

FATİH TAYYİP YAVUZ

THE GENERATION EFFECT ON MEMORY AND METAMEMORY

Bilkent University 2024

INVESTIGATION OF THE GENERATION EFFECT
ON MEMORY AND METAMEMORY THROUGH
SEMANTIC AND PERCEPTUAL CUES

A Master's Thesis

by
FATİH TAYYİP YAVUZ

Department of
Psychology
İhsan Doğramacı Bilkent University
Ankara
April 2024

INVESTIGATION OF THE GENERATION EFFECT ON MEMORY AND
METAMEMORY THROUGH SEMANTIC AND PERCEPTUAL CUES

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

FATİH TAYYİP YAVUZ

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN PSYCHOLOGY

THE DEPARTMENT OF
PSYCHOLOGY
İHSAN DOĞRAMACI BILKENT UNIVERSITY
ANKARA

April 2024

INVESTIGATION OF THE GENERATION EFFECT ON
MEMORY AND METAMEMORY THROUGH
SEMANTIC AND PERCEPTUAL CUES

BY FATİH TAYYİP YAVUZ

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 Arts in Psychology

Miri Besken
Advisor

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 Arts in Psychology

Aslı Kılıç Özhan
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 Arts in Psychology

Jedediah Wilfred Papas Allen
Examining Committee Member

Approval of the Graduate School of Economics and Social Sciences

Refet S. Gürkaynak
Director

ABSTRACT

INVESTIGATION OF THE GENERATION EFFECT ON MEMORY AND METAMEMORY THROUGH SEMANTIC AND PERCEPTUAL CUES

Yavuz, Fatih Tayyip

M.A., Department of Psychology

Supervisor: Asst. Prof. Dr. Miri Besken

April 2024

The generation effect is an experimental finding that self-generated information produces higher memory performance than reading. The effect can be obtained through various semantic and perceptual generation manipulations. Despite numerous studies investigating semantic and perceptual generation tasks separately, studies have not compared the effectiveness of the two formats directly. For Experiment 1, participants read or generated words from rhyming words or highly associated words, comprising a 2 (level of processing: perceptual, semantic) x 2 (generation status: generation, read) within-subjects design. The results showed that participants had higher memory performance for generated words than for words that were read. Moreover, an interaction revealed that memory performance was higher for the semantic generation task than for the perceptual generation task. Experiment 2 aimed

to investigate participants' memory predictions for perceptual and semantic generation tasks. Moreover, we investigated whether making predictions modifies memory performance. Experiment 2 incorporated judgments-of-learning (JOL) and no-JOL groups, yielding that participants accurately predicted and performed better on memory tasks involving generation and semantic manipulations. Additionally, the cued-recall retrieval phase produced higher memory performance for the JOL-group than the no-JOL group, suggesting that predicting one's memory performance enhances actual memory performance. Experiment 3 aimed to see the importance of the test type for the metamemory reactivity. Like Experiment 2, the JOL group still had a higher memory performance than the no-JOL group. The group difference was only observable during a semantic cued-recall test, implying the test type's importance.

Keywords: Generation Effect, Metamemory, Metamemory Reactivity, Levels of Processing, Transfer Appropriate Processing

ÖZET

BELLEK VE ÜSTBELLEK ÜZERİNDEKİ ÜRETİM ETKİSİNİN ANLAMSAL VE ALGISAL İPUÇLARIYLA İNCELENMESİ

Yavuz, Fatih Tayyip

Yüksek Lisans, Psikoloji

Tez Danışmanı: Dr. Öğr. Üyesi Miri Besken

Nisan 2024

Üretim etkisi, kendi kendine üretilen bilginin pasif olarak okunmasından daha yüksek bellek performansı ürettiğine dair kanıt sağlayan deneysel bir bulgudur. Bu etki, çeşitli anlamsal ve algısal üretim manipülasyonları yoluyla elde edilebilir. Anlamsal ve algısal oluşturma görevlerini araştıran çok sayıda çalışmaya rağmen, çalışmalar genellikle algısal ve anlamsal üretmenin etkinliğini doğrudan karşılaştırmamaktadır. Deney 1 için katılımcılar, 2 (işleme düzeyi: algısal, anlamsal) x 2 (üretim: üretim, okuma) denek içi tasarımı içeren, kafiyeli kelimelerden veya yüksek oranda ilişkili kelimelerden kelimeler okudu veya oluşturdu. Tanıma testi sonuçları, katılımcıların oluşturulan kelimeler için okudukları kelimelere göre daha yüksek bellek performansına sahip olduklarını gösterdi. Üstelik bir etkileşim, anlamsal oluşturma görevi için bellek performansının algısal oluşturma görevinden daha yüksek olduğunu

ortaya çıkardı. Deney 2, katılımcıların hafıza tahminlerinin, algısal ve anlamsal üretim görevleri altındaki sonraki hafıza performanslarına göre nasıl değiştiğini görmeyi amaçladı. Ayrıca değerlendirme yapmanın hafıza performansını değiştirip değiştirmediğini de araştırdık. Deney 2, öğrenme yargılarını (ÖY) içeren ve ÖY içermeyen gruplardan oluşturuldu; katılımcıların, üretim ve anlamsal manipülasyonları içeren hafıza görevlerini doğru bir şekilde tahmin ettiklerini ve bu görevler üzerinde daha iyi performans sergilediklerini gösterdi. Ek olarak, ipucuyla geri çağırma aşaması, ÖY grubu katılımcılarının ÖY olmayan gruptan daha iyi performans göstermesiyle sonuçlandı; bu da kişinin bellek performansını tahmin etmenin gerçek bellek performansını artırdığını öne sürdü. Deney 3, üstbellek reaktivitesi için geri getirme aşamasının önemini görmeyi amaçladı. Deney 2'de olduğu gibi, ÖY grubu ÖY olmayan gruptan daha yüksek bir bellek performansına sahipti. Grup farkı yalnızca anlamsal ipucuyla-geri çağırma testi sırasında gözlemlendi, bu da test türünün önemini gösterdi

Anahtar kelimeler: Üretim Etkisi, Üstbellek, Üstbellek Reaktivitesi, İşleme Düzeyi, Aktarıma Uygun İşleme

ACKNOWLEDGMENTS

Once, a professor said the following: we change as we age. That is a decent saying, and it is not just about rhyming. I have spent my days, weeks, and months in the text you are about to read; this is simply my three years. I have changed with each challenge; each struggle requires a new solution, and I may not know the answers then. Fortunately, I was not alone in the strange waters. Throughout this passage, I will express my gratitude for each unique contribution that helped finish my story.

First of all, I want to thank my supervisor Miri Besken. It all started on a random Sunday when she emailed me to conduct an interview. It was not the actual interview for admission; she wanted to know me. It was a great meeting, and then I found myself sitting with her as a Bilkent student. For over three months, we covered a variety of subjects and finally decided on the current thesis. It was an excellent opportunity to learn about the literature and make a decision on your own dissertation. I remember that one time when I collected enough data for Experiment 1, she cheerfully yelled my name from the upper floors with genuine excitement to ask for the results. After Experiment 2, we had unexpected results, and we were sitting on the couches of the entrance and trying to come up with a meaning (spoilers: we designed the appropriate methodology for Experiment 3). My friends from other graduate schools usually told me about the hardness of graduate studies. It was not the case for me. Sure, it was hard sometimes, but I was cheerful, and there is a relationship

between my cheerfulness and Miri Besken. I am just grateful; thanks for the experience.

Secondly, I need to talk about my family. I want to thank my mother, brother, father, and sister. Mainly, I will mention my mother and brother. They were always in contact with me during each challenge, and whenever I needed mental relief, they helped me sustain my education.

I must say a big thank you to Özge Konyalı, who was always with me during my last two years of education. She was always supportive, finding solutions for my problems and helping me with the thesis, even though it was not something she usually enjoyed. As I said to her once, I am very lucky to have her in my life.

Also, I was never just a graduate student but an ultimate frisbee player.

Approximately every month, we were a part of a national tournament. All the training, practices, and competitions were a relaxing escape experience, and graduate students should be encouraged to be socially active. Life is not just lab.

On the other hand, my lab partners helped me immensely. I want to thank Rengin Dalkıran, my first friend at Bilkent University, for being able to laugh at my jokes, still. Although she just came the previous semester, Öykü Aydın is one of the reasons I still believe in making friends after 25. Additionally, I want to thank Nilsu Sağlam for being the first undergraduate student to collect the data for my thesis, followed by Duygu Savaş, İpek Deniz Genç, Zeynep Acar, and so on.

I am grateful to each participant who spent about 45 minutes on my thesis.

Lastly, I would like to thank the Scientific and Technological Research Council of Turkey (TUBITAK) for providing me with financial support under the BIDEB 2210-A program.



TABLE OF CONTENTS

ABSTRACT.....	i
ÖZET	iii
ACKNOWLEDGMENTS	v
LIST OF TABLES.....	xi
LIST OF FIGURES	xii
CHAPTER 1	1
INTRODUCTION	1
1.1 Generation Effect.....	4
1.1.1 Underlying Mechanisms	5
1.1.2. Generation Tasks.....	7
1.2 Metamemory	9
1.2.1 Judgment of Learning.....	10
1.2.2. Fluency Effect.....	11
1.3 Metamemory Reactivity	14
1.3.1 Cue-Strengthening Hypothesis	17
1.4 Aims of the Current Project	22
CHAPTER 2	24
EXPERIMENT 1	24
2.1 Method.....	25
2.1.1 Participants.....	25
2.1.2 Design.....	26
2.1.3 Materials	27

2.1.4 Procedure.....	28
2.2 Results.....	31
2.3 Discussion.....	33
CHAPTER 3	36
EXPERIMENT 2	36
3.1 Method.....	38
3.1.1 Participants.....	38
3.1.2 Materials, Design, and Procedure	39
3.2 Results.....	41
3.2.1 Response Latency	41
3.2.2 Metamemory Performance.....	42
3.2.3 Memory Performance	44
3.3 Discussion.....	45
CHAPTER 4	50
EXPERIMENT 3	50
4.1 Method.....	52
4.1.1 Participants.....	52
4.1.2 Materials, Design, and Procedure	53
4.2 Results.....	55
4.2.1 Response Latency	55
4.2.2 Metamemory Performance.....	56
4.2.3 Memory Performance	58
4.3 Discussion.....	59
CHAPTER 5	62
GENERAL DISCUSSION	62
5.1 Generation Effect.....	62
5.2 Metamemory	66
5.3 Metamemory Reactivity	70
5.4 Limitations and Further Studies.....	76

5.5 Conclusion	77
REFERENCES	79
APPENDICES	90
WORD LIST	90



LIST OF TABLES

Table 1. Means of Response Latency and Memory Performance Results	31
Table 2. Means of Metamemory and Memory Performance Results	42
Table 3. Means of Metamemory Performance Results.....	55
Table 4. Means of Memory Performance Results	55

LIST OF FIGURES

Figure 1. Experiment 1 Schematic Description. Read conditions included both semantic and perceptual associations (e.g. hot-cold, bold-cold). There were mathematics questions between phases.	29
Figure 2. Experiment 1 Response Latency	32
Figure 3. Experiment 1 Memory Performance	32
Figure 4. Experiment 2 Schematic Example. Participants were either in the JOL or no-JOL group and saw all word groups. There were mathematics questions between phases.....	40
Figure 5. Experiment 2 Metamemory Performance	43
Figure 6. Experiment 2 Memory Performance	43
Figure 7. Experiment 3 Cue-Matching Manipulation Schematic Example.....	54
Figure 8. Experiment 3 Metamemory Performance	57
Figure 9. Experiment 3 Memory Performance	57

CHAPTER 1

INTRODUCTION

The generation effect is an experimental finding stating that participants who self-generate information produce higher memory performance than the participants who read information passively. The generation effect has been extensively documented in the literature over the last 40 years (Jacoby, 1978; Slamecka & Graf, 1978). In a typical generation effect paradigm, participants are presented with single words (Glisky & Rabinowitz, 1985) or word pairs (Rabinowitz & Craik, 1986), from which they either have to generate the information or process passively. There are various versions of it. For example, in a semantic generation task, participants may be given a word pair (e.g., hot-c _ _ _) for which they have to generate the antonym target word from the cue word (Fiedler et al., 1992). They passively read the word pair (e.g., hot-cold) for the read condition. There are also perceptual generation tasks (McFarland et al., 1980). For these, participants may be asked to generate words from similar looking cue-words (e.g., bold- c _ _ _) rather than semantic-derived cognitive

processes. Usually, the investigation of semantic and perceptual generation tasks is done separately in the literature. Even though there are meta-analytic studies (Bertsch et al., 2007) that show that perceptual generation tasks have a higher effect size in boosting memory performance, there is no direct comparison of these processes with the same parameters, so it is difficult to conclude whether perceptual or semantic generation tasks produce superior memory. One aim of the current set of studies was to compare these different tasks directly within the same experimental paradigm.

Another aim of the current studies was to investigate the metamemory processes associated with perceptual and semantic generation tasks. Metamemory refers to all the principles that guide our learning and memory processes (Nelson, 1990). In other words, it is related to monitoring, regulating, and controlling our memory performance. It is essential to indicate that we cannot always accurately assess our memory performance, leading to retrieval failures. Various cues affect our metamemory systematically in the same way. One specific cue that systematically affects our metamemory is the cue of fluency, the ease with which we can complete mental operations (Bjork et al., 2013). Literature routinely shows that participants tend to believe that they will remember easy-to-process, fluent stimuli more than difficult-to-process, disfluent stimuli, even though the ease of learning may not always affect actual memory performance in the same manner. From the generation effect point of view, it is more difficult to generate information than to passively read it, as determined by accuracy and reaction time. Furthermore, semantic generation may take longer than perceptual generation because generating a word from a

semantic cue is more difficult than a perceptual cue. Thus, we would like to investigate whether the difficulty level in generating information through semantic and perceptual generation tasks affects metamemory processes.

Lastly, investigating metamemory processes to understand memory mechanisms better leads to a new subject: metamemory reactivity. The mere effect of metacognitive judgments changes memory performance (Double et al., 2018). Since participants can judge their learning process during learning, they may modify their behavior. Although reactivity could be found in early psychology studies under different concepts, such as demand characteristics, Soderstrom et al. (2015) directly investigated the metamemory reactivity for the first time. The study revealed that participants produce higher memory performance when they have to make metacognitive predictions about their subsequent memory performance. In a similar line of research, Mitchum and his colleagues (2016) also found that participants only produce higher memory performance for related materials but not for unrelated materials. The last few years of research reveal that making metacognitive judgments can sometimes boost (Soderstrom et al., 2015; Yan et al., 2015; Myers et al., 2020; Tekin & Roediger, 2020) and sometimes harm memory (Mitchum et al., 2016; Janes et al., 2018) in certain cases. In other cases, making metacognitive judgments produces null results (Tauber & Rhodes, 2012; Tauber et al., 2015; Dougherty et al., 2018). The third and final aim of the current project was to understand whether there was a metamemory reactivity for the generation effect and, if so, how semantic and perceptual cues alter the memory performance.

In order to give a wide range of background information for each subject, the following sections will focus on the generation effect, metamemory, and metamemory reactivity works of literature.

1.1 Generation Effect

The generation effect is a memory manipulation in which participants produce higher memory performance for self-generated stimuli than passively read or experimenter-directed stimuli (Björk et al., 2013). The generation effect has been studied for over 40 years, starting with Slamecka and Graff (1978).

In a typical generation effect paradigm, participants are presented with items they have to apply a cognitive operation to produce or are given the item to view passively. The materials can change anywhere from word lists (Nairne & Widner, 1987; Johnson et al., 1989; Marsh et al., 2001; Weber et al., 2012; Kornell et al., 2009) to non-words, from pictures (Kinjo & Snodgrass, 2000; Leopold & Leutner, 2012; Besken, 2016) to advertisements (Sengupta & Gorn, 2002; Lee & Labro, 2004; Brennan, 2008). The manipulation can be done visually (Soderstrom et al., 2015; Nieznanski, 2011) or auditorily (Broek et al., 2001; Besken & Mulligan, 2014). Across modalities and different item types, studies typically reveal a memory advantage for generating rather than passively viewing items.

Even though the generation effect can be obtained with various different types of materials and modalities, the results also show that certain encoding and retrieval characteristics can enhance or reduce the generation effect. Several studies revealed that there was a more significant memory benefit for several experiment designs, such

as recognition over free-recall (Begg & Snider, 1987), mixed over the blocked list (Begg et al., 1991), or within-subject over between subjects (Hirshman & Bjork, 1988).

