIX A. CATEGORIES AND EXEMPLARS

1) Tahtadan Yapılan Şeyler

1. Kibrit
2. Kalem
3. Merdane
4. Sandalye
5. Bank
6. Gardrop
7. Kitaplık
8. Kayık

2) Çelikten Yapılan Eşyalar

1. Çatal
2. Kaşık
3. Termos
4. Çaydanlık
5. Bisiklet
6. Kılıç
7. Tren Rayı
8. El arabası

3) Sarı olan şeyler

1. Muz
2. Güneş
3. Civciv
4. Limonata
5. Limon
6. Taksi
7. Patates
8. Mısır

4) Beyaz olan Şeyler

1. Bulut
2. Süt
3. Diş
4. Gelinlik
5. Kuğu
6. Kutup Ayısı
7. Yazı Tahtası
8. Pamuk

5) Topsuz Sporlar

1. Yüzme
2. Koşu
3. Yoga
4. Kayak
5. Jimnastik
6. Boks
7. At Binme
8. Karate

6) Top ile Oynanan Sporlar

1. Basketbol
2. Voleybol

7) Yuvarlak olan Şeyler

1. Top
2. Yüzük
3. CD
4. Düğme
5. Para
6. Donut
7. Gezegen
8. Portakal

8) Üçgen olan Şeyler

1. Askılık
2. Külah
3. Yelken
4. Pizza Dilimi
5. Trafik konisi
6. Çam ağacı

3. Futbol
4. Hentbol
5. Badminton

7. Havuç
8. Mısır Piramitleri

9) Gürültü Çıkararak Çalışan Eşyalar

1. Radyo
2. Matkap
3. Elektrik Süpürgesi
4. Saç kurutma makinesi
5. Aspiratör
6. Mikser
7. Megafon
8. Klarnet

10) Sessiz Çalışan Eşyalar

6. Amerikan Futbolu
7. Tenis
8. Golf

1. Saat
2. Lamba
3. Ütü
4. Fotoğraf makinesi
5. Saç Düzleştiricisi
6. Şarj Aleti
7. Tost Makinesi
8. Uydu Alıcısı

11) Kokusu Baskın Olan Şeyler

1. Balık
2. Sucuk
3. Gül
4. Tarçın
5. Parfüm
6. Kahve

7. Sigara
8. Çöp

12) Kokusu Baskın Olmayan Şeyler

1. Su
2. Kabak
3. Kum
4. Avokado
5. Yaprak
6. Cam
7. Kaktüs
8. Taş

APPENDIX A1. ACTUAL NAMES OF NON-WORD CATEGORY NAMES

1. VOMİDOR: Something that is white (cloud, whiteboard, wedding dress, swan, milk, teeth, cotton, polar bear)
2. NERUSA: Something that is yellow (banana, sun, lemonade, lemon, taxi, chick, potato, corn)
3. ZİLAKO: Something that is made of wood (lucifer match, pencil, rolling pin, chair, bank, wardrobe, bookcase, boat)
4. CAKİDE: Something that is made of steel (bicycle, fork, spoon, thermos, teapot, sword, railroad, hand cart)
5. SEDOKA: Devices that make noise (radio, drill, vacuum cleaner, hair dryer, cooker hood, mixer, clarinet, hailer)
6. FESİYA: Devices that work quietly (clock, iron, phone charger, lamb, camera, hair straightener, satellite receiver)
7. HİSAM: Something that is odorless (water, rock, avocado, zucchini, leaf, sand, cactus, glass)
8. LURTES: Something that has a strong odor (rose, pepperoni, cinnamon, fish, garbage, perfume, coffee, cigarette)
9. ERBEZ: Sports which are played without aball (swimming, running, yoga, boks, horsemanship, karate, boxing, gymnastic)
10. YEFARO: Sports which are played with a ball (football, volleyball, handball, badminton, tennis, golf, American football, basketball)
11. URTESA: Something that has a round shape (ball, button, donut, orange, planet, ring, CD, coin)

12. PELATO: Something that has a triangular shape (carrot, pizza slice, road cone, pyramid, pine-tree, coat hanger, cone, sail)