



Faculty of Cognitive Sciences and Human Development

**The Effects of Retrieval Practice Across Levels of Thinking and
Retention Interval on Reading Comprehension**

Corrin Alicia anak Nero

**Master of Science
2022**

The Effects of Retrieval Practice Across Levels of Thinking and Retention Interval on Reading Comprehension

Corrin Alicia anak Nero

A thesis submitted

In fulfillment of the requirements for the degree of Master of Science

(Cognitive Science)

Faculty of Cognitive Sciences and Human Development

UNIVERSITI MALAYSIA SARAWAK

2022

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.


.....

Signature

Name: Corrin Alicia anak Nero

Matric No.: 19020133

Faculty of Cognitive Sciences and Human Development

Universiti Malaysia Sarawak

Date : 15 September 2022

ACKNOWLEDGEMENT

Before anything else, all praise and thanks to God for blessing my journey in completing this work. I thank Him for giving me the strength during difficult times and wisdom, when I lacked it.

I would like to express my sincere appreciation to my supervisor, Dr Norehan Zulkiply for her time, effort, and guidance in the completion of this research. I thank her for the words of encouragement and for the invaluable advice she had given throughout this journey, in both academic research as well as life.

Special thanks to the two most important figures in my life—my beloved parents, Mr Nero Madi and Mdm Bibi Chia for their unwavering love and support. Needless to say, I would not be where I am today without them. My gratitude also extends to my sister, Daine Esther, for her continuous support.

My appreciation also goes out to my friends who have been there for me in my ups and downs. Thank you for continuously motivating, advising, and tirelessly listening to my rants (you know who you are).

To the dramas, artistes, and fandoms that I have got myself into, thank you for being a constant space of comfort especially during stressful times.

Finally, I would like to thank all the students for participating in the study. The completion of this research would not have been possible without their participation.

ABSTRACT

The present study examined the effect of different types of retrieval practice on reading comprehension across levels of thinking and retention interval in students. One hundred undergraduates divided into two retention interval groups (short and long retention interval) were asked to read a passage on a topic in Cognitive Psychology and were then required to engage in two types of retrieval practice learning strategy (production test and recognition test) and different levels of thinking (lower-order thinking and higher-order thinking). A three-way mixed ANOVA statistical test was used to analyse the data and found no significant difference in reading comprehension across the different types of retrieval practice, suggesting that the performance when using the recognition test was equivalent to when using the production test. The difference in reading comprehension between the different types of retention interval also was not observed, indicating that students in the short retention interval group retained just as much information as those in the long retention interval group. Additionally, the present study observed a significant difference in students' reading comprehension between different levels of thinking, signifying that the students' performance for the lower-order thinking questions was better as compared to the higher-order thinking questions. The present finding contributed to the existing body of knowledge in which it suggested that the performance in reading comprehension when using a recognition test, particularly a well-constructed one, with competitive alternatives was equivalent to when using a production test.

Keywords: Retrieval practice, levels of thinking, retention interval, reading comprehension, undergraduates

Kesan Latihan Perolehan Semula, Tahap Pemikiran dan Selang Pengekalan terhadap Pemahaman Membaca

ABSTRAK

Kajian ini mengkaji kesan latihan perolehan semula, tahap pemikiran dan selang pengekalan terhadap pemahaman membaca dalam kalangan pelajar prasiswazah. Seratus peserta dibahagikan kepada dua kumpulan (selang pengekalan pendek dan selang pengekalan panjang) dan mereka diminta untuk membaca petikan tentang satu topik dalam subjek Psikologi Kognitif. Peserta kajian kemudiannya dikehendaki untuk mengulangkaji petikan dengan menggunakan dua jenis strategi pembelajaran latihan perolehan semula (ujian penjaan jawapan secara bebas dan ujian objektif aneka pilihan) dan melibatkan tahap pemikiran yang berbeza (pemikiran aras rendah dan pemikiran aras tinggi). Ujian statistik ANOVA campuran tiga hala digunakan untuk menganalisis data dan didapati tiada perbezaan signifikan dalam pemahaman membaca antara pelbagai jenis latihan perolehan semula, sekaligus menunjukkan bahawa prestasi pemahaman membaca pelajar adalah sama apabila menggunakan ujian objektif aneka pilihan dan ujian penjaan jawapan secara bebas. Hasil analisis ujian statistik juga tidak menemui perbezaan signifikan dalam pemahaman membaca pelajar antara pelbagai jenis selang pengekalan sekaligus menunjukkan bahawa jumlah maklumat yang dikekalkan oleh pelajar adalah sama untuk selang pengekalan pendek dan panjang. Di samping itu, kajian ini menemui perbezaan yang signifikan dalam pemahaman membaca pelajar antara tahap pemikiran yang berbeza, menandakan bahawa prestasi pemahaman pelajar untuk soalan pemikiran aras rendah adalah lebih baik berbanding dengan soalan pemikiran aras tinggi. Penemuan dalam kajian ini menyumbang kepada literatur sedia ada di mana dapatan kajian ini mencadangkan bahawa prestasi dalam pemahaman membaca apabila menggunakan ujian objektif aneka

pilihan, terutamanya jika soalnya dibina dengan baik, dengan alternatif yang kompetitif adalah sama apabila menggunakan ujian penajaan jawapan secara bebas.

Kata kunci: *Latihan perolehan semula, tahap pemikiran, selang pengekal, pemahaman membaca, prasiswazah*

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
<i>ABSTRAK</i>	iv
TABLE OF CONTENTS	vi
LIST OF TABLES	x
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Background of Study	1
1.3 Problem Statement	8
1.4 Objectives	11
1.5 Research Questions	12
1.6 Definition of Terms	13
1.6.1 Retrieval Practice	13
1.6.2 Levels of Thinking	14
1.6.3 Retention Interval	14

1.6.4	Reading Comprehension	15
1.7	Conceptual Framework	15
1.8	Significance of Study	16
1.9	Chapter Summary	17
CHAPTER 2: LITERATURE REVIEW		19
2.1	Introduction	19
2.2	Introduction to Retrieval Practice	19
2.3	Explanation for Benefits of Retrieval Practice in Learning	21
2.4	Effects of Retrieval Practice Relative to Restudying or Rereading	26
2.5	Types of Retrieval Practice, Learning Materials and Study Topics	30
2.5.1	Production Test Versus Recognition Test	31
2.5.2	Retrieval Practice and Levels of Thinking	34
2.5.3	Retrieval Practice and Retention Interval	36
2.6	Chapter Summary	38
CHAPTER 3: METHOD		39
3.1	Introduction	39
3.2	Participants	39
3.3	Design	40
3.4	Research Materials	41
3.4.1	Study Phase	41

3.4.2	Retrieval Practice Phase	41
3.4.3	Final Test Phase	43
3.4.4	Distractor Task Phase	45
3.5	Data Collection Procedure	46
3.6	Data Analysis	47
3.7	Chapter Summary	48
CHAPTER 4:	RESULTS	49
4.1	Introduction	49
4.2	Demographic Analysis	49
4.3	Tests for Normality and Homogeneity	50
4.4	Findings From the Three-Way Mixed ANOVA on Final Test Performance	53
4.4.1	Students' Reading Comprehension Across Different Types of Retrieval Practice	55
4.4.2	Students' Reading Comprehension Between Lower-Order Thinking and Higher-Order Thinking	56
4.4.3	Students' Reading Comprehension Between Different Types of Retention Interval	56
4.4.4	Interaction Effect Between Types of Retrieval Practice and Levels of Thinking	56
4.4.5	Interaction Effect Between Types of Retrieval Practice and Retention Interval	57
4.4.6	Interaction Effect Between Retention Interval and Levels of Thinking	58
4.4.7	Interaction Effect Between Types of Retrieval Practice, Levels of Thinking and Retention Interval	58

4.5	Chapter Summary	59
CHAPTER 5: DISCUSSION AND CONCLUSION		61
5.1	Introduction	61
5.2	Effect of Different Types of Retrieval Practice on Students' Reading Comprehension	61
5.3	Effect of Levels of Thinking on Students' Reading Comprehension	64
5.4	Effect of Different Types of Retention Interval on Students' Reading Comprehension	65
5.5	Interaction Effect Between Types of Retrieval Practice and Levels of Thinking	66
5.6	Interaction Effect Between Types of Retrieval Practice and Retention Interval	67
5.7	Interaction Effect Between Retention Interval and Levels of Thinking	68
5.8	Interaction Effect Between Types of Retrieval Practice, Levels of Thinking and Retention Interval	69
5.9	Theoretical and Practical Implications	71
5.10	Limitations of the Present Research	72
5.11	Suggestions for Future Research	73
REFERENCES		75
APPENDICES		85

LIST OF TABLES

	Page
Table 4.1 Demographic Analysis	50
Table 4.2 Test for Homogeneity of Variance	53
Table 4.3 Three-Way Mixed ANOVA Results on Within-Subject Variables	53
Table 4.4 Three-Way Mixed ANOVA Results on Between-Subject Variable	55
Table 4.5 Means and Standard Deviations of Interaction Between Types of Retrieval Practice, Levels of Thinking and Retention Interval	59

LIST OF FIGURES

	Page
Figure 1.1 Conceptual Framework for the Present Study	16
Figure 3.1 Distribution of Questions in the Retrieval Practice Phase	43
Figure 3.2 Examples of Questions Used in the Retrieval Practice Phase and the Final Test Phase	45
Figure 4.1 Histograms for the Short Retention Interval Group	51
Figure 4.2 Histograms for the Long Retention Interval Group	52
Figure 5.1 An Example of the Recognition Test Question	64

LIST OF ABBREVIATIONS

HOT	Higher-Order Thinking
LOT	Lower-Order Thinking
MCQ	Multiple-Choice Question
MCQs	Multiple-Choice Questions

CHAPTER 1

INTRODUCTION

1.1 Introduction

Chapter One presents the general picture of this research. It describes the background of the study, highlights the problem statement, and defines the research objectives as well as research questions. This chapter also includes the definition of terms used, conceptual framework and the significance of the study.