1.1.1 Underlying Mechanisms

There have been various attempts to explain the generation effect. One of the views depends on *mental effort*. In essence, mental effort is the amount of cognitive resources spent processing information (Kirschner & Kirschner, 2012). It can be conceptualized as the mediator between task characteristics and the subject's information processing capacity (Shenhav et al., 2017). Mental effort can be extended to direction, duration, intensity, and strategy development processes (Bonner & Sprinkle, 2002). There are two different basic components to understand the amount of mental effort self-reports and response latency. The first self-explanatory concept, self-reporting, depends on the subjective judgments of participants. On the other hand, response latency is operationalized as the amount of time spent on the target. A longer response latency is indicative of higher mental effort than a shorter response latency. Early work displayed a positive relationship between mental effort and memory performance in reliance on memory (Tyler et al., 1979). Since self-generating the stimuli requires higher mental effort than reading, the higher spending of effort leads to the generation effect (Griffith, 1976; Jacoby et al., 1979; McFarland et al., 1980). However, there are some limitations to the mental effort theory itself, beginning with measuring effort. Even if the researcher can establish exact and discrete tasks in terms of their predicted effort, the perceived effort by the subjects does not necessarily

display a good fit. Secondly, the later studies on the generation effect and mental effort theory found that the generation effect is not in line with the mental effort assumptions. For instance, Foley and Foley (2007) challenged the mental effort theory with several experiments. Their design used easy and difficult anagram tasks, leading to higher memory performance for the generated stimuli. However, the results did not differ within the difficulty levels, revealing that mental effort theory fell short for the generation effect. In his meta-analysis, McCurdy (2020) states that mental effort falls short and encourages one to consider the type of processing rather than the amount of processing.

Maybe the best-known theory for the generation effect is the *two-factor or multi-factor theory*, which explains improved memory through item-specific and item-relational associations. To better understand the multifactor theory, item-processing accounts should be elaborated. During the encoding of items, the experiment procedure may be more beneficial for one of the following: item-specific and item-relational. Item-specific processing refers to processing the distinctions of each item from the rest of the list. Item-relational processing, on the contrary, refers to processing the similarities among different items (Hunt, 2013). Researchers claim that the generation effect emerges because the generation effect can make use of item-specific and item-relational processing more effectively than the passive reading condition. Particularly, the generation condition may make use of item-relational information more effectively, (Meyers-Levy, 1991), particularly when the test phase also requires processing relationships across or within the lists. (Winstanley, 1996).

While item-specific processing may be more sensitive to recognition tests, free recall is mostly influenced by item-relational processing. Furthermore, the cue-target relation mostly influences the cue-recall tests, which also benefit from the item-specific processing. Supporting a multifactor account, McDaniel and colleagues (1990) displayed the generation effect through free recall in between-subject designs. They later stated that categorizing the generation effect is just an artifact of a radical positioning. Similarly, Hirshman and Bjork (1988) claimed that the generation effect resulted from item-specific and item-relational processes, revealing the generation effect in a test manipulation between cued-recall and free-recall. Sustaining the two-factor theory, a recent meta-analysis by McCurdy (2020) stated that although the magnitudes differ, the generation effect could be observed through recognition, cued-recall, and free-recall. Unlike the item-relational claim for the generation effect by Hirshman and Bjork (1988), the higher memory performance for the cued-recall compared to free-recall was about a cue-target relation, also known as intra-item processing.

1.1.2. Generation Tasks

Apart from the item processing-based theories, the generation tasks could also provide strong evidence for the mechanisms of the generation effect. Participants usually encounter word lists or word pairs in a typical generation task. The nature of the generation task can be divided into two subcategories: perceptual and semantic. Perceptual tasks depend on the physical properties of the stimuli; for example, a typical letter rearrangement task requires replacing the indicated letters to generate the

correct form of the word (e.g., clod – c _ _ _). On the other hand, the semantic tasks focus on the meaning of the words; for instance, a typical antonym task requires participants to find the antonym of the cue word to generate the target word (e.g., new – o _ _). Even though most studies use one type of generation Mulligan (2002) compared two tasks that were semantic and non-semantic. Semantic generation task was realized through associations, and whereas non-semantic generation was realized through letter transposition, rhyme, and word fragments. Although each generation rule led to higher memory performance than reading, the semantic generation had more benefits than the non-semantic generation.

Mulligan's finding for superior memory for semantic than non-semantic generation tasks particularly makes sense in the context of levels-of-processing (LOP) framework. In an early work by Craik and Lockhart (1972) about LOP, they showed that the type of processing that the participant performs during encoding can change memory performance. If the participant is asked to process the materials through their physical attributes, thus process it shallowly, this leads to lower memory performance than deep processing conditions, in which the participants process the items in terms of their meanings. Studies have been replicating the LOP, and one of the most used comparisons was between perceptual and semantic tasks. Since the shallow process usually refers to the attributes regarding the physicality of the material, such as brightness, color, position, and font, a perceptual task can be categorized within the shallow process. On the other hand, semantic tasks grasp the deeper process by applying the material's meaning, interpretation, and implications (Craik, 2002).

Accordingly, in a typical LOP design, the semantic attributes of the stimuli lead to higher memory performance than perceptual attributes. However, Bertsch and colleagues (2007) examined the generation effect, revealing counterfactual meta-analysis results against the LOP predictions. According to the study, perceptual generation tasks such as letter rearrangement, rhyming, and word fragments have higher effect size scores than semantic tasks such as association, synonym, and antonym.

Although research seeks to understand the magnitude of the generation effect through different tasks, there is no direct study to figure out the perceptual generation tasks and semantic generation tasks within the same experimental procedure. Therefore, one of the current research aims is to understand the generation effect through levels of processing manipulation by applying perceptual and semantic generation tasks within the same parameter.

1.2 Metamemory

Metamemory refers to the cognition of our cognition and is related to monitoring, regulating, and controlling memory performance (Nelson, 1990). Metamemory gives insight into the decision-making and evaluation process for human memory and learning abilities.. Metamemory is important from a practical perspective because if we cannot monitor and regulate our learning and memory, we might have suboptimal memory performance for the materials we learn (Veenman et al., 2006). Therefore, it is critical to include metamemory tasks to display the results regarding awareness of one's memory performance.

1.2.1 Judgment of Learning

Although there are different methods to assess metamemory performance, such as "feeling of knowing," "global rating," and "confidence rating," one of the most used techniques is Judgment of Learning (JOL). During a classic immediate JOL procedure (Rhodes, 2016), participants are asked to judge their learning by scoring themselves on a scale after the item's presentation (e.g., "What is your probability of recalling the item?"). Although some studies show a good fit between actual and predicted memory performance, not all judgments about memory are accurate (Rhodes, 2016). In other words, the actual and predicted memory performance may differ from each other, producing a metamemory illusion.

The most commonly used framework to understand how individuals make their JOLs is Koriat's (1997) cue-utilization framework. Koriat (1997) argued that participants can use different types of cues, and to the extent that the used cues are diagnostic of actual memory performance, participants can produce higher accuracy regarding their predictions. Koriat proposed three different types of cues that participants use for making their judgments: intrinsic, extrinsic, and mnemonic. The first type of cue is intrinsic cues. In a typical intrinsic cue utilization, participants consider the intrinsic characteristics of the stimuli and a prior classification of learning the stimuli (e.g. word frequency). The extrinsic cues are the second set of cues crucial for metamemory judgments. Participants may use particular cues that are specific to the ongoing experimental procedure, such as study time, font size or number of repetitions during encoding. Finally, participants may use mnemonic cues, which is the interaction of participants with intrinsic and extrinsic cues. Although the

mnemonic cues are a different category, they are highly affected by the intrinsic and extrinsic cues. For example, participants may use word frequency, an intrinsic cue, as a way to evaluate their ease of learning, which is a mnemonic cue.

Mnemonic cues may depend on belief-based or experience-based processes (Rhodes, 2016). The belief-based process to JOLs, also known as the theory-based processes, refers to participants' prior beliefs, general knowledge, and assumptions about the experiments. If participants have a priori belief about the manipulation, they can guide their memory predictions in line with that belief. This is a conscious process, and participants can modify their JOLs according to their beliefs. On the other hand, the experience-based approach to JOLs, also known as the non-analytical approach, refers to all the experiences that participants confront during the experiment that are directly transformed into memory predictions without conscious effort. For example, participants' difficulties during the experiment may cause them to produce lower memory predictions even if they are not doing this consciously (Koriat & Bjork, 2006). One of the critical methods to understand whether the JOLs given under the experience-based approach is to control the reliance on the ease of learning with response latency.

1.2.2. Fluency Effect

One of the most frequently used cues for making memory predictions is fluency. According to the fluency effect, participants consider their subjective experience for the ease of learning (Oppenheimer, 2008). There are three types of fluency effects differentiating in their positioning for ease of learning: Retrieval fluency (Serra &

Dunlosky, 2005; Son & Metcalfe), perceptual fluency (Hertzog et al., 2003; Rhodes & Castel, 2008; Undorf et al., 2017), and encoding fluency (Koriat & Ma'yan, 2005; Undorf & Erdfelder, 2011). All these accounts reveal that when participants experience easily perceived, retrieved, or easily encoded items, they produce higher JOLs than when they deal with items that are hard to perceive, encode, or retrieve. Yet, the way that fluency affects JOLs is a controversial topic. Some studies show that participants have a priori beliefs about fluency and use these beliefs to make their JOLs (Mueller et al., 2013). This points towards a belief-based account of JOLs for fluency. On the contrary, some other studies show that the time spent in perceiving, encoding or retrieving an item is directly and linearly reflected through JOLs, implying an experience-based account of JOLs (Susser et al., 2017; Yang et al., 2018).

Even though most studies show that fluency and ease of learning are crucial factors in boosting JOLs, certain studies show contrary results. For instance, Castel and his colleagues (2013) applied the production effect manipulation to understand the JOLs. Participants read the items aloud or silently in a typical production effect design. The difference in encoding was revealed to be a significant memory benefit for the loud items compared to the silently read condition. Although reading the items aloud was disfluent compared to silent reading, participants scored higher for the stimulus read aloud. Castel et al. (2013) argued that participants may have a belief that production benefits memory even though reading aloud is more difficult than reading silently. Thus, not all studies show an advantage of fluent over disfluent items on memory predictions.

The generation effect may be a good way to pit the belief-based and experience-based views against each other since generation can be manipulated perceptually and semantically. The generation manipulation allows us to test the effort and time spent and search for any relationship between them and JOLs. The metamemory studies focusing on the generation effect routinely found that participants give lower JOL scores for generated material compared to read one, although generated stimuli produce higher memory performance than read ones (Besken & Mulligan, 2014; Besken, 2016; Yang et al., 2017; Froger et al., 2011). Participants consider the time spent and ease of encoding the stimuli since the generated material leads to higher response latency than the read condition. For example, Besken (2016) found that participants give higher JOLs for intact items than to-be-generated items. In Experiment 4 of the same experiment, the exact experimental procedure was introduced as a questionnaire scenario to participants to understand whether they have prior beliefs. The results of Experiment 4 revealed no difference between intact and to-be-generated items, meaning that participants learned how to judge two groups of material during the experiment. Experience-based processes could be hypothesized as the shorter response latency results in higher JOL.

On the other hand, Begg and her colleagues (1991) found a reverse pattern. In their study, participants considered the generation effect a beneficial manipulation for their later memory performance; therefore, they gave higher JOL scores for the generation tasks than for the read ones, even though generating typically takes a longer time. Thus, participants must have considered their prior beliefs on memory manipulation.

Notably, the current literature only covers the generation effect and metamemory, and the results are varied. We may be able to assess the contribution of experience-based and theory-based processes to the generation effect better when semantic and perceptual generation tasks are used within a single experiment. As the ease of generating semantic versus perceptual items differ, we may be able to see whether the ease of generating is directly transformed into JOLs or whether they are guided by some other belief-based processes

1.3 Metamemory Reactivity

Reactivity is the intentional or unintentional alteration of behavior due to the participant's awareness of the ongoing measurements, stimuli, and goals (Double & Birney, 2019). The historical Hawthorne Effect is one of the earliest studies on reactivity within psychology. The experiment focusing on Western Electric's factory in Hawthorne reveals that workers' awareness of outside observation altered their behavior. In later years, the reactivity component was an issue for the studies applying self-report techniques. This phenomenon is called "demand characteristics". When people are aware of the aims of the research, they alter their characteristics concerning the demands of the study. The reactivity of subjects leads to poor results, revealing unrealistic directions for further research. From a point of view, it is possible to ignore the reactivity by depending on the core feature of data analysis: the distribution of errors- *positive and negative errors will equate with each other*. However, the critical issue is to understand the size and direction of the reactivity under various methods

since the reactivity may also carry unique information in the learning and memory abilities of the human mind.

One of the areas that leads to reactivity in memory research is the JOL procedure. The JOL procedure requires participants to report their own judgments on their learning to understand participants' monitoring and awareness abilities. However, in doing so, participants must reconsider the stimuli and their likelihood of being retrieved in a later memory test, leading to a different process than the regular encoding phase. Even though the JOL procedure is typically used to assess self-monitoring of memory predictions, it also has the potential to produce reactivity for the actual memory performance. Recent studies showed that JOL-reactivity is a common finding across various encoding tests and can alter actual memory performance.

Even though JOLs have been used in memory research for the last four decades, JOL reactivity has not been a topic of investigation until recently. Some studies find JOL-reactivity for their materials. For example, Yang and his colleagues (2015) conducted a JOL study to understand the effect of reactivity. In their study, two groups were manipulated at the encoding stage. Participants in the learning group encountered a word list. On the other hand, participants in the JOL group provided a JOL for a 1 day-delay recognition test after the exposure of each word and the results showed that providing a JOL enhanced the actual memory performance. Janes and her colleagues (2018) used related and unrelated word pairs. Participants were assigned to JOL or no-JOL groups and completed a cued-recall test. They found a memory enhancement for related word pairs in the JOL group compared to the no-JOL one. In another

study, Tekin and Roediger (2020) used the levels of processing (LOP) paradigm. Participants saw words under three types of yes/no questions varying on three different levels of processing. During a case task, participants were asked whether the word was written in lowercase or uppercase, which was the shallowest process. In the deeper processing condition, participants processed words by indicating whether or not the critical word rhymed with the following word. In the deepest condition, participants were required to determine whether the word was in the correct category (e.g., "pencil" is a type of vegetable). Following the literature, they previously expected no reactivity for the shallowest task since the relatedness is quite low. Since the yes/no question would activate a deeper process for the rhyme and the deepest semantic process for the category, there should be a boost for the LOP effect. The results across two experiments with similar procedures showed that the shallowest process benefited more from JOL reactivity than the deepest process. Although the LOP effect was still observable, the magnitude was diminished.

On the contrary, some other studies show that JOL-reactivity does not occur for some other manipulations. For example, Tauber and colleagues (2015) compared the retrieval practice with the JOL procedure in another study. Participants saw unrelated word pairs in each group and completed a cued-recall test. The memory performance did not differ between providing a JOL and retrieval practice.

A meta-analysis by Double and his colleagues (2018) indicated that there was no significant effect of JOL on overall memory performance; however, through a subgroup analysis, they stated that there was a moderate positive reactivity for only

related word pairs and word list meaning that JOL group may outperform no-JOL group. They also reported no reactivity effect for unrelated word pairs or mixed-list material of related and unrelated pairs. The meta-analysis concluded that there was a positive reactivity for the JOL condition if the material consisted of related word pairs.

1.3.1 Cue-Strengthening Hypothesis

There are various hypotheses aiming to explain the JOL-reactivity effect. One of these is the cue-strengthening hypothesis. The cue-strengthening hypothesis states that providing a JOL strengthens the cues that contribute to the JOL score. In doing so, the strengthened cues lead to higher memory performance, informing the JOL reactivity. One of the first direct studies of the metamemory effect on memory performance was structured by Soderstrom (2015), which led to the cue-strengthening hypothesis. Soderstrom proposed that the JOL procedure fosters the higher relatedness of cue-target relationships, leading to a higher memory performance for the read stimuli during a cued-recall test. During the experiment, participants either had to generate (e.g., Window-G_ _ _ _) or passively read (e.g., Spring-Flower) semantically associated words, which typically produces higher memory performance for generated than read words. Participants were assigned to JOL or no-JOL conditions for this experiment. The results were similar for generated words between groups. However, the recall performance of participants in the JOL condition for the reading task was significantly higher than that of those in the no-JOL condition. Soderstrom argued that making a JOL increases the association of an already existing strong intrinsic cue

(i.e., *relatedness*). The increased association of the intrinsic cue enriched the memory performance.