1.2 Background of Study

Tests in educational setting are conventionally used as a tool in the assessment of learning (Amin & Malik, 2013; Brame & Biel, 2015; Moreira et al., 2019) as well as to provide a summarised picture of an individual's knowledge (Stenlund et al., 2016). However, a growing number of studies have revealed that tests can do more than just measuring one's knowledge, that is, it can boost learning as well as promotes long-term memory retention (Stenlund et al., 2016; Karpicke, 2017; Moreira et al., 2019). Studies that examined the effects of test usually compares the condition where learners take test or engage in retrieval practice task on previously studied materials to the condition where learners restudy or reread the previously studied materials. Experimental procedures in examining this effect usually involved the following phases: study phase, retrieval practice task phase and final test phase. Participants were firstly required to study a set of material, for example word pairs, word lists or reading passages and then instructed to take part in a retrieval practice task, either in the form of a free recall test, a multiple-choice question (MCQ) or a short-answer test. After engaging in the retrieval practice task, they have to complete a final test, either immediately or after a delay (e.g., hours, days, or weeks) to measure their performance

in terms of their understanding of the studied material or how much information they can recall from their memory.

Several studies have shown that when learners engage in the practice of retrieving previously studied information, it can result in greater long-term memory retention as compared to merely rereading or restudying that particular information (Roediger & Karpicke, 2006; Karpicke & Roediger, 2008; Stenlund et al., 2016; Koh et al., 2018). In both experiments conducted by Roediger and Karpicke (2006), students were required to study reading passages and then undertake one or three immediate free recall tests without feedback or restudy the material the same number of times as the students who took the tests. Students then took a final retention test five minutes, two days, or one week later. Restudying was included as one of the conditions based on the idea that the students were represented with the information once again. The results, particularly the second experiment which examined repeated testing versus repeated studying revealed that when the final test was given after five minutes, restudying improved recall compared to repeated testing. However, on the delayed tests, testing was revealed to have produced greater memory retention than studying, although repeated studying increased students' confidence in their ability to recall the learned material. Similarly, Stenlund et al.'s (2016) study which compared the effects of repeated practice testing with short-answer items on short- and long-term memory retention to repeated practice testing with multiple-choice questions, as well as rereading the material found that both test formats (short-answer and MCQ) enhanced short- and long-term memory relative to rereading. Rereading or restudying is usually used as the comparison due to the basic assumption that learning happens when learners study and encode material and based on such an assumption, rereading or restudying should increase learning (Karpicke & Roediger, 2008). However, such is not always the case as several studies revealed that using

test or retrieval practice is more effective for memory retention as it involves effortful retrieval, whereas studying or reading the material simply involves encoding (Roediger & Karpicke, 2006; Karpicke & Roediger, 2008).

When one studies and then proceeds to retrieve the previously learned information, it enhances learning and promotes long-term memory retention more than merely restudying the material. This phenomenon is known as the testing effect (McDaniel, Roediger, et al., 2007; Stenlund et al., 2016; Moreira et al., 2019). This phenomenon is also commonly referred to as the retrieval practice effect (Karpicke & Roediger, 2007a; Karpicke, 2017) or retrieval-based learning (Karpicke, 2017; Bae et al., 2019). The explanation behind the effectiveness of this effect is based on several theories. One of the theories is the retrieval effort hypothesis which states that a more difficult and more effortful processing has greater beneficial effects on memory (Gardiner et al., 1973; Pyc & Rawson, 2009). The act of taking a test requires the use of effortful retrieval in which learners have to actively retrieve information from memory (Karpicke & Roediger, 2008). When information is successfully retrieved from memory, the representation in memory is changed in a way that it becomes more likely to be recalled in the future (Bjork, 1975) thus, yielding greater benefit compared to repeatedly studying the materials (Roediger & Karpicke, 2006). Another theory that is commonly used in explaining the retrieval practice effect is transfer-appropriate processing. This theory states that when the cognitive process that a student engages in during the initial learning activity is identical with the processing required during the final assessment, it leads to a better performance in learning and memory tasks (Morris et al., 1977; Kolers & Roediger, 1984). This is based on the assumption that the positive results on long-term memory retention are due to the resemblance of cognitive processes during learning tasks and transfer tasks (Endres & Renkl, 2015), which further explains why test or retrieval

practice has an advantage over restudying or rereading. This advantage exists due to the cognitive processes that are present during the transfer tasks being more similar to the ones used during initial testing, rather than restudying or rereading (e.g., Thomas & McDaniel, 2007). Another theory to explain the retrieval practice effect is the elaborative-retrieval account. This theory explains that semantic elaboration takes place during the process of retrieval which will improve subsequent retrieval (Karpicke, 2017). The idea behind this theory is that when one retrieves a target from a given cue, it activates additional information that is semantically associated to the cue and this information is incorporated with the successfully recalled target to create an elaborate memory trace that becomes easily remembered at a later time (Carpenter & Yeung, 2017; Karpicke, 2017). When one restudies or rereads material on the other hand, the production of additional mediators does not necessarily happen because there is no activity of retrieving the target word during restudy trials (Carpenter, 2009) hence explaining why retrieval practice results in better performance than restudying. Across a wide number of past studies on the retrieval practice effect, consistent findings have reported that retrieval practice has greater beneficial effects than rereading or restudying (Carpenter et al., 2006; McDaniel, Roediger, et al., 2007; Rowland, 2014; Karpicke, 2017) thus indicating that it is an established fact that retrieval practice enhances memory retention relative to rereading or restudying.

It was noted in Greving and Richter's (2018) study that although there is convincing evidence pertaining to the beneficial effects of using retrieval practice, the circumstances in which the effects occur are still open for discussion. One such example is the generalisability of the results, especially in classroom contexts. Therefore, considering that there have been consistent findings on various contexts of studies which reported that retrieval practice had greater benefits than rereading, the present study aims to focus on the effects of the two

different types of retrieval practice format. There are two commonly used test or retrieval practice formats in examining testing effect which are the recognition test and the production test (Butler & Roediger, 2007; Pyc & Rawson, 2009; Larsen & Butler, 2013; Greving & Richter, 2018). The production test is usually in the form of free recall, short-answer essay, and fill-in-the-blanks, whereas the recognition test is in the form of multiple-choice questions (MCQs) and “true or false” (Larsen & Butler, 2013; Moreira et al., 2019). Moreira et al. (2019) noted that the differences with regard to the effectiveness of the different types of tests in drawing out the testing effect remains a question for both laboratory and classroom contexts, wherein there are mixed findings with regard to the effectiveness of different types of test format in producing the testing effect. In other words, some studies reported that tests which involve production task would provide greater benefit than recognition task (Butler & Roediger, 2007; Greving & Richter, 2018) whereas some studies reported otherwise (Little et al., 2012; Smith & Karpicke, 2014; Adesope et al., 2017). In Greving and Richter’s (2018) study which examined the testing effect in a classroom setting whereby the participants were asked to either answer short-answer questions, MCQs or read summarising statements about lecture content, it was found that short-answer testing may have provided an additional advantage in learning in higher education contexts compared to multiple-choice testing. The reason to such finding can be explained in terms of cognitive effort whereby questions that prompted effortful retrieval were likely to produce greater retrieval practice effects, of which in this case would be the short-answer test. However, there are some studies which revealed that MCQs are also as effective as production test in eliciting effects. Smith and Karpicke’s (2014) study found that there was little or no advantages in answering short-answer questions compared to MCQs. Such finding appeared due to the unsuccessful retrievals when using short-answer items whereby most of the time participants seemed to perform better in MCQs

than in short-answer tests. They explained that even though a more difficult test is said to be more advantageous, success of retrieval is equally as important for subsequent performance (Smith & Karpicke, 2014).

In terms of retrieval practice and levels of thinking, there has been an increasing number of studies which examine the effects of retrieval practice on different levels of thinking, that is, lower-order thinking and higher-order thinking (Butler, 2010; McDaniel et al., 2013; Carpenter et al., 2016; Dobson et al., 2018; Agarwal, 2019). In particular, these studies aim to determine whether transfer of knowledge from one level to another level (e.g., lower-level thinking to higher-level thinking or vice versa) would occur if students were instructed to engage in a learning task using the retrieval practice strategy. For example, participants were required to study a set of text passages and then were assigned to the experimental manipulation conditions where they either restudied the passage or engaged in a retrieval practice strategy (i.e., lower-order thinking test or higher-order thinking test). After a delay, they were given a final test that comprised both lower-order and higher-order thinking questions to observe the transfer of knowledge from one knowledge domain to another and to see whether the performance on the lower-order thinking task would enhance performance on the higher-order thinking task. Findings from the studies which examined this matter showed that even though the levels of thinking are usually said to follow a hierarchy, superior performance on lower-order thinking does not necessarily enhance performance of higher-order thinking tasks (e.g., McDaniel et al., 2013, Dobson et al., 2018, Agarwal, 2019). However, considering the majority of previous studies on retrieval practice and levels of thinking, there has yet to be a study that mainly investigates the effect of different types of question format used in retrieval practice (production test versus recognition test) on different levels of thinking. For instance, most studies that observe the

effect of retrieval practice on levels of thinking usually examine the effect of the transfer of thinking skills and most of the time they used only one question format, which is either a production test or a recognition test (Butler, 2010; Dobson et al., 2018; Agarwal, 2019). It is important to study the effects of both types of question formats especially in terms of addressing the effect in an educational setting since different question formats, particularly short-answer questions as well as MCQs are usually implemented in classrooms. Also, there are still ongoing arguments involving the issue of how MCQs and short-answer questions differ in terms of the success rate in the retrieval process and particularly its effect on learning (Karpicke, 2017). Therefore, this underlines the need for further investigation to find out whether there is any effect of the different types of question format on levels of thinking among students.

A growing number of studies investigated the retrieval practice effect and examined the effect on different lengths of delay have shown that when one adopts retrieval practice as a learning strategy, it benefits memory retention (Roediger & Karpicke, 2006; Wiklund-Hörnqvist et al., 2014; Karpicke, 2017). However, in terms of observing the retrieval practice effect especially involving the different types of retrieval practice formats, most studies did not observe the testing benefits across different retention intervals, that is, they either observed the effects of different types of retrieval practice format on short retention intervals or on long retention intervals (e.g., Butler & Roediger, 2007; Little et al., 2012; Smith & Karpicke, 2014). Therefore, not much is known about the long-term retrieval practice effect on the retention of production task and recognition task items (Stenlund et al., 2016). One of the very few studies that observed a retrieval practice effect using different types of formats on different retention intervals was conducted by Stenlund et al. (2016) who revealed that retrieval practice using short-answer items is more beneficial for long-term retention. On the

contrary, a study by Little et al. (2012) found that a well-constructed recognition test with competitive alternatives is also capable of drawing out productive retrieval processes. However, it observed the effect using short retention intervals, therefore, it remains unclear whether the benefits of multiple-choice testing can be retained for longer periods of time.