Until now, the JOL reactivity has been discussed with a specific focus on encoding manipulations. Recent studies found that JOL reactivity also relies on retrieval conditions. The magnitude, direction, and existence of JOL reactivity are sensitive to the final type of the test. For example, Myers and her colleagues (2020) studied related and unrelated word pairs. Participants reacted positively to the related word pairs when they encountered a cued recall and recognition. The magnitude of the effect was larger for the cued-recall than recognition. Interestingly the free-recall test did not induce the memory benefit for the judged word-pairs. This finding implies that JOL reactivity is sensitive to the test type. Perhaps the JOL reactivity is a matter of encoding processes and interactions with the retrieval processes. Indeed, the cue-strengthening hypothesis proposes that JOL reactivity could be observed for a criterion test informing the cues used for JOLs (Soderstrom et al., 2015).

Even though previous studies do not necessarily consider it, another way to interpret these changes in levels of reactivity may be related to the Transfer-Appropriate Processing (TAP) theory (Morris et al., 1977), which plays a crucial role in the similarity of encoding and retrieval conditions. TAP states that increasing similarity between encoding and retrieval results in higher memory performance, and the memory performance should be lower when the encoding processes differ from the retrieval processes. In the original study for TAP, Morris et al. (1997) compared levels of processing with test manipulation. During the encoding phase, participants

made a Yes/No decision for the word according to the task. They were either deciding on a semantically required word (e.g., “The _____ had a silver engine” TRAIN / EAGLE) or with a rhyme required word (e.g., “_____ rhymes with legal” EAGLE / PEACH). As expected, participants were significantly more likely to recognize the stimuli with semantic attributes in a regular recognition test for the target words (e.g., “TRAIN” and “EAGLE”). However, participants were much more successful in recognizing the rhyme required when they were asked to recognize the items rhymed with target words (e.g., “RAIN” and “REGAL”).

With a combination of cue-strengthening, LOP, and TAP theories, Rivers and her colleagues (2023) conducted a series of experiments to understand JOL reactivity. During the encoding phase, participants saw categorically-cued pairs (e.g., A type of gem - Jade) or letter-cued pairs (e.g., Ja-Jade). Depending on the material type, participants were assigned to JOL or no-JOL groups and completed a cued-recall test. (e.g., A type of gem - “Answer”; Ja - “Answer”). The results showed that categorically-cued pairs led to higher memory performance than letter-cued ones. In support of the cue-strengthening hypothesis, the memory benefit for making JOLs was significant for only categorically cued pairs. The letter-cued pairs did not display any memory advantage for the group making JOLs. In search for the TAP, they altered the test types for Experiment 3. The encoding of the phase of Experiment-3 was identical to Experiment-1 except for two study cycle procedures (they switched to one cycle due to a lack of meaningful memory difference between cycles). The main change was related to the test type. In Experiment 3, participants were assigned to

either cued recall or letter-cued recall. Overall, the cued-recall test displayed a higher memory performance than the letter-cued recall test. There was a positive JOL reactivity for the categorically cued pairs in the cued-recall test, but there was no difference in providing JOLs for letter pairs in the cued-recall test. For the letter-cued recall, the memory performance was higher for letter-cue pairs than for categorically-cue pairs. However, there was no effect of JOL presence. In other words, the memory advantage of making JOL for categorically cued pairs was only observable for a criterion test. Also, a test without fostering the same cues that inform JOLs failed to display a reactivity effect. Those outcomes provided a convincing argument for cue-strengthening theory. Lastly, participants were assigned to either JOL or no-JOL groups and were encountered with categorically cued and letter-cued pairs, followed by a free recall test. The results did not display any JOL presence advantage.

Interestingly, letter-cued pairs displayed a negative reactivity. Again, a test without a criterion about the cue-target association failed to display the JOL reactivity. These results are compatible with a TAP perspective. As expected by TAP, categorically cued pairs were significantly less recalled during the letter-cued recall test. Also, the cued pairs reached a higher memory performance when there was a cued-recall test. The JOL reactivity did not show a memory advantage for the free recall test. From the reactivity perspective, the cue-strengthening hypothesis focuses on the criterion test that fosters the cues used for the JOL task (i.e., cue-target semantic relationships). Similarly, the TAP perspective informs a higher memory benefit for a criterion test defined as the similarity between encoding and perceptual cues. In other words, the reactivity predictions of the cue-strengthening hypothesis and TAP are similar. On the

other hand, the results cannot fully support the TAP perspective. Across several experiments, they did not ask participants to recall the letter-cued pairs through a category cue. TAP would ideally predict that a letter-cued pair would reach a higher memory performance through a letter-cued test than categorically-cued pairs.

The third aim of the current research is to understand JOL reactivity on the generation effect using perceptual vs. semantic processing cues from a retrieval perspective.

Although cue-strengthening theory informs that increased relatedness indicates a larger effect of JOL reactivity for a cued-recall test, Tekin and Roediger showed that perceptually processed information benefited the most from JOL-reactivity rather than semantically processed material using a recognition procedure. However, the latter finding may not be a consequence of only encoding but the interaction and the match between encoding and retrieval phases. If JOL reactivity is affected by the match between encoding and testing, as the TAP framework claims, one would expect participants to show the highest benefit when encoding and retrieval cues match each other. Thus, participants who processed information perceptually at encoding should be more successful at retrieval when given the same perceptual cues, and semantic processing at encoding should lead to higher recall when the same type of cue is given at retrieval. In contrast, mismatching cue conditions between encoding and retrieval should typically produce lower recall. If the findings are a consequence only of encoding conditions (and not its match with retrieval), the level of (mis)match should not change the amount of reactivity when the retrieval type changes.

Thus, Experiments 2 and 3 aimed to investigate the underlying processes for JOL-reactivity by pitting encoding-related explanations for the generation effect against the TAP perspective, which focuses on the match between encoding and retrieval.

1.4 Aims of the Current Project

The first critical question was whether the generation effect emerged with Turkish words. To our knowledge, no study focused on the generation effect with Turkish words as the material. Experiment 1 aimed to show this effect in a basic generation manipulation design. In Experiment 1, participants saw word pairs so that each target word was paired with either a perceptual cue (BOLD-COLD) or a semantic cue (HOT-COLD). Participants either generated the word from its semantic or perceptual associate (BOLD-C _ _ _) or processed the word passively. Although Bertsch and colleagues (2007) stated that perceptual generation resulted in larger effect sizes, the LOP paradigm would argue against that meta-analysis. Since generation studies do not usually search for the difference between semantic and perceptual generation tasks, a direct study including both types of tasks within the same design would be beneficial to understanding the generation effect. Nevertheless, there were two opposite predictions for the expected memory performance. First, participants may display higher memory performance for perceptually generated words than semantically generated ones, as reviewed by Bertsch et al. (2007). On the other hand, semantically generated words may lead to higher memory performance than perceptually generated ones (Craig & Lockhart, 1972). A recognition test followed the encoding phase. Foreshadowing for the Experiments: Experiments 1 applied a

recognition test, and Experiments 2 and 3 replaced it with cued-recall. Additionally, Experiments 2 and 3 solicited a JOL task for half of the participants.

The second main question of the current set of studies was to investigate how the generation effect may impact metacognitive judgments. For this, participants provided JOLs after each encoding trial in Experiments 2 and 3. We particularly looked into whether the disfluent generation manipulations produce lower JOLs than the fluent read conditions and whether the pattern could be altered through semantic and perceptual attributes. Moreover, the effects of experience vs theory-based processing were also investigated by looking into the RTs at each encoding condition.

Finally, the third main question of these studies was to investigate the JOL-reactivity for the generation effect. In order to do this, in Exp. 2 and 3, half of the participants provided JOLs at encoding, whereas the others were asked to do a filler task. We aimed to see whether making JOLs increases memory performance differentially for perceptual and semantic cues. This was tested in Exp 2. As a final point, Experiment 3 aimed to see the effects of the match between encoding and retrieval.

CHAPTER 2

EXPERIMENT 1

In Experiment 1, we examined the memory performance of the generation effect through perceptual and semantic cues. For this purpose, perceptual and semantic processing are operationalized as levels of processing and generation manipulation as the generation. Since there was no previously used Turkish word list for the generation manipulation, we created a new material using Turkish words.

Each target was paired with a perceptual cue and a semantic cue. Since the perceptual relation depended on the physicality of the words, a rhyming task was applied for the perceptual processing. On the other hand, an association pairing method was used for the semantic processing. If the levels of processing manipulation are effective, then participants should have longer response latency for semantic processing than perceptual processing. Moreover, applying LOP would indicate a higher memory performance for semantic processing than perceptual processing.

Apart from the level of processing manipulation, Turkish words were not previously used for the generation effect. Although the generation effect was found in languages other than English, it was another external validity concern. Generation manipulation was applied using a word-stem method. If participants consolidate a higher memory

performance for the generated words than read ones, this will imply the generation effect.

Regarding the interaction between levels of processing and generation, there are two possible and opposite directions in memory performance. Levels of processing manipulations typically show that when participants process information through perceptual attributes, their memory performance is lower than when they process information through semantic characteristics of the stimuli (Craik & Lockhart, 1972). In contrast, the meta-analysis performed by Bertsch et al. (2007) has reviewed a higher effect size for perceptual than semantic generation manipulations. Thus, if the LOP manipulation is valid, one would expect higher memory performance in semantic than perceptual generation manipulation. Nevertheless, if our results mimic Bertsch's meta-analysis, the memory performance should be higher in perceptual generation conditions.

2.1 Method

2.1.1 Participants

Ethical approval was obtained from the Bilkent University Ethics Committee. Typical within-subject generation studies used 24-32 participants. In order to have a .90 power at $\alpha = .05$ for a medium effect size of $f\text{-test} = .25$, G-power stated a sample of 30 participants. Since there were eight counterbalancing conditions, we decided to stop at 32. Participants were either recruited from Bilkent University through a general education course to be reimbursed with coursed credits, or they were volunteers.

2.1.2 Design

A 2 (generation: generate, read) x 2 (level of processing: semantic, perceptual) within-subject design was used. Participants encoded the target word using one of the following: generated or read. Participants completed the target word by generating the stimulus for the generate condition. For the read condition, the given target word was completed, and participants typed the target word to ensure the processing of the stimuli. Participants processed the cue-item relationship through one of the following levels of processing. In the semantic level of processing, the cue-item relationship was related to the cue-to-target association strength elicited through the meaning of the cue and item. The semantic relationship was any relationship that resulted in a high association between the cue and target words (e.g., antonym, synonym, and close meaning). In the perceptual level of processing, the cue-item relationship was related to the perceptual characteristics of the cue and item. In the perceptual relationship type, no cue-to-target association was elicited through the meaning of the cue and item to eliminate the possible multiple cue effect.

Participants encountered any of the following tasks:

Semantic Generation Condition: Semantic generation was operationalized as a generation-with-meaning task for the experiment. In the association task, participants were presented with a concept in which the cue word was highly related to the target word. In some cases, most antonyms and synonym words were also highly associated; however, it was not limited to these conditions.

(e.g., *Hot – C _ _ _*)

Perceptual Generation Condition: The perceptual generation condition was operationalized as a generation-with-rhyming task in which participants were given a perceptually similar-looking pair and needed to generate the target from the cue through rhyming.

(e.g., *Bold* – C _ _ _)

Read Condition: They typed the target word to ensure that participants processed the word pairs. Therefore, the read conditions were operationalized as a writing task. Participants were presented with semantic or perceptual processing during the read condition without generating the target word.

(e.g., *Bold* – *Cold*, *Hot* - *Cold*)

2.1.3 Materials

A total of 96 target words were paired with two different cue types, which were semantic and perceptual (Tekcan & Göz, 2005) and frequency ratings were standardized (Göz, 2020). 48 words were randomly chosen for encoding conditions. 48 words were randomly divided into four lists and assigned to one of the four conditions: semantic generation, perceptual generation, semantic read, and perceptual read. The lists were counterbalanced across conditions for the generation and level of processing, such that each target word was presented in each condition to an equal number of participants. An additional 12-word pair was used for the practice phase to acquaint participants with the task, and three trials were performed in each condition. The order of presentation was randomized anew for each participant.

2.1.4 Procedure

The experiment took place in the cognitive lab area of the Psychology Department. Participants were tested individually. The experiment consisted of four phases: The practice phase, encoding phase, distraction phase, and memory test phase. All phases were presented on a monitor using E-prime presentation software. Before the encoding phase, participants were trained in a shortened version of the experiment. In the encoding phase, participants were presented with word pairs across four conditions: Generation (Generate vs Read) and Level of Processing (Semantic vs Perceptual). Figure 1 displays the schematic description of Experiment 1.

After obtaining informed consent from the participants, they saw a computer screen briefly informing them about the experiment. Participants encountered several word pairs requiring the following conditions: semantic generation, perceptual generation, and read conditions. For semantic generation, they were required to use the meaning of the cue word to generate the missing parts of the target word through association. The perceptual generation was required to use the cue word's underlined letters to generate the target word's missing parts through rhyming. For the reading task, they typed the word on the right side of the screen. For all conditions, they typed the whole word. The missing number of letters was indicated with underlines, meaning each underline corresponded to a letter. Before exposure to each word pair, there was an instruction screen for the upcoming condition. There was no time limit for the condition instruction. It stayed on the computer screen until the participants typed the relevant key: Generation with a meaning task: "a" ("Anlam ile üretim" in Turkish), Generation with rhyming task: "k" ("Kafiye ile üretim" in Turkish) and Writing task:

"y" ("Yazma" in Turkish). Lastly, they were informed that there would be a memory test about the words they had seen.

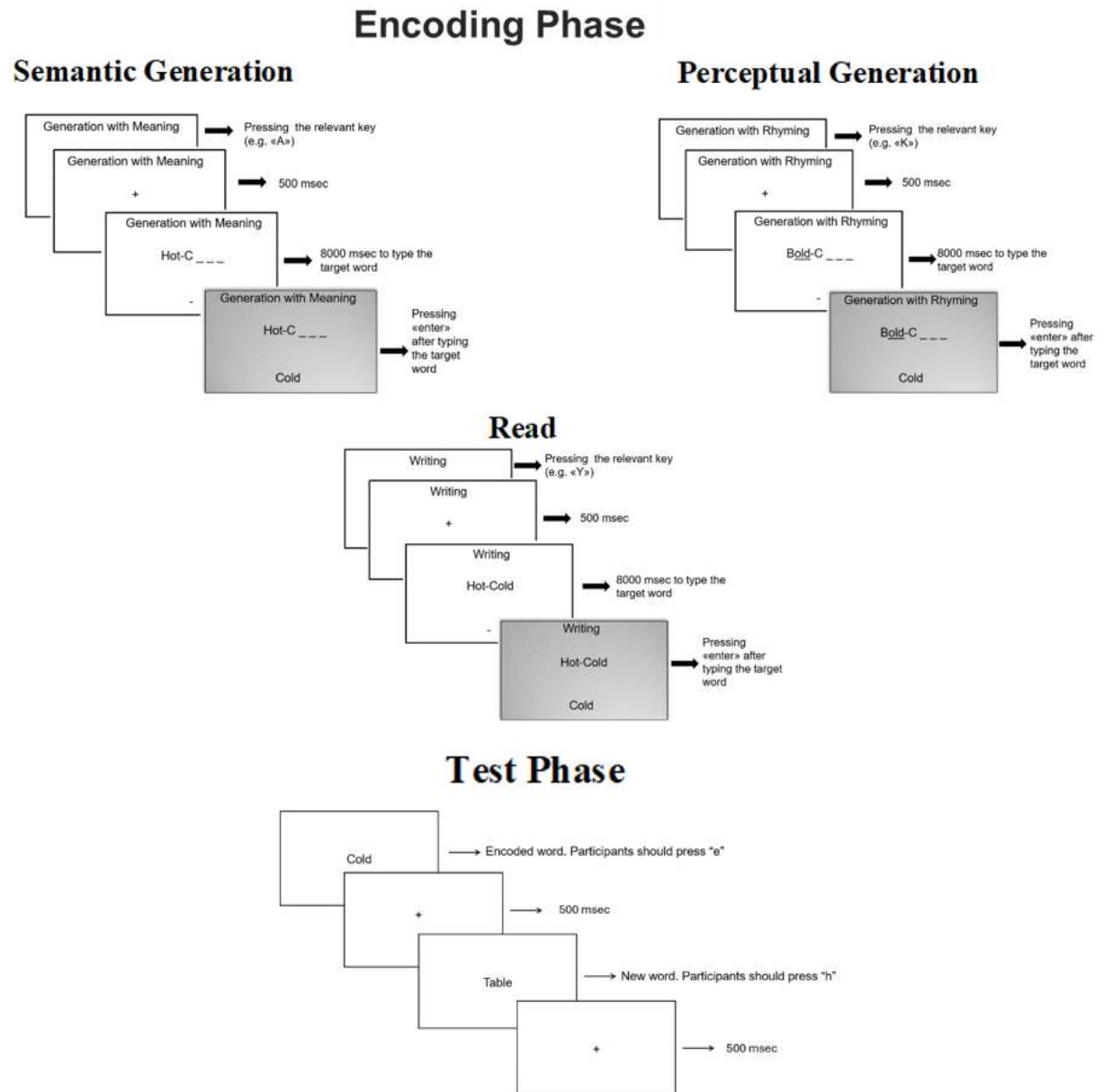


Figure 1. Experiment 1 Schematic Description. Read conditions included both semantic and perceptual associations (e.g. hot-cold, bold-cold). There were mathematics questions between phases.

Before the encoding phase, there was a practice phase in which they were presented with each condition to prepare for it. The practice phase was identical to the encoding phase. After participants typed the appropriate key for the condition instruction screen, a 500 ms fixation was followed by the word pairs. The conditions for presenting the word pairs were randomized. The trials were experimenter-paced and set at 8000 ms. For each condition, participants were allowed to type the target word for 8000 ms, and typing was required to be finalized by pressing the "enter" button. If they could type the target word before 8000 ms, the screen would turn gray, and no further typing was possible. The same presentation time was aimed for each pair, regardless of their typing duration. If they could not type the target word within 8000 ms, the procedure automatically started with a new condition and word pair. The first press following the presentation of the word pairs was the first key reaction time and pressing "enter" was the total reaction time.

The distraction followed the encoding phase, in which participants were asked to do simple mathematical calculations for 3 minutes to prevent them from rehearsing their answers.

After completing the distraction phase, the test phase instructions were presented to participants. For the test phase, participants saw the 48 target words they had encountered during encoding and 48 new words. By pressing a key, they needed to indicate whether they had seen a specific word. If they thought they remembered the item, they were required to press "e" (evet means yes), and if they did not think they recognized the item, they were required to press "h" (hayır means no). After finishing

the test phase, participants were debriefed and thanked on the last screen. If they had any questions, they could ask the experimenter. The experiment was expected to take about 30 minutes.

2.2 Results

Table 1. Means of Response Latency and Memory Performance Results

	Perceptual		Semantic	
	Generation	Read	Generation	Read
Response Latency	1672 (282)	1115 (197)	1924 (613)	1132 (233)
Memory Performance	0.75 (0.19)	0.55 (0.20)	0.84 (0.21)	0.59 (0.24)

Note: Numbers in parentheses refer to the standard deviations

In order to understand whether the LOP paradigm has been effective, the response

latencies have been analyzed through a 2 (LOP: Perceptual vs Semantic) x 2

(Generation: Generate vs Read) repeated measures ANOVA. The alpha level was set

at .05 for this and all subsequent analyses. Additionally, all post hoc analyses were

conducted with Bonferroni correction. The analysis showed the main effect of LOP, F

$(1,31) = 7.97, p = 0.008, \eta^2_p = 0.20$. The post hoc analysis revealed that semantic

processing significantly leads to longer response latencies than perceptual processing t

$= 2.82, p = 0.008, d = 0.37$. The main effect of Generation was also significant, F

$(1,31) = 123.65, p < .001, \eta^2_p = 0.80$, and the post hoc test showed that participants

spent more time before typing in the generation condition compared to read, $t = 11.12,$

$p < .001, d = 1.83$. Furthermore, the interaction of LOP and Generation was also

significant, $F(1,31) = 6.08$, $p < 0.019$, $\eta^2_p = 0.16$. The post hoc comparisons yielded that participants were slower to respond to semantic generation than perceptual generation, $t = 3.74$, $p = 0.002$, $d = 0.68$; however, there was no difference between semantic read and perceptual read conditions $t = 0.26$, $p = 1.000$.

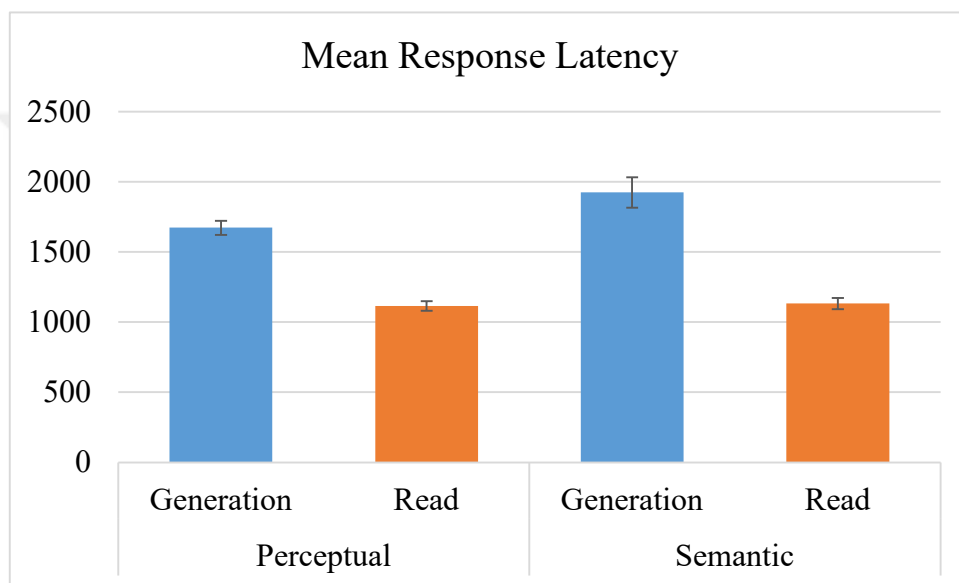


Figure 2. Experiment 1 Response Latency

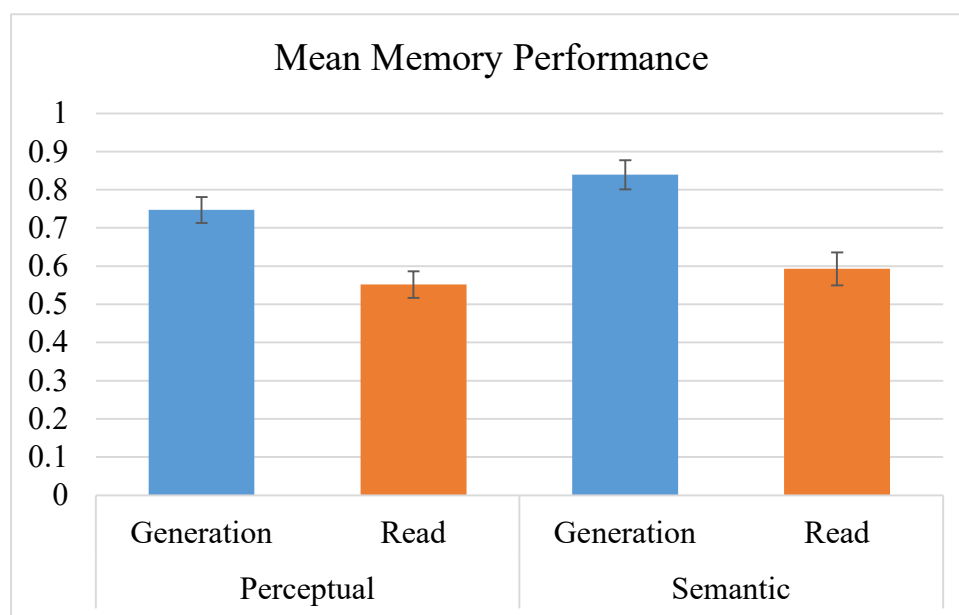


Figure 3. Experiment 1 Memory Performance

The recognition test performance was analyzed for the memory results through a 2 (LOP: Perceptual vs Semantic) x 2 (Generation: Generate vs Read) repeated measures ANOVA. The main effect of LOP, $F(1,31) = 9.18, p = 0.005, \eta^2_p = 0.23$, and the post hoc test revealed that semantic processing leads to higher memory performance than perceptual processing, $t = 3.03, p = 0.005, d = 0.31$. Also, the Generation was significant, $F(1,31) = 59.23, p < .001, \eta^2_p = 0.66$, and the post hoc test found that participants had higher memory performance for the generated words than the read ones, $t = 7.69, p < .001, d = 1.04$. There was no interaction effect, $F(1,31) = 1.14, p = 0.29$, but a planned comparison was conducted to understand whether there was a difference between the perceptual generation and semantic generation. The results showed that semantic generation produced higher memory performance than perceptual generation, $t = 2.84, p = 0.036, d = 0.43$. However, there was no difference between the perceptual read and semantic read, $t = 1.28, p = 1.000$.

2.3 Discussion

In the current research, participants encountered whole word pairs or generated the target word through cue words, depending on the task type, to recognize them later. Results showed that semantic processing operationalized as an association task leads to longer response latency than perceptual processing operationalized as a rhyming task. The current methodology fits the LOP hypothesis well since response latency has been considered when comparing the deep and shallow processes. Apart from LOP, participants responded slower when they needed to generate the word rather than just typing. Interestingly, the increased response latency was only particular to

cue types during the generation task. Participants spent similar time for the cue types when they needed to type, indicating that the reading task was easy to process for both perceptual and semantic cues.

The good fit of the LOP paradigm was also observable through memory performance. Consistent with the LOP paradigm, participants had higher memory performance for semantically processed words than perceptually processed ones. Likewise, the generation effect was successfully applied to Turkish words, indicating a higher memory performance for generated words than passively read ones. One of the crucial aims of the current research was to understand whether perceptual generation was a better manipulation to display the generation effect compared to semantic generation; the results were against the meta-analysis (Bertsch et al., 2007). Participants showed higher memory performance for the semantically generated words than the perceptually generated ones. Thus, the results were consistent with the LOP paradigm predicting that memory performance would be higher for the semantic attributes of stimuli than perceptual ones.

One of the possible explanations for the opposing memory results to meta-analysis is related to the mixed-list effect. First, a few studies focus on perceptual and semantic generation manipulation in the same experimental paradigm. The lack of comparisons within the same list may have affected the magnitude of the generation effect. In order to elaborate, we can consider two different experimental designs investigating the generation effect. In the first experiment, participants either generate the word pairs through a rhyming task or just passively read the word pairs. When participants need

to remember the word pairs either by perceptually generating or passively reading, it is more likely to have a larger effect size for the generation condition since they do not have any advantage during the read condition. In another experiment, the encoding phase would require participants to generate the word pairs through an association task or read the associated pairs passively. Although participants will probably remember the generated words similar to the former experimental design, they will be more likely to remember more words from the read conditions. In short, a mixed-list design provided participants with an opportunity to consider both cue types at the same time.

CHAPTER 3

EXPERIMENT 2

Experiment 1 showed that levels of processing and generation alter memory performance. Participants had higher memory performance when the word pairs were semantically processed than perceptually processed. Similarly, they were more likely to recognize the generated words than read ones. Importantly, there was a meaningful difference between the semantic generation and perceptual generation tasks. The results indicated a counter pattern to Bertsch and her colleagues' meta-analysis (2007) by providing evidence for a higher memory performance of the semantically generated words. The meaningful difference in cue types did not extend to the read conditions. A second aim of the project was to see how participants' memory predictions changed regarding their subsequent memory performance under perceptual and semantic generation tasks. Moreover, we investigated whether making predictions about one's own memory performance modifies memory performance. These questions were tested through Experiment 2.

Experiment 2 was similar to Experiment 1, except for two major differences. First, participants were randomly assigned to either the judgments-of-learning (JOL) or the no-JOL group. Second, participants recalled their responses through a cued-recall test

rather than a recognition test. There were two reasons for replacing the recognition test with the cued-recall. Although the generation effect is an item-specific processing and the recognition test is a good fit for understanding memory performance, a cued-recall test would also include the cue-target relation, a.k.a. intra-item processing. Also, JOL reactivity was found to be sensitive to the similarity of test type to the encoding procedure (Myers et al., 2020).

It is possible to observe a memory performance change due to JOL reactivity. Since all word pairs in the current study were related, there should be a higher memory performance across all conditions when participants provided a JOL (Double et al., 2018). Notably, Tekin and Roediger's (2020) article investigating the JOL reactivity through the LOP paradigm indicated that the JOL reactivity was larger for the shallow process when compared to the deep process. Nonetheless, JOL reactivity could not eliminate the LOP effect. Therefore, we expected a larger positive JOL reactivity for perceptual processing than semantic processing but still a higher memory performance for semantic processing. In line with that, the read condition would have a higher positive JOL reactivity for memory performance than the generation condition; however, the generation condition would still elicit a higher memory performance than the read condition.

There were two directions for the metamemory judgments. The fluency perspective would consider the ease of processing the stimuli operationalized through response latency. Since the semantic processing and the generation condition were conceived as more effortful and less fluent, the more fluent perceptual processing and the read

condition should produce higher JOLs. Regarding the interaction between the level of processing and generation, the JOLs were expected to align with response latencies. For the read condition, there would be no meaningful difference between semantic and perceptual conditions; however, the generation condition was expected to differ. The semantic generation should produce lower JOLs than the less fluent perceptual generation condition.

On the other hand, the production effect perspective informs the ability of participants to correctly assess the manipulation as an assistant for their later memory performance (Castel, 2013). If the current research displayed a production effect, memory, and metamemory performance would follow a similar trend.

Experiment 2 was preregistered prior to the creation of data and can be found through the link: <https://osf.io/e8hz4>

3.1 Method

3.1.1 Participants

Participants were recruited from Bilkent University through a general education course. Participants were native Turkish speakers and did not take any courses from Miri Besken at the time of data collection. They were reimbursed with course credits. A power analysis with G-power 3.1 showed that to obtain a power of .90 at $\alpha=.05$ for an f-test: .25, each group should include 30 participants, so we needed 60 participants. As each counterbalance consisted of 8 conditions, a sample size of 64 participants was settled. Thirty-two were assigned to the JOL group, and the remaining 32 were assigned to the no-JOL group. Since we added at least 70%

success criteria per each condition during the encoding phase, 8 participants were replaced following their dedicated counterbalanced conditions.

3.1.2 Materials, Design, and Procedure

The materials were identical to Experiment 1. However, the previous within-subject design was altered with a between-subject manipulation. Experiment 2 was a 2 (level of processing: semantic, perceptual) x 2 (Generation: generate, read) x 2 (judgment: JOL, no-JOL) mixed factor design. Similar to Experiment 1, the level of processing and generation was manipulated within subjects, whereas the judgment group was manipulated between subjects. Figure 3 depicts the schematic examples for the respective groups

The procedure was similar to Experiment 1, with two major differences. The first of them was about the encoding phase. Participants in the JOL condition were provided a self-paced JOL for each item within a new screen, asking them how likely they think they will recall the given item on a scale between scale ranging from 0 (I am sure I will not remember it) to 100 (I will definitely remember it). The pairs did not appear on the screen during the JOL process. In order to eliminate any possible time confound in the no-JOL condition, participants in the no-JOL condition entered random numbers between 0 and 100.

Secondly, the cued-recall test was applied. Participants saw each of the 48 cue words individually. For each cue word, they typed the corresponding target word. The test phase was self-paced. After finishing the test phase, participants were debriefed and

thanked on the last screen. If they had any questions, they asked and received answers from the experimenter. The experiment took approximately 30 minutes.