1.3 Problem Statement

Across a wide number of studies on retrieval practice effect that were conducted over the years, it is an established fact that retrieval practice enhances retention compared to rereading or restudying (Carpenter et al., 2006; Roediger & Karpicke, 2006; McDaniel, Roediger, et al., 2007; Rowland, 2014; Karpicke, 2017). When information is successfully retrieved from memory, the representation in memory is changed in a way that it becomes highly likely to be recalled more in the future (Bjork, 1975) thus, yielding a greater benefit compared to repeatedly studying or reading the materials which only involves encoding (Roediger & Karpicke, 2006). Previous studies stated that the mental process involved in actively retrieving information from memory when one is taking a test, is the mechanism driving the retrieval practice effect, wherein when one actively retrieves information, it strengthens retrieval routes in memory, hence permitting easier retrieval of the particular information in the future. (McDaniel & Masson, 1985; Carrier & Pashler, 1992; Carpenter & DeLosh, 2006; McDaniel, Roediger, et al., 2007; McConnell et al., 2018). Although restudying or rereading the passage exposes the students to the entire set of information again, engaging in retrieval practice allows the practice of the skills necessary for upcoming tests, resulting in greater performance after a delay (Roediger & Karpicke, 2006). While there is compelling evidence for the effects of retrieval practice, there are still questions with regard to the circumstances under which the effects arise, particularly in generalising the benefits of this effect in classroom setting (Greving & Richter, 2018). Therefore, in

consideration of the fact that there have been consistent findings on various contexts of studies which reported that retrieval practice has greater benefits than repeated reading or studying, the present study intends to focus mainly on the effects of the two types of retrieval practice format.

As mentioned earlier, few studies that looked into the different types of test (production test versus recognition test) have produced mixed findings. Some studies have found that tests which involve a production task will provide greater benefit compared to a recognition task (Butler & Roediger, 2007; Pyc & Rawson, 2009; Greving & Richter, 2018) whereas others have not (Little et al., 2012; Smith & Karpicke, 2014). A study by Greving and Richter (2018) which observed the retrieval practice effect in a classroom setting whereby the participants were required to either answer short-answer questions, MCQs or read summarising statements about lecture content discovered that the short-answer test may have additional benefits in learning compared to the multiple-choice test. Such a finding can be attributed to the role of cognitive effort by which questions that induce effortful retrieval are likely to produce greater retrieval practice effects, and in this case these would be the short-answer tests. On the contrary, some studies showed that MCQs are also effective in drawing out the retrieval practice effect. For example, Smith and Karpicke's (2014) study found that there was little or no advantages in answering short-answer tests over MCQs. This finding was based on the unsuccessful retrievals when using short-answer questions. The researchers explained that even though a more difficult test format is said to be more beneficial, the retrieval success rate is equally as crucial for later performance (Smith & Karpicke, 2014).

There has been a growing number of studies of the retrieval practice effect on different levels of thinking, which specifically examine the transfer of knowledge from lower-order thinking to higher-order thinking or vice versa (e.g., McDaniel et al., 2013, Dobson et al., 2018, Agarwal, 2019). However, looking at these studies, there has yet to be a study that mainly investigates the effect of different types of question format used in retrieval practice (production test versus recognition test) on different levels of thinking. For example, most studies that examined the retrieval practice effect on levels of thinking usually investigated the effect in terms of the transfer of thinking skills from one level to another and they usually used only one question format, which was either production test or a recognition test (e.g., Butler, 2010; Dobson et al., 2018; Agarwal, 2019). Considering the fact that different question formats, notably short-answer questions and MCQs, are commonly adopted in classrooms, it is important to explore the effect of both types of question formats, particularly in terms of addressing the retrieval practice effect in educational contexts. Besides that, student's higher-order thinking skills development is considered as an important element of education and recently, educators as well as scientists have been actively developing strategies and instructional approaches that aim to increase higher-order thinking (Agarwal, 2019). Thus, further examining the effects of different question formats on levels of thinking, can shed light on how educators can design learning materials that will improve students' performance and learning experience either for lower-order thinking questions or higher-order thinking questions.

The existing body of literature that observes the effect of retrieval practice on different lengths of delay has shown its benefits for memory retention which further proves that there is an added advantage when one uses retrieval practice as a learning strategy over rereading or restudying (Roediger & Karpicke, 2006; Wiklund-Hörnqvist et al., 2014;

Karpicke, 2017). However, with regard to the retrieval practice effect, in particular examining the effects of different types of retrieval practice format, most studies did not observe the testing benefits by comparing the different retention intervals. These studies either employed short retention intervals or long retention intervals (e.g., Butler & Roediger, 2007; Little et al., 2012; Smith & Karpicke, 2014) hence, little is known about the long-term effects of different types of retrieval practice format. A study by Stenlund et al. (2016) found that retrieval practice using a production task rather than a recognition task had more beneficial effects for long-term memory retention. Conversely, a study by Little et al. (2012) revealed that the multiple-choice test (recognition task), particularly one that was well-constructed, was also capable of eliciting productive retrieval processes compared to the cued-recall test (production task). Such finding is in contrast to the popular belief that the production task brings a greater benefit than the recognition task. However, the study observed the effect over short retention interval, and consequently, the long-term benefits of using multiple-choice tests remains uncertain.

1.4 Objectives

1.4.1 Main Objective:

To examine the effects of different types of retrieval practice on reading comprehension across levels of thinking and retention interval.

1.4.2 Specific Objectives:

- i. To examine if there is any significant difference in students' reading comprehension across the different types of retrieval practice.

- ii. To examine if there is any significant difference in students' reading comprehension between lower-order thinking and higher-order thinking.
- iii. To examine if there is any significant difference in students' reading comprehension between the different types of retention interval (short retention interval versus long retention interval).
- iv. To examine the interaction effect between types of retrieval practice and levels of thinking.
- v. To examine the interaction effect between types of retrieval practice and retention interval.
- vi. To examine the interaction effect between retention interval and levels of thinking.
- vii. To examine the interaction effect between types of retrieval practice, levels of thinking and retention interval.

1.5 Research Questions

- i. Is there any significant difference in students' reading comprehension across the different types of retrieval practice?
- ii. Is there any significant difference in students' reading comprehension between lower-order thinking and higher-order thinking?
- iii. Is there any significant difference in students' reading comprehension between the different types of retention interval (short retention interval versus long retention interval)?

- iv. Is there any interaction effect between the types of retrieval practice and levels of thinking?
- v. Is there any interaction effect between the types of retrieval practice and retention interval?
- vi. Is there any interaction effect between retention interval and levels of thinking?
- vii. Is there any interaction effect between the types of retrieval practice, levels of thinking and retention interval?

1.6 Definition of Terms

1.6.1 Retrieval Practice

Conceptual Definition of Term:

According to Roediger and Butler (2011), retrieval practice is a learning strategy which involves the act of recalling previously studied information. This process of recalling previously studied information is usually present when one takes a test (Karpicke, 2017).

Operational Definition of Term:

In the present study, retrieval practice refers to the test used in recalling previously learned items. Specifically, there were two types of tests used—production test and recognition test. Production test is defined as a test which requires the production of answers such as short-answer questions whereas recognition test is a test that requires the recognition of answers such as multiple-choice questions (MCQs).

1.6.2 Levels of Thinking

Conceptual Definition of Term:

Levels of thinking is defined according to a framework created by Bloom et al. (1956) which is usually known as Bloom's Taxonomy. In the taxonomy, one's level of thinking is said to be hierarchal, that is, it proceeds from lower levels to higher levels (Bloom et al., 1956). A revision of the Bloom's taxonomy was made in 2001 wherein verbs replaced the nouns in the original taxonomy: *remember*, *understand* (formerly known as comprehension), *apply*, *analyse*, *evaluate*, and *create* (formerly known as synthesis). Based on the revised taxonomy, higher-order thinking falls under *apply*, *analyse*, *evaluate*, and *create* categories, whereas lower-order thinking falls under the *remember* and *understand* categories (Anderson et al., 2001; Agarwal, 2019).

Operational Definition of Term:

In the present study, 'levels of thinking' refers to the lower-order thinking and higher-order thinking that are measured based on participants' performance in answering lower-order thinking questions and higher-order thinking questions in the tests. Lower-order thinking questions comprised of questions that fall under *remember* and *understand* categories in the Bloom's taxonomy whereas higher-order thinking questions were defined as questions that fall under *apply*, *analyse*, *evaluate*, and *create* categories in the Bloom's taxonomy.

1.6.3 Retention Interval

Conceptual Definition of Term:

Retention interval is defined as the period between when a participant is exposed to information and when they are tested for the retention of that particular information (American Psychological Association, n.d.).

Operational Definition of Term:

In the present study, retention interval refers to the testing done after the second distractor task. In particular, there were two types of retention interval used in the study, namely short retention interval and long retention interval. Short retention interval refers to the immediate testing after the second distractor task while long retention interval refers to the delayed testing completed three weeks after the second distractor task.

1.6.4 Reading Comprehension

Conceptual Definition of Term:

Reading comprehension refers to the process where one simultaneously obtains and constructs meaning via interaction with written language (Snow, 2002).

Operational Definition of Term:

In the present study, reading comprehension refers to the ability to read and understand a passage. Specifically, reading comprehension performance was measured based on the scores obtained by each participant in the final test phase.

1.7 Conceptual Framework

The present study consisted of three independent variables: retrieval practice, levels of thinking and retention interval. The dependent variable was the performance in the reading

comprehension task which was measured by the scores obtained in the final test phase (Figure 1.1).