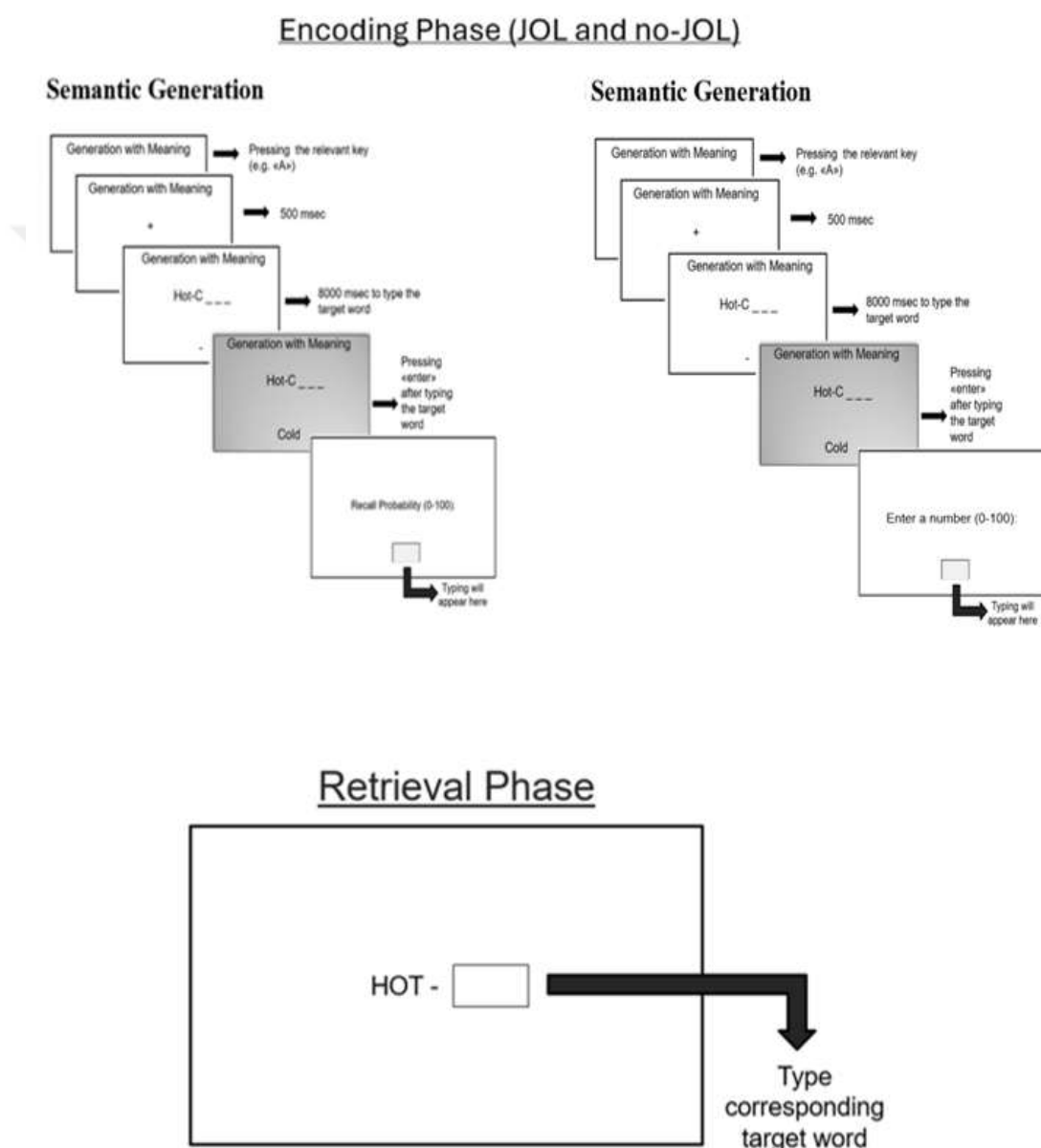


Figure 4. Experiment 2 Schematic Example. Participants were either in the JOL or no-JOL group and saw all word groups. There were mathematics questions between phases

3.2 Results

3.2.1 Response Latency

A 2 (level of processing: semantic, perceptual) x 2 (Generation: generate, read) x 2 (judgment: JOL, no-JOL) repeated measures ANOVA was conducted in which the judgment group was calculated as the between subjects. Similar to Experiment -1, LOP, $F(1,62) = 8.34, p = 0.005, \eta^2_p = 0.12$, and Generation, $F(1,62) = 262.44, p < .001, \eta^2_p = 0.81$ provided significant main effects. Also, there was an interaction effect of within manipulations LOP*Generation, $F(1,62) = 10.91, p = 0.002, \eta^2_p = 0.15$. Between-subject manipulation was insignificant for any interaction. The post hoc tests revealed that semantic processing was slower than perceptual processing, $t = 2.89, p = 0.005, d = 0.22$, and generation learning was slower than read learning, $t = 16.2, p < .001, d = 1.63$. In order to elaborate on the interaction of LOP*Generation, a planned comparison was conducted in which semantic generation was significantly slower than the perceptual generation, $t = 4.35, p < .001, d = 0.44$. However, there was no difference between perceptual read and semantic read, $t = 0.019, p = 1.000$.

3.2.2 Metamemory Performance

Table 2. Means of Metamemory and Memory Performance Results

	Perceptual				Semantic			
	Generation		Read		Generation		Read	
	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL
Metamemory	50.65 (19.31)	44.46 (14.20)	45.99 (19.2)	43.75 (13.96)	63.48 (18.92)	43.23 (15.29)	60.79 (18.13)	40.33 (13.74)
Memory	0.3 (0.21)	0.25 (0.15)	0.27 (0.20)	0.16 (0.18)	0.80 (0.15)	0.72 (0.16)	0.66 (0.26)	0.41 (0.23)

Note: Numbers in parentheses refer to the standard deviations

In order to understand the metamemory performance, a 2 (LOP: perceptual, semantic) x 2 (Generation: generation, read) repeated measures ANOVA was conducted for each judgment group. Since the no-JOL group was informed to enter random numbers, their results did not differ across conditions (See Figure 4).

On the other hand, the JOL group has shown significant differences. The results showed the main effects of LOP, $F(1,31) = 31.47, p < .001, \eta^2_p = .50$ and Generation, $F(1,31) = 9.83, p = 0.004, \eta^2_p = 0.24$. There was no interaction effect. Post hoc tests revealed that participants gave higher JOL scores for semantic processing than perceptual processing $t = 5.61, p < .001, d = 0.77$. Likewise, the generated words received higher JOL scores than the read words, $t = 3.13, p = 0.004, d = 0.20$.

Although the interaction was insignificant, a planned comparison was conducted to understand the trends. Each comparison was significant except for the semantic

generation and semantic read. The semantic processing did not differ across types of learning, $t = 1.54$ $p = 0.781$.

None of the JOL score analyses can be explained through a fluency perspective.

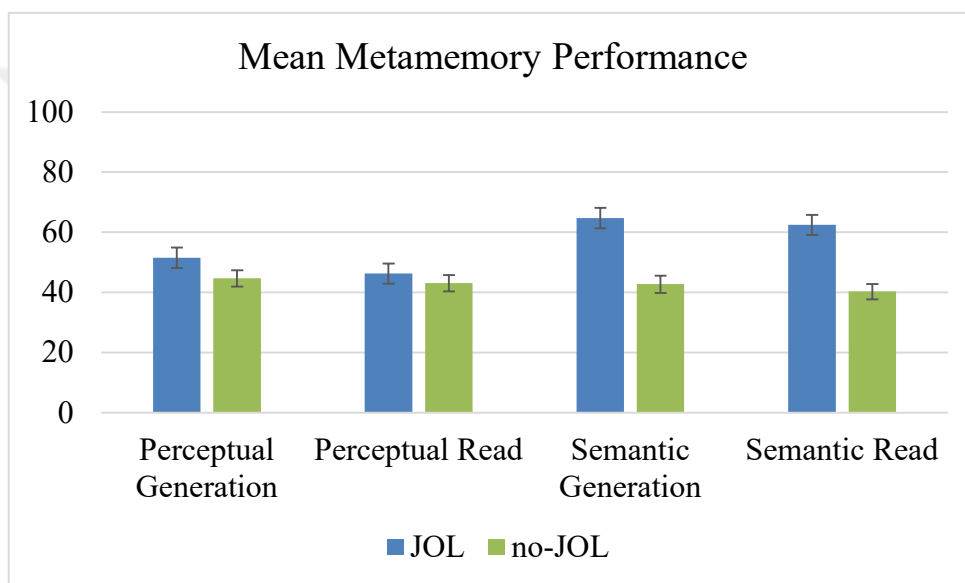


Figure 5. Experiment 2 Metamemory Performance

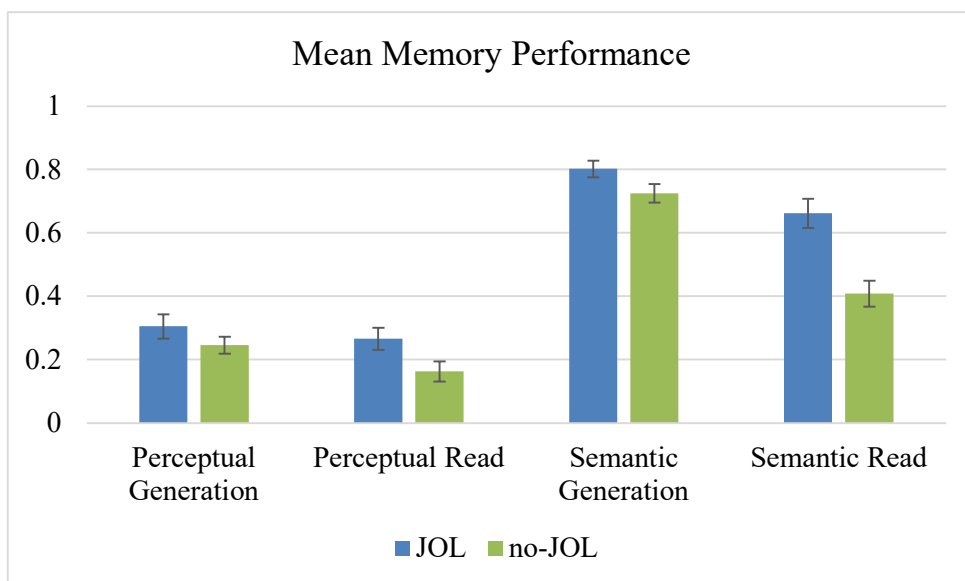


Figure 6. Experiment 2 Memory Performance

3.2.3 Memory Performance

A 2 (level of processing: semantic, perceptual) x 2 (Generation: generate, read) x 2 (judgment: JOL, no-JOL) repeated measures ANOVA was conducted in which the judgment group was calculated as the between-subjects (see Figure 5). The memory performance results replicated Experiment 1. The main effects of LOP, $F(1,62) = 463.49$, $p < .001$, $\eta^2_p = 0.88$ and Generation, $F(1,62) = 47.549$, $p < .001$, $\eta^2_p = 0.43$. The post hoc test revealed that semantic processing led to higher memory performance than perceptual, $t = 21.53$, $p < .001$, $d = 2.04$. As expected, the generated words also provided higher memory performance than read ones, $t = 6.90$, $p < .001$, $d = 0.73$. Also, the interaction of LOP*Generation was significant $F(1,62) = 26.06$, $p < .001$, $\eta^2_p = 0.30$. The post hoc comparisons showed that semantic generation produced higher memory performance than perceptual generation, $t = 19.57$, $p < .001$, $d = 2.47$; interestingly, there was no generation effect for perceptual processing, $t = 2.28$, $p = 0.18$. Lastly, semantic read led to higher memory performance than perceptual read, $t = 12.85$, $p < .001$, $d = 1.62$.

The between-subject effects were analyzed to assess the metamemory reactivity. The results showed that participants in the JOL group had higher memory performance than those in the no-JOL group, $F(1,62) = 10.98$, $p = 0.002$, $\eta^2_p = 0.15$. Post hoc comparisons of groups displayed that there was no reactivity of perceptual processing for either generated words or read ones. Also, semantic generation did not alter across groups. However, participants in the JOL group had higher memory performance for semantic read words than participants in the no-JOL group, $t = 5.13$, $p < .001$, $d = 1.28$.

3.3 Discussion

Experiment 2 replicated the response latency results of Experiment 1. Participants displayed the levels of processing manipulation by spending more time before pressing the first key in deeper conditions than in shallow conditions. Also, their response latency was longer when generating the words than reading. On the other hand, participants provided evidence for categorical awareness of levels of processing only when they generated the words. There was no difference between semantic read and perceptual read conditions.

The results of memory performance for Experiment 2 were also replicated. Participants displayed a higher memory performance for semantic processing over perceptual processing. The same trend was observed for the generated words. Also, semantic generation led to higher memory performance than perceptual generation. Bertsch et al. meta-analysis would predict that perceptual generation should have resulted in higher memory performance than semantic generation. However, both Experiment 1 and Experiment 2 revealed significant support for a higher memory performance of semantic generation over perceptual generation. Critically, the recognition results of Experiment 1 extended to the cued-recall results of Experiment 2, which displayed that proposed effect sizes attributing a larger effect size for a perceptual generation were not observable when both generation tasks were applied within the same experimental design.

Although the generation and LOP effects were consistent with Experiment 1, the current study had novel results. First of all, perceptual processing did not display a

generation effect. The overall performance of perceptual processing was significantly lower than in Experiment 1. For example, the no-judgment group, which was supposed to reveal results equivalent to Experiment 1, displayed a different pattern. The semantic read task had a higher memory performance over the perceptual generation. One of the possible explanations was that transfer-appropriate processing may have played a significant role in memory performance. Since perceptual processing of cue-target association was not as powerful as semantic processing, the outcome of the same cued-recall task may have been beneficial for semantic processing.

Secondly, there was a reactivity effect separating the groups. Participants in the JOL group had higher memory performance than those entering random numbers. However, planned comparisons revealed that the reactivity effect was only observable for the semantic read task. The remaining tasks of perceptual read, perceptual generation, and semantic generation did not benefit from providing a metamemory judgment. There are some crucial notes for the metamemory research. First, the separate analysis of semantic processing perfectly replicated Experiment 2 of Soderstrom (2015). Soderstrom used semantically related words during a generation task in that experiment and displayed a memory benefit for only read conditions. Participants did not display any difference in semantic generation, which already resulted in higher memory performance. Like that experiment, current research displayed a highly significant reactivity for semantic reading, and the increased memory performance still displayed a generation benefit. Interestingly, the reactivity

of perceptual processing was not significant for either generation or reading. If the results could have replicated the findings of Tekin and Roediger (2021), we should have observed a larger reactivity for the shallowest process, the perceptual read task. In contrast, the reactivity for the shallowest process was insignificant. As Double (2018) suggests, the relatedness was the key to positive reactivity, and the lack of significant reactivity for perceptual processing of generation and read task displays that their relatedness was lower than semantic processing. However, it should be considered that the current experiment used a cued-recall test favoring the cue-target relation, whereas Tekin and Roediger's study (2021) applied a recognition test. Again, the test type seems to play an important role not considered before within the LOP or generation effect reactivity.

Regarding metamemory performance, the current research found that participants knew about the memory benefits. Although prior research on the generation effect found a fluency effect on metamemory judgments, Experiment 2 displayed that participants had accurately judged their expected memory performance. Apart from the generation effect, they were also successful with LOP by considering semantic processing as more likely to be remembered than perceptual processing. Also, all cues had an additive effect, meaning that participants effectively considered the memory benefit for each and combined manipulations supporting the cue integration perspective on metamemory (Undorf & Bröder, 2020).

With regard to the fluency perspective, the metamemory judgments were not affected by the response latency. In other words, depending on the ease of learning the

material, the fluency effect is inadequate for understanding the current metamemory results. Although participants were still overconfident about their expected memory performance, their actual and expected memory performance displayed the same pattern supporting the production effect. The mentioned pattern means that participants gave higher scores for semantic processing over perceptual processing and generation learning over reading. In addition to that, they gave higher scores for the semantically read words over the perceptual generated ones, indicating that they did not only select one cue but also were aware of their overall memory performance.

A possible explanation for the reason that semantic processing was confidently selected over perceptual processing could be related to the fact that participants had a prior belief in LOP manipulation. As Undorf and Bröder (2020) stated, participants' immediate JOL scores were affected by their pre-study JOLs and *beliefs*. The belief that the level of processing as higher relatedness leads to better memory performance may be strengthened through experiments. When they scored on their expected memory performance, they might have considered the relatedness, which is more powerful between semantic processing than perceptual.

However, the same explanation cannot be stated for the generation effect. When Froger et al. (2011) asked participants to give a global rating between the recall probability of generated and read words before the experiment, they gave higher recall probability scores for the read condition. Therefore, it was a roadblock to understanding the accurate metamemory judgments for the generation effect since the fluency effect or prior belief cannot explain it. Participants knew that they spent more

time generating but were confident that the manipulation would help their later memory performance.



CHAPTER 4

EXPERIMENT 3

Experiment 2 strongly supported the Experiment 1 memory results by displaying that LOP and Generation Effect led to higher memory performance. Opposed to the meta-analysis of Bertsch et al. (2007), semantic generation still had a better memory benefit than perceptual generation in a cued-recall test. However, two important notes should be considered. First, perceptual processing does not display a generation effect. Secondly, LOP had a significantly larger effect size than the generation effect in explaining the memory performance.

Although a higher memory performance for the JOL group was expected, the reactivity effect was not in line with the expectation in sub-word groups. (Tekin & Roediger, 2020). The JOL reactivity was only significant for the semantic read task.

In addition, the metamemory judgments were unorthodox. Unlike most generation effect literature, participants accurately predicted their expected memory performance. The same pattern was observed for semantic processing over perceptual processing. The metamemory judgments of LOP and the generation effect cannot be understood through the fluency effect.

Experiment 2 displayed an extraordinarily successful memory performance regarding semantic processing. The LOP manipulation was so strong that the generation effect could not benefit perceptual processing, resulting in low memory performance. One of the reasons beyond the larger memory benefit of semantic processing might be the test type favoring semantic processing. Participants could easily develop a better cue-target relation for the semantic process over the perceptual one during the encoding process. Since the semantic process relied on the meaningful association of word pairs, it would not be surprising that participants did not care to learn the cue-target relation of perceptual processing, which was only rhyming. Like the cue-strengthening hypothesis (Soderstrom et al., 2015), the cued-recall test fostered the cue and target memory rather than just the target. Since participants failed to strengthen the cue and target relation for perceptual processing, the memory results did not differentiate between learning the target and learning the cue. Also, the generation effect did not have the largest effect size for memory performance. The perceptual processing should have resulted in higher memory performance for generated words over read words; however, there was no perceptual generation effect. Considering that the same cued-recall task favored semantic processing through the cue-strengthening hypothesis, Experiment 3 applied a cue-matching and cue-no-matching manipulation for retrieval. We expected that generated words would lead to higher memory performance than read ones. From the LOP perspective, we expected to replicate the results of Experiment 2 for cue-matching retrieval, and the cue-no-matching retrieval would result in lower memory performance.

From the reactivity perspective, the JOL group would provide evidence for higher memory performance than the no-JOL group. In other words, the results needed to replicate the Experiment 2. Additionally, we expected an interaction effect between matching and JOL presence. If the cue-strengthening theory were valid for the results of Experiment 2, the matching retrieval would lead to a larger effect size of reactivity than no matching retrieval. Since the JOL reactivity was found to be sensitive to the contextual drift between the encoding phase and the test type, the more related matching test would be a more significant benefit for the reactivity.

On the other hand, if cue-strengthening theory was not valid for the unexpectedly larger effect size of LOP, then semantic processing should lead to higher memory performance than perceptual processing regardless of the matching manipulation. In other words, there should be no interaction between matching manipulation and JOL presence.

4.1 Method

4.1.1 Participants

Participants were native Turkish speakers who had not attended any Miri Besken course during the experiment semester. They were either volunteers or recruited through a general education course. Participants attending the general education course were reimbursed with course credit. Similar to Experiment 2, we used 64 participants assigned into two groups randomly: JOL and no-JOL. Additionally, participants had to type at least 70% of words for each condition to be considered for

results. Five participants were replaced before the analysis according to their experiment conditions.

4.1.2 Materials, Design, and Procedure

The materials were identical to Experiment 1 and Experiment 2. However, we eliminated half of the materials. Since participants struggled with typing the correct words, we analyzed the error rates of word pairs. Then, we listed all word pairs in descending order according to the number of times the answer was mistaken. Afterward, we eliminated the first 48-word pairs from the list, resulting in 48-word pairs less prone to cause an error.

Experiment 3 included a new within-subject manipulation. Therefore, Experiment 3 was a 2 (LOP at encoding: perceptual, semantic) x 2 (Generation: generation, read) x 2 (LOP at retrieval: perceptual, semantic) x 2 (judgment group: JOL, no-JOL) mixed-factor design. Similar to Experiment 2, judgment groups were manipulated between subjects, whereas the rest were within subjects.

The procedure differed from the previous experiment with the additional manipulation of the cued-recall test. During the test phase, the cued-recall test was altered with a matching manipulation for cue words depending on the LOP (see Figure 5).

Participants saw the matching cue word with the encoding phase for half the time, whereas the rest was a no-match. The order of the presentation was randomized anew for each participant. Before the retrieval phase, participants were informed that:

“You will need to recall the target words. In order to ease the recall process, there will be cue words. Although some cue words were already presented with their targets,

there will also be novel ones. Nevertheless, your task is to type the target words. Let us assume that you have seen the word pair of *window-glass*. You may see the cue word window or the novel cue word class during the recall. In both cases, you are expected to type the target word *glass*.”

The self-paced test phase began after participants approved their understanding of retrieval manipulation. Following the ending, they were debriefed and thanked on the last screen, and their questions were answered. The experiment lasted about 30 minutes.

<i>Cue-Match Retrieval (Perceptual Generation)</i> At encoding: <u>Bold</u> – C _ _ _ At retrieval: Bold –	<i>Cue-No-Match Retrieval (Perceptual Generation)</i> At encoding: <u>Bold</u> – C _ _ _ At retrieval: Hot –
<i>Cue-Match Retrieval (Semantic Generation)</i> At encoding: Hot – C _ _ _ At retrieval: Hot –	<i>Cue-No-Match Retrieval (Semantic Generation)</i> At encoding: Hot – C _ _ _ At retrieval: Bold –

Figure 7. Experiment 3 Cue-Matching Manipulation Schematic Example

4.2 Results

4.2.1 Response Latency

A 2 (level of processing: semantic, perceptual) x 2 (Generation: generate, read) x 2 (judgment: JOL, no-JOL) repeated measures ANOVA was conducted in which the judgment group was calculated as the between subjects. The results were similar to Experiments 1 and 2.

Table 3. Means of Metamemory Performance Results

	Perceptual				Semantic			
	Generation		Read		Generation		Read	
	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL
Metamemory	53.39	40.23	49.62	40.84	59.52	39.88	58.43	42.36
Performance	(17.73)	(16.88)	(18.76)	(17.79)	(20.77)	(16.86)	(19.87)	(19.29)

Note: Numbers in parentheses refer to the standard deviations

Table 4. Means of Memory Performance Results

	Perceptual Encoding				Semantic Encoding			
	Generation		Read		Generation		Read	
	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL	JOL	no-JOL
Perceptual	0.32	0.26	0.22	0.18	0.09	0.01	0.04	0.06
Retrieval	(0.23)	(0.27)	(0.22)	(0.23)	(0.13)	(0.17)	(0.07)	(0.13)
Semantic	0.44	0.39	0.35	0.23	0.85	0.73	0.67	0.53
Retrieval	(0.19)	(0.28)	(0.24)	(0.28)	(0.17)	(0.25)	(0.21)	(0.26)

Note: Numbers in parentheses refer to the standard deviations

4.2.2 Metamemory Performance

In order to understand the metamemory performance, a 2 (LOP: perceptual, semantic) x 2 (Generation: generation, read) repeated measures ANOVA was conducted for each judgment group (see Figure 6). Since the no-JOL group was informed to enter random numbers, their results did not differ across conditions.

Similar to Experiment 2, the metamemory scores significantly differed for LOP manipulation, $F(1,31) = 5.49, p = 0.03, \eta^2p = 0.15$. The post hoc test revealed that semantic processing received higher JOL scores than perceptual processing, $t = 2.34, p = 0.03, d = 0.39$. Surprisingly, there was only a marginally significant effect of Generation, $F(1,31) = 3.74, p = 0.06, \eta^2p = 0.11$. The t-test analysis revealed that marginal significance was a trend towards giving higher scores for generated words compared to read ones. There was no interaction effect.

Although the difference between generated and read words was only marginally significant, the results replicated the unique metamemory findings of Experiment 2.

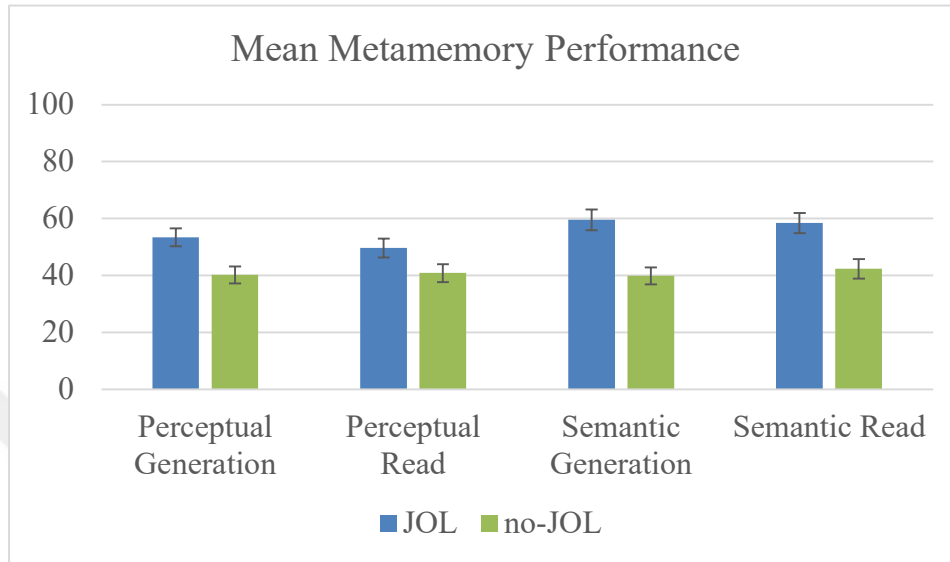


Figure 8. Experiment 3 Metamemory Performance

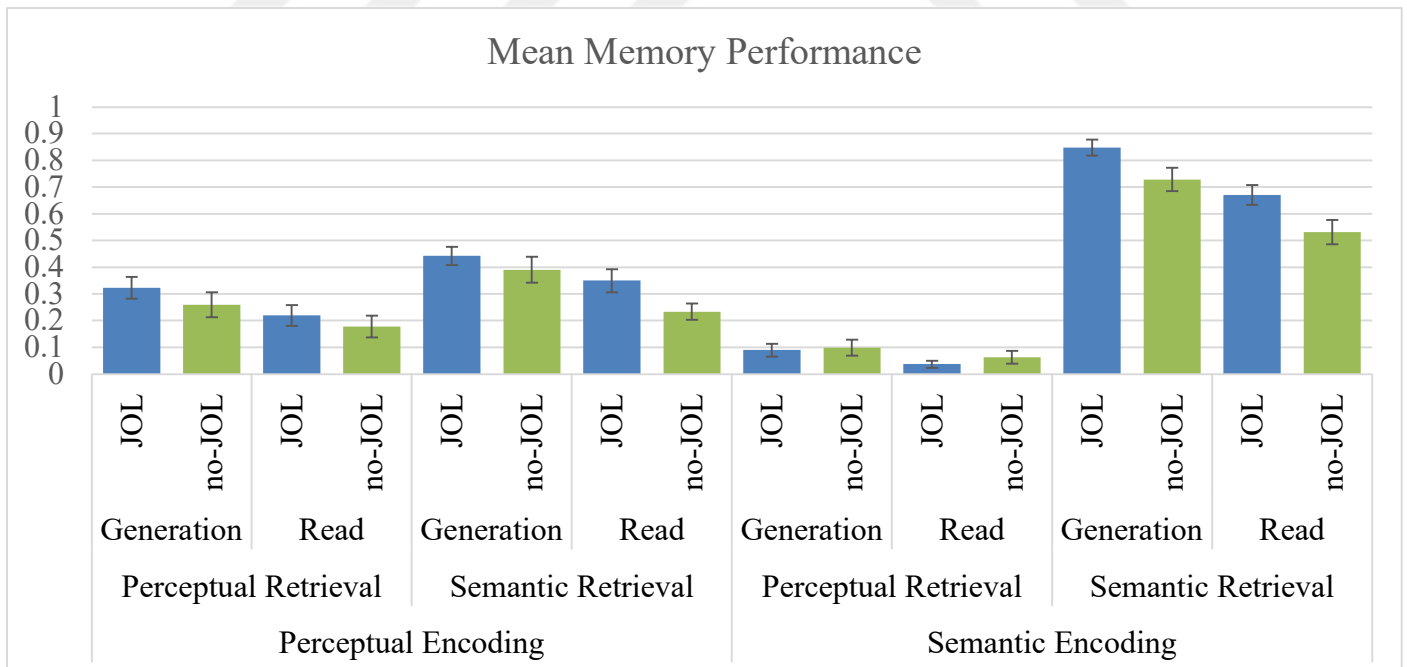


Figure 9. Experiment 3 Memory Performance

4.2.3 Memory Performance

A 2 (LOP at encoding: perceptual, semantic) x 2 (Generation: generation, read) x 2 (LOP at retrieval: perceptual, semantic) x 2 (judgment group: JOL, no-JOL) repeated measures ANOVA was conducted in which the judgment group was assessed as the between-subjects (see Figure 7). The main effects of LOP at encoding, $F(1,62) = 25.28, p < .001, \eta^2_p = 0.29$ (semantic processing > perceptual processing, $t = 5.03, p < .001, d = 0.4$), Generation, $F(1,62) = 58.45, p < .001, \eta^2_p = 0.4$ (generated words > read words, $t = 6.48, p < .001, d = 0.54$) and LOP at retrieval, $F(1,62) = 362.12, p < .001, \eta^2_p = 0.85$ (semantic cues > perceptual cues, $t = 19.03, p < .001, d = 1.74$) were significant.

Although LOP at Encoding*Generation was insignificant, a planned comparison was conducted to observe the memory benefits for semantic and perceptual generation tasks. Similar to Experiment 2, semantic generation led to higher memory performance than perceptual generation task, $t = 16.94, p < .001, d = 1.95$.

There was an interaction effect between LOP at encoding and LOP at retrieval, $F(1,62) = 190.49, p < .001, \eta^2_p = 0.75$. The post hoc test revealed that semantic processing led to higher memory performance than perceptual, $t = 17.67, p < .001, d = 2.14$ during a matching cue recall. Contrarily, perceptual processing displayed a higher memory performance than semantic processing during the no-match cue recall condition, $t = 11.08, p < .001, d = 1.34$. Additionally, there was a three-way interaction: LOP at encoding*LOP at retrieval*Generation, $F(1,62) = 4.06, p = 0.048, \eta^2_p = 0.06$. The post hoc test revealed a memory benefit for the semantic

generation task during a semantic cue recall, $t = 6.25, p < .001, d = 0.89$ and a memory benefit for the perceptual generation task during a semantic cue recall, $t = 4.17, p = 0.001, d = 0.59$. The results were marginally significant for the perceptual generation task during a perceptual cue recall, $t = 3.12, p = 0.056, d = 0.45$ and insignificant for the semantic generation task during a perceptual cue recall, $t = 1.47, p = 1.00$.

The between-subject effects were analyzed to assess the metamemory reactivity. The results displayed a significant memory benefit for the JOL group compared to no-JOL group, $F(1, 62) = 4.44, p = 0.039, \eta^2 p = 0.067$. There was only one interaction: LOP at Retrieval and Group, $F(1, 62) = 5.44, p = 0.023, \eta^2 p = 0.081$.

A post hoc test revealed that perceptual cue recall did not differ between groups, $t = 0.48, p = 1.00$. However, the JOL group displayed a higher memory performance than the no-JOL group during a semantic cue recall, $t = 3.04, p = 0.018, d = 0.51$.

Metamemory reactivity is boosted when there is a semantic cue recall, and there is no difference during a perceptual cue recall.

4.3 Discussion

First, it should be noted that response latency results from Experiment 1 and Experiment 2 were replicated in the last experiment. Also, Experiment 2 memory results were similar to Experiment – 3 (matching) condition. Out of interest, semantic generation and perceptual generation tasks were compared, and the previous results of Experiments 1 and 2 were replicated. Unlike Bertsch et al. (2007), semantic generation tasks provided better memory benefits than perceptual generation tasks.

Although the overall memory results displayed a similar pattern with TAP by showing a higher memory benefit for matching cues than no-matching ones, the subgroup results did not support a TAP perspective. If participants were aligned with a TAP perspective, they should always recall more words when the encoding and retrieval cues matched. Although the encoding phase consisting of semantic associations led to higher memory performance when presented with a matching (semantic) cue, there was a reverse pattern for perceptual processing. The number of words recalled from perceptually encoded words was higher when participants were solicited to recall through a no-matching (semantic) cue.