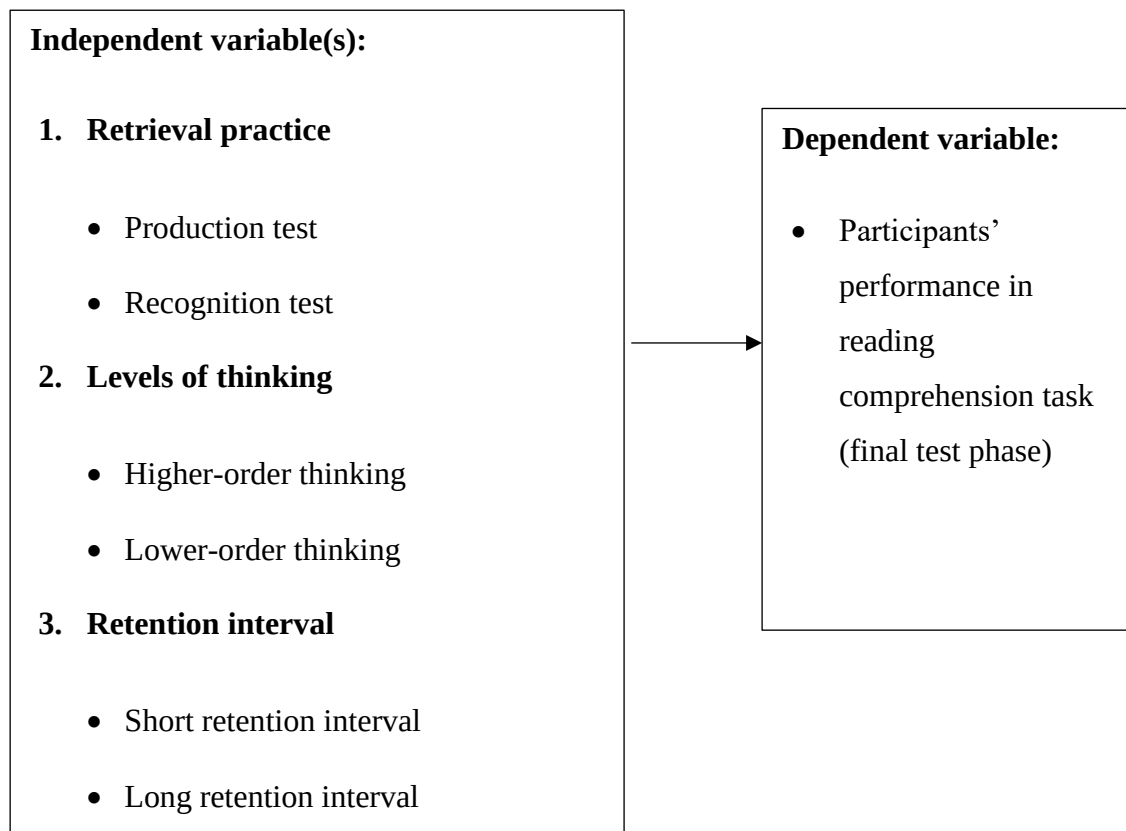


Figure 1.1: Conceptual Framework for the Present Study

1.8 Significance of Study

The findings from the present study in terms of the benefits of using tests or retrieval practice will provide further insight into the roles of testing in facilitating reading comprehension as well as learning in general. It will also shed light on the types of test question format or retrieval practice task that is applicable or suitable in enhancing learning as well as improving memory retention among students. Moreover, it can provide evidence that testing should not be used solely for the purpose of measuring the student's knowledge, an approach widely adopted in our local education institutions and usually conducted a

couple of times per semester (mid-term or final exam). Instead, it should be regarded as one of the learning strategies that can be incorporated regularly in classrooms either in the form of low-stake quizzes or short tests.

Furthermore, it was noted that students' development in higher-order thinking skills is the key component of education. Therefore, the findings from the present study in terms of the effects of different question formats on levels of thinking can shed light on which question format will yield greater benefits for the different types of levels of thinking (i.e., lower-order thinking or higher-order thinking). Such finding can assist educators in the designing of learning materials and instructional approaches that can facilitate students' higher-order thinking skills which in turn will enhance their learning experience. Also, the findings also allow both students and educators to see if the benefits can be retained for long periods of time.

Furthermore, by taking into account the issues that need further exploration, particularly in terms of observing the retrieval practice effect that involves different types of retrieval practice format across different retention intervals, more can be learnt pertaining to the long-term effects of different types of retrieval practice format. Additionally, the findings from the present study can also contribute to the existing body of knowledge on issues about retrieval practice effects and extends the findings from previous studies, in particular concerning the beneficial effects of using test or retrieval practice in educational settings.

1.9 Chapter Summary

This chapter has provided the general overview of the present study. It includes the background of the study, the problem statement that highlights on the issue that needs to be addressed, the research objectives as well as the research questions. This chapter has also

presented both conceptual and operational definitions of the terms used, the conceptual framework and also the significance of the present study.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Chapter Two presents a review of the literature related to the present study and it is arranged in subsections. It begins by introducing the retrieval practice effect, followed by the theoretical explanation for the benefits of retrieval practice on learning and the studies which compared the effects of using retrieval practice to rereading or restudying. It also includes a review of studies examining the different types of retrieval practice, learning materials and study topics that have been used in examining the retrieval practice effect. This chapter then discusses the related literature on the specific topics that the present study aims to examine (i.e., types of retrieval practice, levels of thinking and retention interval).

2.2 Introduction to Retrieval Practice

According to cognitive psychologists, the human memory process generally involves three stages which comprise encoding, storage and retrieval. Encoding is described as a process in which an individual acquires information and transfers it into memory. The encoded information is then kept in the memory and this process is called storage. To remember the information stored, it involves accessing the information or pulling out the information that has been encoded earlier and this process is referred to as retrieval (Goldstein, 2008; Sternberg & Sternberg, 2012). The same process also occurs when it comes to learning. When we read or study something, we encode the information and then the information is transferred and stored into our memory. Following that, we will be asked to recall or retrieve the information learnt usually in the form of a test or examination.

However, the popular idea surrounding tests especially in the classroom setting is that it is usually considered as an assessment tool (Brame & Biel, 2015; Moreira et al., 2019) as well as a tool in providing a summarised view of an individual's knowledge (Stenlund et al., 2016). Similarly, in the Malaysian education system, be it in schools or in higher learning institutions, tests are commonly used as a method of evaluating students' performance in the mid-term and final exams. In other words, our local education system commonly uses tests as a measure of students' level of understanding of the things they have learnt prior to the exam. Nevertheless, a growing number of studies have revealed that a test can do more than merely assess one's knowledge, that is, it can be used as a learning strategy which facilitates learning and promotes long-term retention in memory.

Numerous studies have found that when learners engage in the practice of retrieving previously learned information (i.e., retrieval practice), it can result in greater long-term memory retention compared to merely rereading or restudying that particular information; this phenomenon is called the testing effect (Carpenter et al., 2006; McDaniel, Roediger, et al., 2007; Moreira et al., 2019). Besides being termed as the testing effect, this phenomenon is also commonly referred to as the retrieval practice effect (Karpicke & Roediger, 2007a) or retrieval-based learning (Karpicke, 2017). Experimental procedures in examining this effect usually require participants to study a set of materials, either word pairs, word lists or reading passages. Then, they are instructed to take part in a retrieval practice task, either in the form of a free recall test, MCQs or a short-answer test. After engaging in the retrieval practice task, they have to complete a final test, either immediately or after a delay (e.g., hours, days, or weeks).

Studies on the retrieval practice effect usually observe and discuss final test performance, that is, the statistical analysis test employed is mainly focused on the data collected in the final test phase to measure participants' performance based on the retrieval practice conditions administered to them (Roediger & Karpicke, 2006; Cranney et al., 2009; Pyc & Rawson, 2009; Carpenter et al., 2016; Stenlund et al., 2016; Dobson et al., 2018; Thomas et al., 2018; Agarwal, 2019). With regard to the final test phase, the format of the retrieval practice task does not necessarily need to match the format of the final test, for example there have been several past studies which have used different formats for the retrieval phase and the final test phase (Lyle & Crawford, 2011; Bekkink et al., 2012; Dobson & Linderholm, 2015; Carpenter et al., 2016; Thomas et al., 2018). Additionally, numerous literatures have also stated that the format of the retrieval practice task or the initial test does not have to match the format of the final test for the beneficial effects of retrieval practice to occur (Karpicke & Blunt, 2011; Blunt & Karpicke, 2014; McDermott et al., 2014; Rowland, 2014; Karpicke, 2017).

2.3 Explanation for Benefits of Retrieval Practice in Learning

There are numerous theories that are used in explaining the effectiveness of tests. The first theory is the retrieval effort hypothesis which is driven by the desirable difficulty framework which states that a more difficult but successful retrieval has greater beneficial effects on memory compared to easier, successful retrievals (Gardiner et al., 1973; Pyc & Rawson, 2009). Pyc and Rawson (2009) conducted a study to test the effects of retrieval practice in promoting memory retention by using a cued-recall test of 70 Swahili-English word pairs. The study was done by manipulating two variables—interstimulus interval (ISI, in this study it was defined as the number of items between each next practice trial) and criterion level (the number of correctly retrieved items). Both manipulations were based on

the assumed relationship between the manipulated variables and retrieval difficulty. The first assumption was that correct retrieval of items is said to be more difficult after a longer ISI than after a shorter ISI, hence, contributing to a greater final test performance while the second assumption was that the number correctly retrieved items would affect the result in the final test performance. Findings from the study revealed that participants' performance in the final test increased when the difficulty of retrieval increased. Regarding the second manipulated variable (i.e., criterion variable), it was revealed that when the criterion level increased, retrieval became less difficult, thus resulting in a decrease in the final test performance. These findings therefore confirmed the assumptions on the retrieval effort hypothesis which state that successful but difficult retrievals are more beneficial in enhancing memory, contrasted with successful but less difficult retrievals. Prior to Pyc and Rawson's study (2009), a study by Gardiner et al., (1973) also provides evidence that retrieval difficulty can enhance memory retention. Participants in their study were tested on 50 definition-word pairs and findings showed that difficult retrievals of words in the initial test were recalled significantly better in the final recall test, thus implying that greater retrieval effort yields greater benefits on memory. With regard to the desirable difficulty framework from which the retrieval effort hypothesis is derived, the framework suggests difficult retrievals are more desirable than easy ones (Pyc & Rawson, 2009). Based on the desirable difficulty framework, learning strategies which requires more effort and are cognitively demanding during the encoding or learning phase compared to activities that requires minimal cognitive demand (i.e., restudying or rereading) may hamper short-term recall but can assist in constructing necessary networks in memory to improve long-term recall (Bjork, 1994).