Our main aim for Experiment 3 was to understand the importance of matching cues between encoding and retrieval for the reactivity effect. In our predictions, we stated two plausible perspectives. If the results display an interaction between JOL presence and matching, this will indicate the cue-strengthening hypothesis. In the opposite direction, no observable interaction between JOL-presence and matching will defer cue-strengthening theory. From the TAP perspective on JOL reactivity predictions, the literature states that relatedness is a key factor in higher reactivity benefits. That said, the relatedness between encoding and retrieval phases could indicate a higher reactivity benefit. The current results find no interaction between JOL-presence and matching. These predictions fell short. TAP could not account for the expected size and direction of the reactivity effect.

However, the lack of interaction did not correspond to the overall memory benefit for semantic encoding processing over perceptual one. The interaction between LOP at

Retrieval and JOL presence indicated a reactivity effect depending on the retrieval type. The JOL group had no memory benefit during a perceptual cue recall. However, semantic cue recall led to a higher memory performance for the JOL group than the no-JOL group. The JOL reactivity was boosted during the retrieval phase of semantically meaningful testing. The current research found that JOL reactivity depends not only on the encoding process but also on the retrieval phase.

Last but not least, the metamemory judgments replicated the Experiment-2 results. As was said before, the metamemory results were unorthodox compared to the literature. Participants did not follow the fluency effect perspective, and their predictions were in the same trend as their actual memory performance.

CHAPTER 5

GENERAL DISCUSSION

The current research aimed to investigate humans' learning abilities, their self-awareness, and the consequences of self-awareness on memory performances.

Throughout the discussion, the results of the experiments will be discussed within the conceptual frameworks of the generation effect, metamemory performance, and metamemory reactivity.

5.1 Generation Effect

Although the generation effect has been investigated over 40 years, understanding the magnitude of different sets of cues was not one of the main aims in the literature. A mixed-generation tasks design was established for the current research, including semantic and perceptual generation tasks. In a meta-analysis, Bertsch et al. (2007) found a memory superiority for perceptual generation tasks over semantic generation tasks. Across experiments, perceptual generation tasks were operationalized as rhyming tasks, and the semantic generation ones were association tasks (including synonyms and antonyms). Although meta-analysis favored perceptual generation tasks, the levels of processing perspective (Craik & Lockhart, 1972) acknowledged the importance of semantic attributes for a higher memory performance compared to

perceptual attributes. Within the duality of predictions, the current research prioritized comparing the two different types of tasks within the same list. Our study revealed that semantic generation tasks were better for memory performance than perceptual generation tasks. Furthermore, there was a different test type in each experiment. In Experiment 1, there was a recognition test, and Experiment 2 applied a cued-recall retrieval phase; lastly, Experiment 3 used another cued-recall test with a matching manipulation. Nonetheless, the direction was not altered through the application of a different test type, and the results were consistent across three experiments: semantic generation task produces higher memory benefits than perceptual generation tasks.

There were three underlying mechanisms and possible explanations for the generation effect. The first was the mental effort theory (Tyler et al., 1979; Griffith, 1976; Jacoby et al., 1979; McFarland et al., 1980; Kirschner & Kirschner, 2012; Shenhav et al., 2017; Bonner & Sprinkle, 2002), which states that there will be a higher possibility of recalling an item if the item requires a higher mental effort. The difference between a high-effort and a low-effort task is not always a simple operation. As measured in current research, the mental effort theory proposes to use the response latency component. When participants had a longer response latency, this automatically indicates a requirement for a higher mental effort, leading to a higher memory performance for that stimulus. Across all experiments, the response latencies were measured. The current research showed that participants had a longer response latency when typing the generation tasks than when reading them. Also, they spent longer before typing the first letter for the semantic generation than the perceptual

generation. However, they did not display any difference between semantic reading and perceptual reading tasks.

For Experiment 1, using a recognition test, the mental effort theory was valid. Similar to response latency results, the memory results displayed a higher memory benefit for generation over the reading condition and semantic generation over perceptual generation, and there was no difference between semantic and perceptual reading tasks.

However, the mental effort theory falls short of understanding Experiments 2 and 3. The latter experiments applied a cued-recall retrieval phase. Although the response latency trends were similar to Experiment 1, the memory results were not in the same direction. In Experiment 2, semantic reading tasks led to higher memory performance than perceptual generation and perceptual reading tasks. Moreover, the semantic reading tasks also outperformed the perceptual reading tasks, and there was no difference between semantic reading tasks and perceptual generation tasks. The difference between the expected direction through mental effort theory and the actual memory results led to the consideration of other possible explanations and underlying mechanisms for the generation effect.

The two-factor or multi-factor theory (Hirshman & Bjork, 1988; McDaniel et al., 1990; McCurdy et al., 2020) states the importance of both item-specific and cue-target (i.e., intraitem) relationships for understanding the underlying mechanisms of the generation effect. One of the ways to test the multifactor theory is to apply different types of retrieval phases. If there are different factors affecting the outcome of the

generation effect, then there should be different directions after different retrieval phases. Similarly, the generation effect is observable through recognition and different recall phases, but using pairs in current research fosters the undeniable cue-target relationship. Furthermore, the alteration of the recognition test of Experiment 1 with cued-recall tests in the remaining Experiment 2 and 3 gives insight into the cue-target relationship. Similar to the meta-analysis by McCurdy (2020), the generation effect was observed through recognition and cued-recall. However, the subgroup effects differed from response latency predictions through cued-recall, indicating the importance of the cue-target relationship and multi-factor theory.

Experiment 2 displayed the generation effect as the main effect, and semantic generation still produced better memory benefits than perceptual generation. However, the results also showed no generation effect for perceptual tasks. In other words, the same encoding procedure applied in Experiment 1 did not result in a memory benefit for Experiment 2, solely indicating the importance of the test type, which was the cued-recall retrieval phase.

In Experiment 3, the manipulation of matching and no-matching cued recall altered the memory performance again. The generation effect was still a significant memory benefit as a main effect and semantic generation led to higher memory performance than the perceptual generation. Interestingly, the insignificant results of perceptual generation tasks during Experiment 2 were not replicated for Experiment 3 when participants saw a semantic cue to recall the perceptually encoded item. In other words, perceptual generation tasks were a significant memory benefit during

encoding; however, there should be a semantically meaningful cue to succeed in the recalling process.

The investigation of the generation effect raises counterfactual arguments against the Bertsch (2007) meta-analysis by providing consistent results of higher memory performance through semantic generation tasks compared to perceptual generation tasks. Also, the memory results indicated that mental effort theory was only valid when there was a recognition test; with the application of the cued-recall, the results can be understood through the multi-factor theory.

In the current research, we have created a new material set. To our knowledge, no study used Turkish words to test the generation effect. Across three experiments, our materials consisting of word pairs revealed the generation effect. This means that the materials can be used for further research.

5.2 Metamemory

The current research involved metamemory tasks for Experiments 2 and 3. In both experiments, participants provided a JOL score for their likelihood of recalling the item after exposure to it. In other words, there was an immediate JOL procedure to assess the metamemory performance. Most of the literature on the generation effect investigating the metamemory judgments (Besken & Mulligan, 2014; Yang et al., 2017; Froger et al., 2011) supported the fluency hypothesis: In most research, more fluent, easy-to-process, easy-to-encode items lead to higher metamemory judgments than less fluent, hard-to process items. As the generation effect typically takes longer and requires more mental effort, the fluency hypothesis should predict that the less

fluent generated items should lead to lower judgments than the read items. However, the current research found that participants were aware of their memory performance by providing higher JOL scores for generated words over read ones. Therefore, the fluency effect perspective cannot explain the metamemory judgments.

From Koriat's cue-utilization perspective (1997), providing a JOL score depends on mnemonic cues, and there are two different ways by which a participant can produce JOLs: belief-based cues and experience-based cues. The former focuses on general knowledge and assumptions, whereas the latter refers to the ease of processing the stimuli. The fluency effect is centralized around the fluency cue operationalized as the subjective experience for the ease of learning. If participants gave higher JOL scores for a set of stimuli, the encoding of stimuli could be associated with being a more fluent process. On the other hand, the more difficult process of the stimuli would lead to a lower fluency, indicating a lower JOL score. Current research measured the first key response latency to assess the ease-of-processing component. If participants take longer before beginning to type the word, their processing of the stimuli is associated with a less fluent expression. A less fluent process would have led to a lower JOL score. However, the results did not support the fluency hypothesis.

Across both experiments, participants provided higher JOL scores for generated words over read ones. Like semantic processing, the generated words required a longer response latency than read ones, indicating a less fluent process. Despite a higher response latency, participants predicted that generation manipulation would produce higher memory performance. Thus, the metamemory results for the

generation effect displayed a different pattern than most of the findings in the literature, which routinely displays that participants produce higher JOL scores for intact than incomplete or fragmented stimuli (Besken & Mulligan, 2014; Besken, 2016; Yang et al., 2017). The current results point out a new perspective on the metamemory judgments of the generation effect falling under the belief-based approach. In other words, participants either had prior beliefs about the generation effect such that they ignored the fluency cue. Similarly, Castel (2013) found that participants gave higher JOL scores for items read aloud even though it is a less fluent process than silently reading them: production effect.

Moreover, a similar pattern was observed for the levels of processing manipulation. The metamemory scores of Experiments 2 and 3 showed that participants gave higher JOL scores for words processed semantically over perceptually processed ones. In both experiments, there was a longer response latency for semantic processes over perceptually processed, meaning that participants were faster in typing the required answer for the perceptual process. From a fluency perspective, participants should have provided a higher JOL score for a more fluent set of words; however, they predicted that semantic processing would be a better memory benefit for their upcoming retrieval phase. Although participants favored semantic processing over perceptual processing, and the results cannot be explained through the fluency effect perspective, there are examples of similar expectations in the literature. One of the reoccurring findings of the literature states that associative relatedness is a key cue for participants to assess their JOL scores (Mueller et al., 2013). From a relatedness

perspective, semantic processing fosters a stronger association between cue-target compared to perceptual processing. Therefore, participants used the stronger cue-target association cue to provide a higher JOL score. From a cue-utilization perspective, participants favored their beliefs over fluency cues.

Last but not least, participants used the levels of processing (semantic, perceptual) and generation (generated, read) cues additively. The literature (Besken & Mulligan, 2014; Mueller et al., 2013) provides information on separate studies of generation effect and levels of processing. However, no study has been conducted on metamemory judgments of semantic and perceptual generation. The current research displayed that participants could predict the benefits of semantic generation over perceptual generation. The pattern was similar for the intact words as well. The semantic read words produced higher JOL scores than perceptual ones. The results were completely different from the response latency results, indicating they were completely different from fluency effect predictions. One of the key issues is the combination of multiple cues. Apart from scoring semantic processing over perceptual and generated words over read ones, they gave higher scores for semantic generation over perceptual generation. This indicates that participants were aware of different cues and accurately assessed their predicted memory performance by combining them strategically.

The current results showed that the generation effect and levels of processing manipulations did not follow the fluency effect perspective. In both cases, participants successfully predicted their memory performance by providing a similar trend

between predicted and actual memory performance. Particularly, the generation effect results in metamemory judgments, one of the unique examples indicating a belief factor on the higher memory performance for the generated words.

5.3 Metamemory Reactivity

Metamemory reactivity is the alteration of memory performance due to providing JOL scores. In the current research, Experiments 2 and 3 used an immediate JOL procedure for half of the participants. In contrast, the other half were instructed to type random numbers and used as a control group. The experiments revealed an overall higher memory performance for the JOL group. Supporting the meta-analysis (Double et al., 2018), the current material of related word pairs resulted in a positive reactivity effect from JOL-presence. The literature discusses two important perspectives, which should be discussed to elaborate on the results.

The cue-strengthening hypothesis (Soderstrom et al., 2015) states that providing a JOL will increase the association of intrinsic cues (i.e., relatedness), resulting in higher memory performance. In other words, JOL acts as a help tool for the material that could benefit from it. On the other hand, the changed goal hypothesis (Mitchum et al., 2016) has different predictions for metamemory reactivity. After participants realize that they must remember the materials, they actively choose to learn the most memorable stimulus at the expense of the harder-to-remember stimulus (i.e., unrelated). In doing so, participants do not differ significantly on related materials but fail to recall the unrelated materials. As a result, JOL acts as an overall memory failure.

The results of Experiment 2 supported the cue-strengthening hypothesis by revealing a positive JOL reactivity as a main effect. Additionally, the memory performance for semantic processing on generation and reading in both groups displayed a similar replication to the original work from Soderstrom (2015).

On the other hand, participants could only benefit from JOL's presence in the semantic read-word group. The rest of the conditions did not provide any significant memory reactivity. We also expected to observe a reactivity effect for the perceptual read-word group, and there were two reasons for our expectation: the shallowest process will have a larger benefit (Tekin & Roediger, 2020), and the perceptual relatedness should have benefited from providing a JOL according to the cue-strengthening hypothesis (Soderstorm et al., 2015).

In their study (Tekin & Roediger, 2020) involving three different levels of processing, participants provided JOLs after encoding the stimuli. The results showed that stimuli operationalized as the shallowest process benefitted more from providing JOLs than the deeper and deepest processes. However, our results did not follow the same pattern, and the shallowest process could not reach the biggest reactivity effect. One of the underlying reasons may be the retrieval phase difference. Compared to the cued-recall test of the current research, they conducted a recognition test.

It should be noted that the relatedness factor was previously discussed several times in the literature (Soderstorm et al., 2015; Mitchum et al., 2016; Double et al., 2018); however, researchers have only examined relatedness in terms of semantic associations to date. However, the operationalization of the relatedness was not

limited to semantic associations in the current research. We have used word pairs that could either have semantic or perceptual associations. Similar to the cue-strengthening theory, our semantic association revealed a metamemory reactivity; however, the new use of perceptual association did not confirm the predictions of the cue-strengthening theory.

As discussed in Experiment 2, semantic processing may benefit more from a cued-recall test. Therefore, we altered the cues in the retrieval phase of Experiment 3. There were two main reasons for revising the cued-recall test.

First of all, metamemory reactivity is heavily discussed through encoding processes. concerning the relatedness factor amongst materials, self-paced or experiment-paced study time, and likeliness for item-specific tasks. Few recent studies found that the positive JOL reactivity premise of the cue-strengthening hypothesis can be achieved with a criterion test. In other words, the metamemory reactivity is not just a variable of encoding; the retrieval phase also plays a crucial role. By altering the test type, we aimed to understand the reactivity effect's underlying mechanisms, size, and directions.

Secondly, the criterion test notion has previously been discussed as a factor of recognition, cued-recall, or free-recall. In particular, cued-recall is considered to be the most sensitive test type for observing a reactivity effect for word pairs. The results of Experiment 2 could only partially support the cue-strengthening hypothesis since the perceptual processing did not lead to any metamemory reactivity even with the same cues available. Although a similar manipulation was applied previously (Rivers

et al., 2023) aiming to understand the metamemory reactivity through levels of processing and transfer appropriate processing, the current manipulation was unique. In their study, participants saw category word pairs and letter pairs following either (same) cued-recall or letter-cued-recall tests. Through current research, we had the opportunity to discover the semantic cued recall of perceptually encoded-word pairs; thus, Experiment 3 aimed to display further definitions for the criterion test.

The main effect of the positive JOL reactivity from Experiment 2 was replicated; providing a JOL score as a metamemory measurement task led to higher memory performance. Compared to Experiment 2, the main effect of metamemory reactivity was decreased but not diminished. Notably, the only positive reactivity for the semantic read word group did not display any benefit of providing a JOL in Experiment 3. The results supported the idea that the retrieval phase is crucial in understanding the underlying mechanism of JOL reactivity. Also, it is critical to observe a change in metamemory reactivity without applying a new retrieval form, such as recognition or free recall. The effect of the retrieval phase changed drastically when Experiment 3 only used a variation of the same recall procedure.

The cue-match and cue-no-match recalls during the retrieval phase expand our understanding of the boundary conditions of the cue-strengthening theory. We aimed to test the importance of the retrieval phase by acknowledging the cue-strengthening theory and Transfer Appropriate Processing (Morris et al., 1977). If the JOL reactivity is more sensitive to the similarity between encoding and retrieval phases, it should produce the highest memory performance increase for matching cues. The results did

not display an interaction between matching and JOL presence. Participants in the JOL group did not benefit from encountering the matching cues any different than participants in the no-JOL group.

Most interestingly, the interaction between JOL-presence and retrieval type revealed that the criterion test matters. The results showed no group difference for the retrieval test using perceptual cues. Contrarily, participants in the JOL group had a significantly higher memory performance than the no-JOL group when the retrieval test used semantic cues.

The results of Experiment 3 make two novel contributions regarding the cue-strengthening theory (Soderstrom et al., 2015): criterion test and relatedness.

First of all, current research redefined the criterion test. The cue strengthening theory states that the increased similarity between the encoding and retrieval phases will be a better measurement to observe if there is a reactivity effect. We found that there is indeed a criterion test to observe the reactivity effect; however, it is not the similarity between encoding and retrieval but rather a semantically meaningful testing. As stated above, there was no reactivity effect for perceptual processing in Experiment 2.

Following the cue-strengthening theory, there should have been a reactivity effect for perceptually encoded associations since the perceptual associations are still a relatedness factor.

We could speculate on the lack of reactivity effect for the perceptual associations:

First, the perceptual associations are not as strong as semantic associations and, thus, will not show similar trends. In other words, observing a reactivity effect is

impossible unless there is a semantic association between cue and target. Second, the retrieval tests that have been used in previous research may not have allowed us to see whether there is a reactivity effect for perceptual associations since the cued-recall testing could be more beneficial only for semantic associations. We tested the second possibility in Experiment 3; participants recalled the perceptually encoded associations from either a perceptual or semantic cue, and the same method was applied to semantically encoded associations. By revealing an interaction between JOL presence and retrieval type, Experiment 3 found that the retrieval phase needs to be a semantically meaningful test to observe a reactivity effect. Thus, the lack of reactivity effect for perceptual processing in Experiment 2 may be due to perceptual testing. The proposed criterion test was not just a similarity between the encoding and retrieval phase but rather applying semantically meaningful testing.

Secondly, the most crucial drawback of studying only semantic association should be revisited with a further point on the relatedness factor. In the current study, we used perceptual and semantic associations. Supporting the cue-strengthening hypothesis, all related material of the current research led to metamemory reactivity in Experiments 2 and 3. However, the subgroups were only significant for the semantic read condition in Experiment 2. Even though perceptual associations are not as powerful as semantic ones, they are still associations and should have led to similar expectations, e.g., *a smaller reactivity effect*. The null reactivity findings of perceptual processing indicate that the relatedness of the word pairs is equal to semantic associations. Since the metamemory literature of word pairs aimed at the difference between related and

unrelated material and the relatedness is equaled to semantic associations, there is a lack of research to understand the possible directions for another type of relatedness.

Our study revealed that providing a JOL score will boost the actual memory performance for the related material. Along with semantic association, the current research is the first to operationalize relatedness as a perceptual association. For semantic association, we found supporting results for the cue-strengthening hypothesis; there was a significant memory benefit for the semantic read-words. However, perceptual association did not follow the same pattern and failed to display a subgroup reactivity effect. In the last experiment, we found that the test type matters in observing if there is a reactivity effect. However, the key to the criterion test was not the similarity between the encoding and retrieval phases. The results showed that the JOL group benefits more than the no-JOL group from semantic testing, regardless of the encoding type.

5.4 Limitations and Further Studies

We did not conduct a follow-up study for Experiment 1. The metamemory reactivity can also be investigated through a recognition test. Experiments 2 and 3 did not find a significant reactivity effect for perceptually encoded stimuli; it should be revisited with a recognition test to understand the effect of test type on the reactivity mechanisms.

Experiments 2 and 3 found that participants assessed the generation effect as a good memory benefit for their actual memory performances. The current results should

encourage further research to consider the prior beliefs and knowledge over fluency cues.

As a modification for Experiment 3, future studies can focus on a between-subject test manipulation. Due to the limits of finding participants, we aimed to understand the importance of test type as a within-subject manipulation and still found an interaction effect. However, semantic recalling of the perceptually encoding pairs between groups was not significantly beneficial. A more powerful study can explain the benefit of providing a JOL for perceptual associations during a between-subject test manipulation.

Last but not least, through current results, metamemory reactivity literature can focus on defining relatedness and the criterion test. It is crucial to understand the positioning of different association types on relatedness and how to observe the reactivity effect with regard to the retrieval phase.

5.5 Conclusion

Across three experiments, current research aimed to understand humans' learning abilities and memory performances. We found that the generation effect is a valid manipulation of Turkish words. Additionally, we conducted one unique study design to compare the semantic and perceptual generation tasks. We found that the semantic generation task has a superior memory benefit over the perceptual generation task. Then, we search for the self-awareness of humans for their own memory performances prospectively. We found that participants are aware of the memory benefits of the generation effect and levels of processing. Interestingly, the current

study is one of the few studies that have found an awareness of the generation effect. Later on, we found that the sole act of predicting the recalling of a stimulus in a later memory test improved the actual memory performance. Metamemory researchers should consider the reactivity effect for future studies, particularly with related materials. Lastly, we sought to understand the metamemory reactivity by manipulating the retrieval phase. The results showed that semantically meaningful testing is the more sensitive way of observing metamemory reactivity.

REFERENCES

- Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 553.
- Begg, I., Vinski, E., Frankovich, L., & Holgate, B. (1991). Generating makes words memorable, but so does effective reading. *Memory & cognition*, 19(5), 487-497.
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & cognition*, 35, 201-210.
- Besken, M. (2016). Picture-perfect is not perfect for metamemory: Testing the perceptual fluency hypothesis with degraded images. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1417.
- Besken, M., & Mulligan, N. W. (2014). Perceptual fluency, auditory generation, and metamemory: analyzing the perceptual fluency hypothesis in the auditory modality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(2), 429.
- Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, 64, 417-444.

- Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, organizations and society*, 27(4-5), 303-345.
- Brennan, I. (2008). Brand placement in novels: A test of the generation effect. *International Journal of Advertising*, 27(4), 495-509.
- Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition*, 41, 28-35.
- Craik, F. I. (2002). Levels of processing: Past, present... and future?. *Memory*, 10(5-6), 305-318.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.
- Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology*, 10, 494909.
- Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, 26(6), 741-750.
- Dougherty, M. R., Robey, A. M., & Buttaccio, D. (2018). Do metacognitive judgments alter memory performance beyond the benefits of retrieval practice? A comment on and replication attempt of Dougherty, Scheck, Nelson, and Narens (2005). *Memory & cognition*, 46, 558-565.

- Fiedler, K., Ackerman, R., & Scarampi, C. (2019). Metacognition: Monitoring and controlling one's own knowledge, reasoning and decisions. *The psychology of human thought: An introduction*, 89-111.
- Foley, M. A., & Foley, H. J. (2007). Source-monitoring judgments about anagrams and their solutions: Evidence for the role of cognitive operations information in memory. *Memory & cognition*, 35(2), 211-221.
- Froger, C., Sacher, M., Gaudouen, M. S., Isingrini, M., & Taconnat, L. (2011). Metamemory judgments and study time allocation in young and older adults: Dissociative effects of a generation task. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie exp rimentale*, 65(4), 269.
- Glisky, E. L., & Rabinowitz, J. C. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(2), 193.
- G z,  . (2020). *Yazılı T rk enin kelime sıklığı s zl   *. T rk Dil Kurumu.
- Griffith, D. (1976). The attentional demands of mnemonic control processes. *Memory & Cognition*, 4, 103-108.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 22.

- Hirshman, E., & Bjork, R. A. (1988). The generation effect: Support for a two-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3), 484.
- Hunt, R. R. (2013). Precision in memory through distinctive processing. *Current Directions in Psychological Science*, 22(1), 10-15.
- Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of verbal learning and verbal behavior*, 17(6), 649-667.
- Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both?. *Psychonomic bulletin & review*, 25, 2356-2364.
- Johnson, M. M., Schmitt, F. A., & Pietrukowicz, M. (1989). The memory advantages of the generation effect: Age and process differences. *Journal of Gerontology*, 44(3), P91-P94.
- Kinjo, H., & Snodgrass, J. G. (2000). Does the generation effect occur for pictures?. *The American journal of psychology*, 113(1), 95.
- Kirschner, P. A., & Kirschner, F. (2012). Mental effort. In *Encyclopedia of the Sciences of Learning* (pp. 2182-2184). Springer.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of experimental psychology: General*, 126(4), 349.

- Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: a comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1133.
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of memory and Language*, 52(4), 478-492.
- Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 989.
- Lee, A. Y., & Labroo, A. A. (2004). The effect of conceptual and perceptual fluency on brand evaluation. *Journal of Marketing Research*, 41(2), 151-165.
- Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, 22(1), 16-26.
- Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, 7, 175-195.
- Marsh, E. J., Edelman, G., & Bower, G. H. (2001). Demonstrations of a generation effect in context memory. *Memory & cognition*, 29(6), 798-805.

- McCurdy, M. P., Viechtbauer, W., Sklenar, A. M., Frankenstein, A. N., & Leshikar, E. D. (2020). Theories of the generation effect and the impact of generation constraint: A meta-analytic review. *Psychonomic Bulletin & Review*, 27, 1139-1165.
- McDaniel, M. A., Riegler, G. L., & Waddill, P. J. (1990). Generation effects in free recall: Further support for a three-factor theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 789.
- McFarland Jr, C. E., Frey, T. J., & Rhodes, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior*, 19(2), 210-225.
- Meyers-Levy, J. (1991). Elaborating on elaboration: The distinction between relational and item-specific elaboration. *Journal of Consumer Research*, 18(3), 358-367.
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145(2), 200.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior*, 16(5), 519-533.
- Mulligan, N. W. (2002). The generation effect: Dissociating enhanced item memory and disrupted order memory. *Memory & Cognition*, 30(6), 850-861.

- Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, 48, 745-758.
- Nairne, J. S., & Widner, R. L. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1), 164.
- Nelson, T. O. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (Vol. 26, pp. 125-173). Academic Press.
- Nieznański, M. (2011). Generation difficulty and memory for source. *Quarterly Journal of Experimental Psychology*, 64(8), 1593-1608.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in cognitive sciences*, 12(6), 237-241.
- Rabinowitz, J. C., & Craik, F. I. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language*, 25(2), 226-237.
- Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. *The Oxford handbook of metamemory*, 1, 65-80.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of experimental psychology: General*, 137(4), 615.

- Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory*, 29(10), 1342-1353.
- Rivers, M. L., Janes, J. L., Dunlosky, J., Witherby, A. E., & Tauber, S. K. (2023). Exploring the role of attentional reorienting in the reactive effects of judgments of learning on memory performance. *Journal of Intelligence*, 11(8), 164.
- Sengupta, J., & Gorn, G. J. (2002). Absence makes the mind grow sharper: Effects of element omission on subsequent recall. *Journal of Marketing Research*, 39(2), 186-201.
- Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1258.
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual review of neuroscience*, 40, 99-124.
- Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and Memory*, 4(6), 592.

- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 553.
- Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition*, 33, 1116-1129.
- Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, 95, 116-123.
- Sutherland, J. A., Krug, D., & Glover, J. A. (1988). The selective displaced rehearsal hypothesis and failure to obtain the generation effect. *Bulletin of the Psychonomic Society*, 26, 413-415.
- Tauber, S. K. U., Dunlosky, J., & Rawson, K. A. (2015). The influence of retrieval practice versus delayed judgments of learning on memory: Resolving a memory-metamemory paradox. *Experimental psychology*, 62(4), 254.
- Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgements of retention (JORs). *Quarterly Journal of Experimental Psychology*, 65(7), 1376-1396.
- Tekcan, A. İ., & Göz, İ. (2005). *Türkçe kelime normları: 600 kelimenin imgelem, somutluk, sıklık değerleri ve çağrışım setleri* (No. 849). Boğaziçi Üniversitesi.
- Tekin, E., & Roediger III, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie*.

- Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, 5(6), 607.
- Undorf, M., & Bröder, A. (2020). Cue integration in metamemory judgements is strategic. *Quarterly Journal of Experimental Psychology*, 73(4), 629-642.
- Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1264.
- Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language*, 92, 293-304.
- Van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & cognition*, 29(8), 1081-1087.
- Veenman, M. V., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and learning*, 1, 3-14.
- Weber, E., Woods, S. P., Kellogg, E., Grant, I., Basso, M. R., & HIV Neurobehavioral Research Program (HNRP) Group. (2012). Self-generation enhances verbal recall in individuals infected with HIV. *Journal of the International Neuropsychological Society*, 18(1), 128-133.

- Winstanley, P. A. D. (1996). Generation effects and the lack thereof: The role of transfer-appropriate processing. *Memory*, 4(1), 31-48.
- Yang, C., Sun, B., & Shanks, D. R. (2018). The anchoring effect in metamemory monitoring. *Memory & Cognition*, 46, 384-397.
- Yang, H., Cai, Y., Liu, Q., Wang, Q., Chen, C., & Xue, G. (2015). Differential neural correlates underlie judgment of learning and subsequent memory performance. *Frontiers in Psychology*, 6, 147448.

APPENDICES

WORD LIST

Semantic Cue	Target	Perceptual Cue
Çabuk	Acele	İskele
Uyum	Ahenk	Kepenک
Cihaz	Alet	Devlet
Blok	Apartman	Orman
Arsa	Arazi	Terazi
Elma	Armut	Somut
Hukuk	Avukat	Sakat
Anne	Baba	Araba
Özgür	Bağımsız	Rahatsız
Dükkan	Bakkal	Sakal
Su	Bardak	Dudak
Patik	Bebek	Göbek
Yakıt	Benzin	Hazin
Bilek	Bilezik	Müzik
Hamur	Börek	Dikmek
Uzay	Boşluk	Yorgunluk
Karınca	Çalışkan	İmkan
Çömlek	Çanak	Tırnak
Kaşık	Çatal	Dijital
Bahar	Çiçek	Gerçek
Evlat	Çocuk	Sucuk

Metal	Demir	Tamir
Bilim	Deney	Güney
Terazi	Denge	Yenge
Felaket	Deprem	Krem
Gazete	Dergi	Sergi
Davet	Düğün	Öğün
Sigara	Duman	Liman
Giyim	Elbise	Lise
Kadın	Erkek	Ürkek
Yeni	Eski	Viski
Slogan	Eylem	Düzlem
Etken	Faktör	Monitör
Tuhaf	Garip	Tahrip
İzlenim	Gözlem	İşlem
Çirkin	Güzel	Dizel
Özlem	Hasret	Ücret
Duygulu	Hassas	Esas
Ütopya	Hayal	Sosyal
Sürpriz	Hediye	Saniye
Ot	İnek	Dernek
Bina	İnşaat	Kanaat
Dikiş	İplik	İyilik
Nehir	İrmak	Koşmak
Elektrik	İşık	Kaşık
Kağıt	Kalem	Özlem
Kelebek	Kanat	Teminat

Gemi	Kaptan	Vatan
Cinayet	Katil	Tekstil
Hüzün	Keder	Minder
İskelet	Kemik	Dinamik
Bıçak	Keskin	Çirkin
Gurur	Kibir	Hiçbir
Bilgi	Kitap	Etap
Espri	Komik	Seramik
Misafir	Konuk	Donuk
Hayvan	Köpek	İpek
Bela	Kötülük	Gözlük
Çubuk	Kraker	Spiker
Islak	Kuru	Kanguru
Öncülük	Liderlik	Delik
Cisim	Madde	Cadde
Çamaşır	Makine	Hazine
Balkon	Manzara	Yara
Hoşnut	Memnun	Kanun
Ülser	Mide	Pide
Vatan	Millet	Tablet
Örnek	Misal	Kutsal
Tabu	Oyun	Boyun
Ceket	Pantolon	Salon
Borçlu	Para	Mağara
Bölüm	Parça	Fırça
İşaret	Parmak	Çakmak

Zeytin	Peynir	Sinir
Pilav	Pirinç	Sevinç
Torba	Poşet	Manşet
Konfor	Rahat	Nasihat
Tanıtım	Reklam	Sağlam
Çerçeve	Resim	Cisim
Kültür	Sanat	Kainat
Uğur	Şans	Dans
Standart	Sapma	Yapma
Kadeh	Şarap	Çorap
Altın	Sarı	Uyarı
Asker	Savaş	Yavaş
Bayram	Şeker	Teker
Başıboş	Serseri	Beceri
Dert	Sıkıntı	Alıntı
Tabanca	Silah	Külâh
Kuşku	Şüphe	Cephe
Otel	Tatil	Stil
Çikolata	Tatlı	Kayıtlı
Kapak	Tencere	Dere
Erozyon	Toprak	Durak
İshal	Tuvalet	Bilet
Pilot	Uçak	Kaçak
Kafiye	Uyak	Dayak
Bulut	Yağmur	Memur
Kural	Yasa	Masa

Nefes	Yaşam	Akşam
Beceri	Yetenek	Örnek
Havuz	Yüzme	Çizme