The second theory that is commonly utilised in explaining the retrieval practice effect is transfer-appropriate processing. According to this theory, the resemblance in an individual's cognitive process during the initial learning activity and the cognitive process required in the final assessment results in greater performance in the memory task (Morris et al., 1977; Kolers & Roediger, 1984). This theory is based on the assumption that the favourable effects on long-term memory retention are due to the similarity of cognitive processes that occur during learning tasks and transfer tasks (Endres & Renkl, 2015; McDermott, 2021), which also explains why retrieval practice outperforms restudying or rereading. This advantage exists due to the similarities in the cognitive processes that are present during the transfer tasks and during the initial testing, compared to the processes during restudying or rereading (e.g., Thomas & McDaniel, 2007). A study that supported this theory was accomplished by Thomas and McDaniel (2007) who aimed to investigate whether the congruency of processes produced during study and required during retrieval would affect memory, metacomprehension, and control processes. In the first experiment, participants were required to read six text passages and for each passage, they had to perform the assigned encoding task (read, letter-reinsertion or sentence-sorting), followed by metacomprehension predictions for the previously learnt materials whereby participants were presented with the passage again so that they could make the predictions, and then perform a cued-recall test (detailed test or conceptual test). There were three conditions observed in the first experiment pertaining to the congruency of process during the encoding task and test type: congruent (i.e., letter-reinsertion–detailed test), incongruent (i.e., letter-reinsertion–conceptual test), or neutral (i.e., read–detailed test or read–conceptual test). The second experiment replicated the first one but without the read condition and with an additional stage whereby participants were presented with a section of the passage similar to

the ones used for them to make metacomprehension predictions. The participants were told they could spend as much time or as little time restudying the section of the passage whichever they thought was necessary. This additional stage was included to find out whether participants would allocate more study time to the items deemed to be more difficult in both congruent and incongruent task-test processing conditions. In both experiments, the study revealed that when cognitive processes during study and test were similar, cued-recall test performance and metacomprehension prediction accuracy were better than when study and test were incongruent or when the participants merely read the text passages. Another study that supports this theory was conducted by Johnson and Mayer (2009) whereby participants were asked to watch a narrated animation about lightning formation and then for the experimental manipulation conditions, they either had to watch the presentation again (restudy), take a practice-retention test (recall), or take a practice-transfer test (short-answer questions). Participants were then given a final test which comprised both retention question (recall) and transfer questions (short-answer questions). Results from this study revealed a significant interaction between the study and the test condition which is consistent with the transfer-appropriate processing theory whereby individuals in the practice-retention group did better in the delayed retention test whereas those in the practice-transfer group did better in the delayed transfer test. Findings from these studies (e.g., Thomas & McDaniel, 2007; Johnson & Mayer, 2009) further proved the assumption on the transfer-appropriate processing which states that when there is a similarity between the student's cognitive processing during an initial learning activity and the processing required during the final assessment, it enhances learning and memory task performance (Morris et al., 1977; Kolers & Roediger, 1984).

The third theory is elaborative-retrieval account which is another theory to explain the retrieval practice effect. This theory suggests that semantic elaboration which occurs during the process of retrieval will improve later recall (Karpicke, 2017). This theory is driven by the notion that when one retrieves a target from a given cue, it activates additional information that is semantically associated to the cue and this information is incorporated with the successfully recalled target to create an elaborate memory trace that becomes easily remembered at a later time (Carpenter & Yeung, 2017; Karpicke, 2017). Restudying or rereading material on the other hand, does not always result in the production of additional mediators since there is no activity of retrieving the target word during restudy trials (Carpenter, 2009), thus explaining why retrieval practice enhances memory retention more than just restudying or rereading. A study that provides evidence for this theory was conducted by Carpenter (2009). In this study, participants were required to study cue–target pairs, which were either strongly associated (e.g., *Toast: Bread*) or weakly associated (e.g., *Basket: Bread*). After this, they were instructed to either restudy the pairs (e.g., *Toast: Bread*) or take an initial cued recall test (e.g., *Toast: _____*). A final test in the form of free recall of the targets was given, and the results showed significant positive effects of retrieval practice on memory retention whereby tested items were retrieved better than restudied items. The comparison between strong cues and weak cues on the other hand revealed that even though strong cues provided an added advantage during initial recall, items retrieved from weak cues were more likely to be remembered over time. One reason for this finding could be that strong cues do not encourage the activation of elaborative information that benefits later retrieval. In the case of a strong cue (e.g., *Toast: _____*), the target item (e.g., *Bread*) is easily retrieved during initial recall, which may generate unfavourable effects on memory retention afterwards. However, for weakly associated cues such as “*basket*”,

participants would have to generate numerous related words in an attempt to recall the target item (e.g., *fruits*, *sandwich*, etc.) thus making the items retained in memory better over time. The finding from this study, particularly in terms of testing versus restudying provides support for the elaborative-retrieval account which states that items that were retrieved were retained in the memory better than the items that were restudied due the activation of elaborative information that occurs to a greater degree during testing than during restudying. This explanation also extends to the comparison between strong cues and weak cues whereby weak cues called for greater activation of elaborative information. For weak cues, additional cue-relevant elaborations were needed to be encoded with the target item in order to produce an elaborate representation in the memory, hence resulting in better memory retention (Carpenter, 2009; Karpicke, 2017).

2.4 Effects of Retrieval Practice Relative to Restudying or Rereading

A growing body of literature in both laboratory and educational settings ranging from elementary school students to university students has found that retrieval practice has greater benefits compared to merely rereading or restudying (Roediger & Karpicke, 2006; Butler & Roediger, 2007; Karpicke & Roediger, 2008; Carpenter et al., 2009; Brame & Biel, 2015; Jaeger et al., 2015; Stenlund et al., 2016; Moreira et al., 2019). Studies comparing the two conditions have observed the effect on various issues, such as multiple testing versus multiple reading or studying opportunities (Roediger & Karpicke, 2006; Karpicke & Roediger, 2007b), different retrieval practice formats versus restudying (Butler & Roediger, 2007; McDaniel, Anderson, et al., 2007; Stenlund et al., 2016) and also on retrieval practice with feedback versus restudying (Carpenter et al., 2008; Wiklund-Hörnqvist et al., 2014).

Findings from the studies observing the retrieval practice effect have shown that testing has more beneficial effects on long term memory retention compared to repeatedly studying the learned materials (Roediger & Karpicke, 2006; Karpicke & Roediger, 2007b). For instance, a study by Roediger and Karpicke (2006) which aimed to examine the testing effect under an educationally relevant setting, whereby they conducted two experiments among university students. In the first experiment, participants were required to take a test on the learned material or study it again before taking a final retention test five minutes, two days, or one week later. In the second experiment on the other hand, participants were instructed to study a passage once and they were then assigned to either one of three conditions, namely repeated study (SSSS), single test (SSST), or repeated test (STTT). They were then required to take a final test five minutes or one week later. Results from both experiments showed that immediate testing after reading a passage enhanced long-term memory retention compared to studying the passage once or studying the passage multiple times. Regarding the issue of repeated testing versus repeated studying, findings showed that on the final tests given after a five-minute retention interval, repeated studying enhanced performance compared to repeated testing; however, the effect was reversed on delayed tests. In short, when one restudies a material, it only gives short-term benefits to memory but when one engages in test or retrieval practice, it has positive effects on long-term memory retention. This study therefore shows that although there is benefit in repeated study soon after learning, repeated testing produces strong beneficial effects on a delayed test.

In addition, there are also studies that have compared the different types of retrieval practice format with rereading or restudying (Butler & Roediger, 2007; McDaniel, Anderson, et al., 2007; Stenlund et al., 2016; Moreira et al., 2019). In particular, a study by Stenlund et al. (2016) investigated the short- and long-term benefits of repeated practice

testing with short-answer items in comparison to repeated practice testing with multiple-choice items and rereading the material among secondary school students. Participants in the study were required to read a 16-pages long study passage taken from a textbook on the driver's education programme during the 30-minute learning session. The learning session was then followed by two practice sessions in which the facts from the textbook were presented either as multiple-choice items, short-answer items, or statements that students were asked to read (restudy). The final test sessions took place at three different time delays: after five minutes, after one week and after four weeks. Results of the experiment showed that both the multiple-choice test and the short-answer test formats had positive effects on short- and long-term memory retention compared to merely rereading the material. Such finding was also present in a study by McDaniel, Anderson, et al. (2007) which observed the effects of retrieval practice in a college course. The researchers found that both multiple-choice and short-answer formats provided advantage in comparison to additional reading opportunities. These findings provide evidence that as long as one engages in retrieving the information learned regardless of the format, it brings more benefits to memory retention as compared to only repeatedly reading the material.

One of the varied issues examined when comparing retrieval practice with restudying or rereading includes the effects of retrieval practice with feedback relative to rereading (Carpenter et al., 2008; Wiklund-Hörnqvist et al., 2014). To this end, Wiklund-Hörnqvist et al. (2014) investigated whether repeated testing with feedback promoted learning better than rereading of cognitive psychology key concepts among undergraduates. This study was based on the belief that feedback can boost learning as it may act as an additional learning opportunity which helps to reduce repeated retrieval failures. In the learning phase, participants were required to study key concepts from three topics in a cognitive-psychology

course (i.e., attention, memory, and perception) during the lecture. Immediately after the topic lecture, they were randomly assigned to either the repeated testing group with feedback or the restudy group. The effect of testing was observed using the short-answer question format and at three different retention intervals: immediately after practice, after 18 days, and after a five-week delay. Results from the study found that repeated testing with feedback significantly improved students' test performance in comparison to rereading the material at all three delays, indicating that repeated retrieval boosts short- and long-term memory retention more than repeatedly reading the materials. This finding therefore proves that retrieving an item is more beneficial than repeatedly encoding the learned information.

Most studies observing retrieval practice effects usually examine the effect on university students (Roediger & Karpicke, 2006; Butler & Roediger, 2007; McDaniel, Anderson, et al., 2007; Wiklund-Hörnqvist et al., 2014) and also secondary school students (Carpenter et al., 2009; Stenlund et al., 2016). Additionally, the positive effects of retrieval practice in comparison to rereading also extends to primary school students. For instance, a study by Jaeger et al. (2015) investigated the retrieval practice effect among children by using educationally related materials. The experiment was conducted on third-grade children with ages ranging from 8 years old to 10 years old who were required to read a short text containing 321 words about the sun during the study phase and after that, were assigned to the test group where they were instructed to perform the cued recall test on target facts from the text. On the other hand, participants in the restudy group had to read the text passage twice. The findings from the final MCQ test which took place seven days after the experimental manipulation showed that students who took the initial cued recall test demonstrated greater performance than the students who only restudied the text twice, thus

suggesting that the effects of retrieval practice over restudying or rereading are also present among younger learners.

The explanation for why retrieval practice is more beneficial than rereading or restudying is because when information is successfully retrieved from memory, the representation in memory is changed in a way that it becomes more highly likely to be recalled in the future (Bjork, 1975), and this improvement is usually greater than engaging in additional study or reading (Roediger & Karpicke, 2006). Previous studies stated that the mental process of actively retrieving information from memory when one is taking a test, is the mechanism driving the retrieval practice effect, in which, when one actively retrieves information, it strengthens retrieval routes in memory, thus allowing easier retrieval of the particular information in the future (McDaniel & Masson, 1985; Carrier & Pashler, 1992; Carpenter & DeLosh, 2006; McDaniel, Roediger, et al., 2007; McConnell et al., 2018). Even though restudying the passage gives exposure of the entire set of information to the students, engaging in retrieval practice permits the practice of the skills required on upcoming tests, thus improving the performance after a delay (Roediger & Karpicke, 2006). The wide range of empirical studies conducted over the years in comparing the effects of engaging in retrieval practice task to rereading or restudying conclusively show that it is an established fact that retrieval practice results in more beneficial effects in learning compared to merely rereading or restudying (Rowland; 2014; Karpicke, 2017).

2.5 Types of Retrieval Practice, Learning Materials and Study Topics

Studies in the past have observed the testing effect or retrieval practice effect using various retrieval practice tasks such as free recall, cued-recall, fill-in-the-blank questions,

MCQs and short-answer questions (Hinze & Wiley, 2011; Smith et al., 2016; Agarwal, 2019; Moreira et al., 2019).

The effects of retrieval practice have also been examined using several different learning materials which include language word pairs (Carpenter et al., 2008; Pyc & Rawson, 2009), face-name pairs (Landauer & Bjork, 1978; Carpenter & DeLosh, 2005), text passages (Agarwal et al., 2008; Little et al., 2012; Smith & Karpicke, 2014; Agarwal, 2019), word list (Carpenter & DeLosh, 2006), term definition pairs (Carpenter et al., 2016), lectures (Lyle & Crawford, 2011; Burdo & O'Dwyer, 2015), video-based learning (Butler & Roediger, 2007; Yong & Lim, 2016) and map-learning tasks (Rohrer et al., 2010).

In addition, previous studies have also examined this effect using a variety of topics, for instance cognitive psychology (Wiklund-Hörnqvist et al., 2014; Grieving & Richter, 2018), general knowledge (Carpenter et al., 2008), anatomy and physiology (Dobson & Linderholm, 2015), statistics (Lyle & Crawford, 2011), biology (Carpenter et al., 2016) as well as language (Pyc & Rawson, 2009; Kang et al., 2013).

To summarise, the previous studies that observed on retrieval practice effect showed that the positive effects of retrieval practice is not limited to certain criteria and that it can be generalised to different formats and to a wide range of learning materials as well as study topics.

2.5.1 Production Test Versus Recognition Test

In terms of the test format used in examining the testing effect or retrieval practice effect, the commonly used test formats can be generally categorised into two types which are production test and recognition test (Butler & Roediger, 2007; Larsen & Butler, 2013;

Greving & Richter, 2018). The production test is usually in the form of a free recall test, cued-recall, short-answer, essay or fill-in-the-blank, whereas the recognition test comprises MCQs, true or false and answer matching (Larsen & Butler, 2013). Moreira et al. (2019) noted in their recent study that the differences with regard to the effectiveness of the different types of test in drawing out testing effects remains a question for both laboratory and classroom contexts. Few studies have found that tests which involve production tasks, rather than recognition tasks, provide greater benefit (Butler & Roediger, 2007; Stenlund et al., 2016; Greving & Richter, 2018) while some studies have reported otherwise (Little et al., 2012; Smith & Karpicke, 2014; Adesope et al., 2017).

In a recent study, Greving and Richter (2018) examined the testing effect in a university lecture whereby the participants were asked to either answer short-answer questions, MCQs or read summarising statements about core lecture content (restudy) during the last 10 minutes of each lecture session to review the lecture content. The retention of the learning content was measured at different times (one week, 12 weeks and 23 weeks after the last lecture) and the results suggested that short-answer testing may have benefitted learning in higher education contexts more than multiple-choice testing, independent of the time of test (Greving & Richter, 2018). This finding was also consistent with the finding by Butler and Roediger (2007). In their study, they observed the benefits of testing in a simulated college classroom setting in which undergraduate participants were required to watch a series of three lectures on art history for three consecutive days and then were asked to perform different learning activities after each lecture (i.e., taking a multiple-choice test, taking a short-answer test, or studying a lecture summary). Findings from the final short-answer test which took place a month after the learning session revealed that when the participants engaged in the short-answer test as a learning strategy, they had significantly

better memory retention of the learned material compared to studying a lecture summary or taking a multiple-choice test. Previous findings which found that the short-answer test is more advantageous than MCQs are supported by the retrieval effort hypothesis which suggests that a more difficult and effortful retrieval process results in greater benefits (Gardiner et al., 1973; Pyc & Rawson, 2009).

However, in Smith and Karpicke's (2014) study which examined the effects of retrieval practice by comparing the different question formats on learning among university students, it was found that there was little or no advantages in answering short-answer questions over MCQs. In this study, participants were required to read text passages and then practised retrieval either using short-answer question format, MCQ format, hybrid questions format (i.e., short-answer and MCQ format) or study-only control condition. Results from the final short-answer assessment which took place a week after the experimental manipulation showed that practising retrieval in all format conditions improved memory retention, in comparison to the study-only control condition. However, there were no differences in the effectiveness of initial retrieval formats (short-answer, multiple-choice, or hybrid formats) observed in the study. In other words, retrieval practice effects among the different formats were almost similar to one another. Although multiple-choice tests are usually assumed to yield little to no retrieval practice effects, the finding from the study suggested that retrieval practice with multiple-choice tests can be quite helpful in enhancing subsequent performance. Such finding was due to unsuccessful retrievals when using short-answer test, in which, it was reported that participants did better in MCQ than in short-answer testing for most of the time. Smith and Karpicke (2014) explained that although a more difficult test is said to be more beneficial, success of retrieval is equally as important for later performance. Moreover, Little et al. (2012) discovered that properly constructed

multiple-choice tests can result in productive retrieval processes. Properly constructed MCQs in this case were utilised by including competitive incorrect options in the choices of answers. However, the study by Little and colleagues was done over a short-term basis and one of the issues raised in the study was how long can the benefits of multiple-choice testing last and whether the testing effect can be retained for longer periods of time.

2.5.2 Retrieval Practice and Levels of Thinking

Levels of thinking is usually categorised according to The Taxonomy of Educational Objectives, a framework published by Bloom, Engelhart, Furst, Hill and Krathwohl in 1956 (Agarwal, 2019), otherwise known by many as Bloom's Taxonomy. A revision of Bloom's Taxonomy was made in 2001 wherein verbs replaced the nouns in the original taxonomy: *remember*, *understand* (formerly known as comprehension), *apply*, *analyse*, *evaluate*, and *create* (formerly known as synthesis). Based on the revised taxonomy, higher-order thinking falls under *apply*, *analyse*, *evaluate*, and *create* categories, whereas lower-order thinking falls under the *remember* and *understand* categories (Anderson et al., 2001; Agarwal, 2019).

With regard to retrieval practice and levels of thinking, there have been an increasing number of studies examining the effects of retrieval practice on different levels of thinking (i.e., lower-order thinking and higher-order thinking). Particularly, these studies aimed to determine whether transfer of knowledge from one level to another level (e.g., lower-level thinking to higher-level thinking or vice versa) would occur if students were instructed to engage in a retrieval practice task as the learning strategy (McDaniel et al., 2013; Dobson et al., 2018; Agarwal, 2019). For instance, a recent study was conducted by Agarwal (2019) where she observed the relationship between fact learning (lower-order learning) and higher-order learning using retrieval practice as the learning strategy as well as examining the

foundation of factual knowledge framework which states that it is important to concentrate on strengthening basic factual knowledge before fostering students' higher order learning. Another series of experiments was conducted to observe whether the transfer of knowledge from one level to another level (e.g., lower-order thinking to higher-order thinking or vice versa) would occur when the participants engaged in the retrieval practice strategy. In one of the experiments, participants were asked to study a set of text passages and then they were required to partake in four retrieval practice conditions: a study once condition, a study twice condition, study once followed by a fact quiz, and study once followed by a higher-order quiz. After two days, students completed fact and higher-order tests for each condition. Results from the experiment found that retrieval practice with fact questions did not improve delayed higher-order test performance and retrieval practice with higher-order questions also did not improve delayed fact test performance. Findings from the studies provided evidence that although the levels of thinking are usually said to be hierarchical, superior performance on lower-level thinking task does not necessary enhance performance in higher-order thinking task and that as long as retrieval practice is used to engage with higher-order thinking task, it will bring more benefits to the participants' performance compared to starting with basic facts or definitions (McDaniel et al., 2013; Dobson et al., 2018; Agarwal, 2019).

By reviewing past studies however, there has yet to be a study that specifically investigates the effect of different types of question format used in retrieval practice (production test or recognition test) on different levels of thinking. In particular, most studies that observe the effect of retrieval practice on levels of thinking usually examine the effect on the transfer of knowledge and most of the time they use only one question format, which is either a production test or a recognition test. Therefore, this highlights the need to further

explore if there is any effect of the different types of question format on different levels of thinking among students.

2.5.3 Retrieval Practice and Retention Interval

Numerous studies have shown that in comparison to rereading or restudying, retrieval practice either in the form of short-answer test, multiple-choice test or free recall has greater benefits for long-term memory retention. (Carpenter et al., 2006; Karpicke & Roediger, 2007a; McDaniel, Roediger, et al., 2007). Studying is said to be more effective when the information being learned is only needed for a short time whereas testing is more effective for better long-term memory retention (Roediger & Karpicke, 2006; Dobson & Linderholm, 2015). A study by Roediger and Karpicke (2006) which observed the testing effect on different lengths of delay found that when the final test was given after five minutes, repeated studying improved retrieval compared to repeated testing. However, on the delayed tests (two days or one week), testing produced noticeably greater retention as compared to studying. Such finding can be explained by the desirable difficulty framework that states a more difficult and effortful retrieval during the learning phase (i.e., testing condition) may impair performance during short-term recall but is helpful in building the necessary memory retrieval network to improve long-term retention (Bjork, 1994; Dobson & Linderholm, 2015). Similarly, repeated studying or reading requires less cognitive demand and is less difficult compared to testing (Larsen & Butler, 2013; Dobson & Linderholm, 2015). Therefore, engaging in such a learning strategy only allows for short-term benefits. Additionally, a study by Wiklund-Hörnqvist et al. (2014) which observed the retrieval practice effect on different length of delay—immediately after practice, after 15 days and after five weeks, discovered that repeated retrieval practice enhanced learning

compared to rereading at all delays. This finding thus implies that repeated retrieval enhances retention in comparison to repeated encoding both in the short and the long-term.

As for the effects of type of test across different retention intervals, most studies however, did not observe the testing benefits across different retention intervals, that is, they either observed over a short-term retention interval or a long-term retention interval. For instance, the study by Smith and Karpicke (2014) investigating the effects of different question formats on learning had observed the effect after one week whereby the final assessment was administered one week after the experimental manipulation session. Likewise, Butler and Roediger (2007) observed the retrieval practice effect across different formats at one month delay after the learning session (retrieval practice session using either MCQ test, short-answer test, or restudying). Therefore, not much is known about the short- and long-term effects of different types of retrieval practice format. One of the few studies that observed the effects of different retrieval practice formats on different retention intervals was carried out by Stenlund et al. (2016). Specifically, their study was to test the hypotheses of retrieval effort and transfer appropriate processing by comparing different retention intervals relating to different retrieval practice formats in which they observed the effect at three different retention intervals (i.e., after five minutes, after one week and after four weeks). The findings from their study showed that both retrieval practice formats, particularly short-answer and multiple-choice improved short- and long-term memory more than merely rereading. More significantly, the study revealed that retrieval practice using short-answer items was more beneficial for learning as well as long-term retention. On the contrary, a study by Little et al. (2012) found that a properly constructed recognition test (i.e., multiple-choice test) can also result in productive retrieval processes and has more advantage over a cued-recall test. Such finding contradicts the popular belief that production

task brings greater benefit than recognition tasks. However, Little and colleagues' (2012) study observed a testing effect over short retention interval, so, it remains unclear whether the benefits of multiple-choice testing can be retained for longer periods of time.

2.6 Chapter Summary

This chapter has reviewed past studies related to the retrieval practice effect which included the theoretical explanation of using retrieval practice on learning as well as the benefits of using retrieval practice over rereading or restudying. This chapter has also reviewed previous studies associated with types of retrieval practice, levels of thinking and retention interval. The review has pointed out that there are still underlying issues particularly in terms of production test versus recognition test whereby past studies have reported mixed findings. In this regard, there are several studies which reported that the production test is more beneficial than the recognition test whereas other studies have proved that the recognition test is just as effective as the production test. Additionally, this chapter also highlighted the need to address the issue regarding retrieval practice and levels of thinking in which there has yet to have a study that mainly investigates the effect of different question formats used in retrieval practice (production test versus recognition test) on different levels of thinking. Most studies usually examine the effect on the transfer of knowledge and most of the time they use only one question format. The review of past studies in this chapter also underlined the issues on retrieval practice and retention interval whereby most studies did not observe the retrieval practice benefits across different retention intervals and that they either observed the effect under short retention interval or long retention interval.

CHAPTER 3

METHOD

3.1 Introduction

Chapter Three presents the method used in conducting the research. This chapter begins by describing the target population of the study, in particular, the participants involved, followed by the description of the research design as well as the research materials used in conducting the experiment. This chapter also outlines the data collection procedure as well as explains the method of data analysis.

3.2 Participants

The target population for this study comprised undergraduate students from a local public university. A total of 201 participants among the first-year students enrolled in the Cognitive Psychology course were involved in the study. To ensure that the experiment was equitable, pre-screening of the participants was done based on their Malaysian University English Test (MUET) result. Students with MUET Band 3 and above were selected to participate in the study. The reason why MUET Band 3 was set to be the minimum requirement for participation was to ensure that participants had at least a moderate level of English proficiency so as to reduce the chances of language proficiency affecting their performance. Out of the 201 students who were initially recruited, 100 students were identified to meet the MUET requirement, and thus participated in the actual experimental manipulation of this study.

3.3 Design

The experiment followed a 2 (Retrieval practice: production test, recognition test) x 2 (Levels of thinking: higher-order thinking, lower-order thinking) x 2 (Retention Interval: short retention interval, long retention interval) mixed-subjects design. Retention interval was varied between subjects whereas retrieval practice phase and levels of thinking were varied within subjects. The present study comprised three independent variables which included retrieval practice, levels of thinking and retention interval whereas the dependent variable was performance in a reading comprehension task which was measured by the scores obtained in the final test phase. The experiment included the following phases: a) Study phase: participants were asked to read a passage on Problem Solving, b) First distractor task: participants were asked to do a word search puzzle, c) Retrieval practice phase: participants were asked to answer 10 recognition test questions (MCQs) and 10 production test questions (short-answer questions) on the reading passage. Each of the ten questions consisted of five lower-order thinking (LOT) questions and five higher-order thinking (HOT) questions. For this phase, participants were asked to answer a total of 20 questions, d) Second distractor task: participants were asked to do a word search puzzle, and e) Final test: participants were required to answer a total of 20 multiple-choice questions (MCQs), comprising 10 lower-order thinking questions and 10 higher-order thinking questions, similar to the questions in the retrieval practice phase whereby it covered the same aspect, but with slight modifications in terms of the structure of questions. For example, the recognition test questions from the retrieval practice task were changed by rearranging the positions of answer choices and some words were rephrased whereas production test questions were converted to MCQs format (see Appendix C)

3.4 Research Materials

The materials used in the present study are discussed according to the phases of the study.

3.4.1 Study Phase

The material for the study phase consisted of a reading passage from one of the chapters in the Cognitive Psychology course content—Problem Solving (see Appendix A). This topic was chosen because it had not been covered in the class lecture and the approaches to problem solving were something that the participants were unfamiliar with. The length of the reading passage was 980 words which was appropriate for university students (e.g., Little & Bjork, 2012; Agarwal, 2019;).

3.4.2 Retrieval Practice Phase

In this phase, participants were asked to restudy the passage using retrieval practice that comprised both recognition and production tests. Then, participants were required to answer a total of 20 questions. The recognition test used 10 MCQs, whereas the production test used 10 short-answer questions (see Appendix B). In both the recognition and production tests, the questions asked tested the participants' comprehension on the reading passage given earlier. In particular, each of the 10 questions in each test (recognition and production) consisted of five lower-order thinking questions and five higher-order thinking questions (Figure 3.1). Based on the findings by Little et al. (2012) which discovered that properly constructed MCQs can elicit a productive retrieval process, the present study therefore decided to follow the same method, that is to include competitive alternatives in the MCQs to ensure that the MCQs cannot be attributed to being relatively easier than short-answer questions. All the questions used in the present study were validated by the course instructor

(an expert in Cognitive Psychology) to ensure that each question matched the level of thinking (LOT versus HOT) required. As for the scoring, MCQs were scored with one point for each correct response and zero point for any incorrect response. For the short-answer questions, a grading rubric was created by consulting the course instructor. A fully correct response that contained keywords or phrases based on the rubric was graded with one point whereas half a point was given for a partially correct response. Zero points were given for any incorrect response or no response.

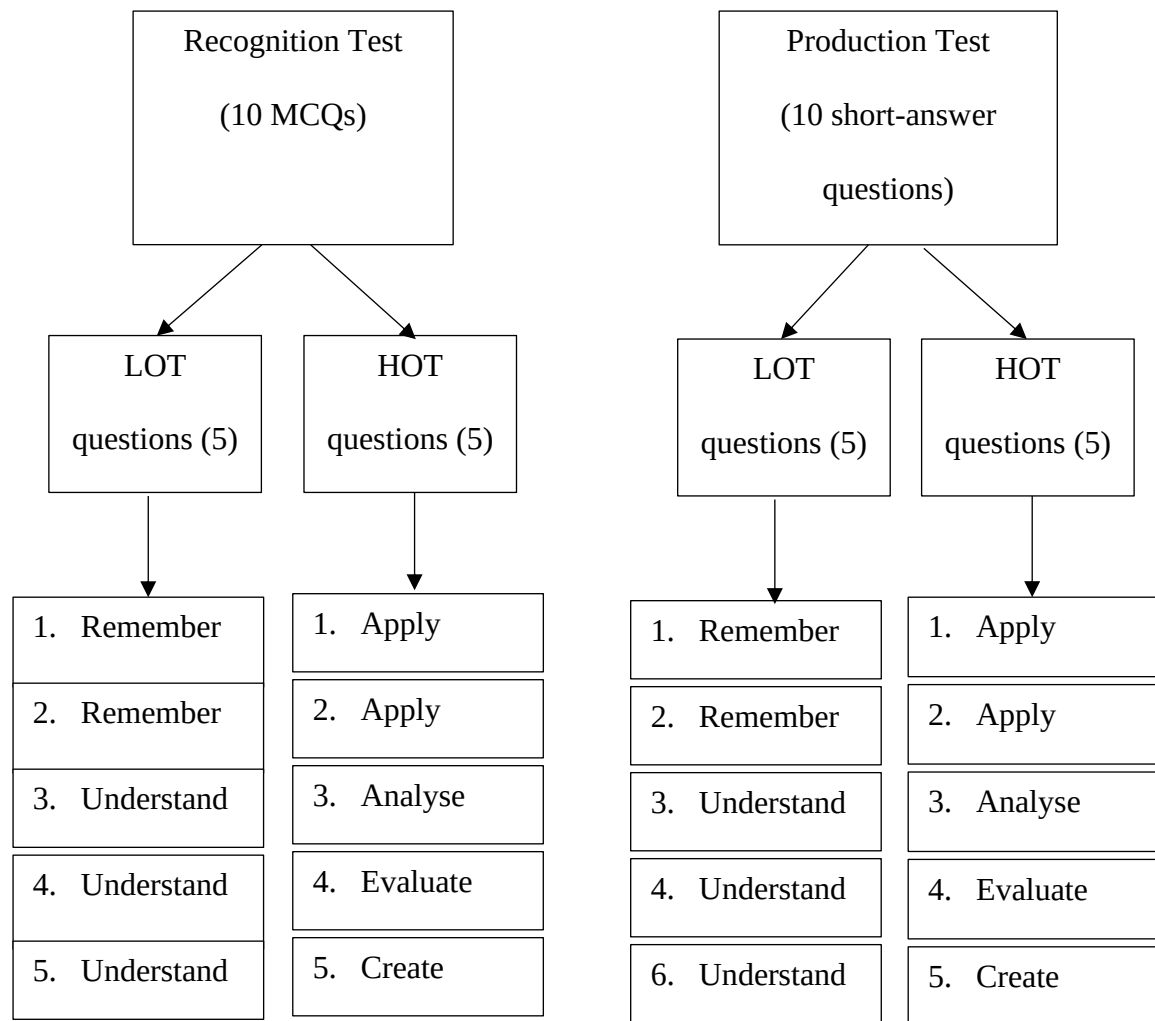


Figure 3.1: Distribution of Questions in the Retrieval Practice Phase

3.4.3 Final Test Phase

A total of 20 MCQs were asked. They comprised 10 lower-order thinking questions and 10 higher-order thinking questions, similar to the ones asked during the retrieval practice phase but with slight changes (see Appendix C). The structure of the questions in this phase was slightly altered by converting short-answer questions to MCQ format, rephrasing certain words, and rearranging the positions of the answer choices, but the context of the questions remained the same as the ones in the retrieval practice phase. The questions taken from the recognition test were modified by rearranging the position of the alternatives and rephrasing

certain words whereas the questions taken from the production test were translated to MCQ format, while still retaining the context of the questions so that it remained similar to the ones in the retrieval practice phase (see example of questions in Figure 3.2). The structure of questions was slightly modified to reduce the chances of students memorising the answers from the retrieval practice phase, particularly from the recognition test. The MCQ format was chosen for the final test format because there have been past studies that observed a testing effect when MCQs were used as the final test (Dobson & Linderholm, 2015; Carpenter et al., 2016). Additionally, the MCQ type was chosen in the final test because it merely acts as an evaluation of the participants' performance based on the different types of retrieval practice administered to them. Therefore, the types of formats used for the final test does not matter (Karpicke, 2017). Moreover, there were previous studies on the testing effect that used different formats for both the retrieval phase and final test phase (see Dobson & Linderholm, 2015; Carpenter et al., 2016). Furthermore, several literatures have also noted that the format of the initial test did not have to match the format of the final test for testing benefits to occur (Karpicke & Blunt, 2011; Blunt & Karpicke, 2014; McDermott et al., 2014; Karpicke, 2017). The scoring for the MCQs in the final test followed the system used in the retrieval practice phase—each question was scored with one point for a correct response and zero points for any incorrect response.

Retrieval practice phase	Final test phase
Production test (Lower-order thinking [LOT] question) “Occurs when there is an obstacle between a present state and a goal, and the solution to get around the obstacle is not immediately obvious”. The above description refers to ...	MCQ (Lower-order thinking [LOT] question) “The solution to get around the obstacle is not immediately obvious”. The above description refers to ... A. Representation B. Analogy C. Restructuring <i>D. Problem</i>
Recognition test (Higher-order thinking [HOT] question) “A car starts from rest and accelerates uniformly over a time of 5.21 seconds for a distance of 110 m. Determine the acceleration of the car.” The above mentioned problem demonstrates A. A well-defined problem B. Problem solving involving insight C. An ill-defined problem D. Problem solving involving restructuring	MCQ (Higher-order thinking [HOT] question) “A man's speed with the current (rate of movement in the water) is 15 km/hr and the speed of the current is 2.5 km/hr. The man's speed against the current is?” The above mentioned problem demonstrates A. an ill-defined problem B. problem solving involving insight C. problem solving involving restructuring <i>D. a well-defined problem</i>
<i>Note: The correct response is in italics</i>	

Figure 3.2: Examples of Questions Used in the Retrieval Practice Phase and the Final Test Phase

3.4.4 Distractor Task Phase

The word search puzzle was used as the distractor task. The level of difficulty of the puzzle was moderate and suitable for university students. The function of this task was to

clear participants' working memory and to prevent them from mentally rehearsing the studied material.

3.5 Data Collection Procedure

The experiment followed a mixed-subjects design in which the participants were assigned into two retention interval groups (i.e., 50 participants in the short retention interval group and 50 participants in the long retention interval group). In each group, participants were exposed to all the conditions (retrieval practice phase and levels of thinking). The experiment was conducted synchronously in two groups whereby all the participants in each group took part in the experimental procedures at the same time. All participants were given a consent form for them to sign as an indication that they agreed to participate in this study. Participants were informed that they could withdraw from the study before, during, and after the experiment for any reason. Given that the experiment was conducted via online, all participants were given an Integrity Declaration Form for them to sign as an indication that they would not refer to sources from the Internet or use other assistance during the experiment. Once they had given their consent and filled in the Integrity Declaration Form, the participants were briefed on what they were required to do. After the briefing, the procedures for the experiment took place. These involved five phases: study phase, first distractor task, retrieval practice phase, second distractor task and final test phase. For the study phase, participants were asked to read one reading passage on Problem Solving (a chapter in the Cognitive Psychology course content) which was presented on screen using Google Doc. They were informed that they would be tested on the information. However, they were not given the details of the test. The participants were given 15 minutes to read the passage. Once the 15-minute time limit was over, participants were asked to solve a word search puzzle task in the first distractor task phase for one minute. The purpose of this phase

was to clear out the participant's working memory. After the one-minute distractor task finished, the next phase, the retrieval practice phase involving the manipulation of the retrieval practice phase (production test and recognition test) took place. Participants were asked to answer 10 recognition test questions (MCQs) and 10 production test questions (short-answer questions) on the reading passage. The questions were presented using Google Form Quiz. Each of the 10 questions consisted of five lower-order thinking questions and five higher-order thinking questions. For this phase, participants were required to answer a total of 20 questions within 15 minutes. Then, the participants were asked to undertake the second distractor task in which they had to solve another word search puzzle for one minute. Once the one-minute time limit for the task was up, the next phase was the final test phase. Similar to the retrieval practice phase, the questions in the final test phase were also presented using Google Form Quiz. Participants in the short retention interval group were tested immediately after the second distractor task whereas those in the long retention interval group were tested after a delay of three weeks. All participants had to answer a total of 20 MCQs for the final test phase. After completing the final test phase, the participants were debriefed and allowed to dismiss. The whole experiment took approximately 60 minutes.

3.6 Data Analysis

The data that were collected were analysed using Statistical Package for Social Sciences (SPSS) which used three-way mixed ANOVA. Following the procedure of most studies which examined retrieval practice (Cranney et al., 2009; Carpenter et al., 2016; Stenlund et al., 2016; Dobson et al., 2018; Thomas et al., 2018; Agarwal, 2019), the statistical test was used to analyse the data collected in the final test phase to measure the participants' performance based on the different types of retrieval practice conditions

administered to them. The three-way mixed ANOVA is used when measuring three independent variables in which the same participants are used to manipulate one or more independent variables, but different participants are used to manipulate the other or more independent variables. In the present study, retention interval was varied between participants whereas retrieval practice condition and levels of thinking were varied within participants. The main research objective was to examine the effect of different types of retrieval practice (production test and recognition test) on reading comprehension across levels of thinking (higher-order thinking and lower-order thinking) and retention interval (short retention interval and long retention interval). By performing this analysis, the present study aimed to answer the research questions.

3.7 Chapter Summary

This chapter has explained the method used in conducting the research. It includes the description of the participants involved in the study, the research design and also describes the research materials used in conducting the experiment. This chapter has also outlined the data collection procedure as well as provided a description of the method of data analysis.

CHAPTER 4

RESULTS

4.1 Introduction

Chapter Four presents the results obtained from the experiment that was conducted for the research. This chapter begins by presenting the demographic analysis of the research, followed by normality and homogeneity test. A test for normality and homogeneity was conducted to ensure the assumptions for ANOVA were fulfilled beforehand. A three-way mixed ANOVA statistical test was then used to analyse the collected data, mainly focusing on the final test performance to measure the participants' performance based on the different types of retrieval practice conditions administered to them. The statistical test was used to find out the difference in students reading comprehension across the different types of retrieval practice, the difference in students' reading comprehension between the different levels of thinking and the difference in students' reading comprehension between the different types of retention interval. The three-way mixed ANOVA was also used to examine the interaction between the types of retrieval practice and levels of thinking, the interaction between the types of retrieval practice and retention interval, the interaction between retention interval and levels of thinking as well as the interaction between the types of retrieval practice, levels of thinking and retention interval.

4.2 Demographic Analysis

Table 4.1 presents the demographic analysis of the present study which includes gender, course programme and MUET results of the participants. The table shows that among the 100 participants, 78 of them were female students whereas 22 of them were male students. In terms of the course programme, 16 participants were enrolled in the Psychology

course programme, 10 participants were from the Counselling programme, 28 students were from the Cognitive Science programme and the remaining 46 participants were enrolled in Human Resource Development course programme. As for the MUET results, it was mentioned that those with MUET Band 3 and above were selected to participate in the study. Among the 100 participants, 52 of them were those who got Band 3, 39 of them were those who got Band 4 and the remaining nine participants were those of Band 5.

Table 4.1: Demographic Analysis

		Number of participants	%
Gender	Female	78	78.0%
	Male	22	22.0%
Course Programme	Psychology	16	16.0%
	Counselling	10	10.0%
	Cognitive Science	28	28.0%
	Human Resource Development	46	46.0%
MUET	Band 3	52	52.0%
	Band 4	39	39.0%
	Band 5	9	9.0%

4.3 Tests for Normality and Homogeneity

Tests for normality and homogeneity were conducted to check whether the data were drawn from a normally distributed population and to see if the data collected were homogenous. Based on the histograms as shown in Figure 4.1 (short retention interval) and Figure 4.2 (long retention interval), the shape of the graphs shows an approximate bell-shaped, thus indicating that the data collected were from an approximately normal

population which meets the assumption of ANOVA. As for the homogeneity test, Table 4.2 shows the results from the Levene’s Test for Homogeneity of Variance. The results in the table show that for all the groups, the significance values were greater than .05 which indicates that the variances were homogenous. In other words, the values for all the levels of the within-subject group (retrieval practice and levels of thinking) were more or less similar for both short and long retention interval group. The results for the tests of normality and homogeneity therefore suggest that the assumptions for ANOVA have been fulfilled.

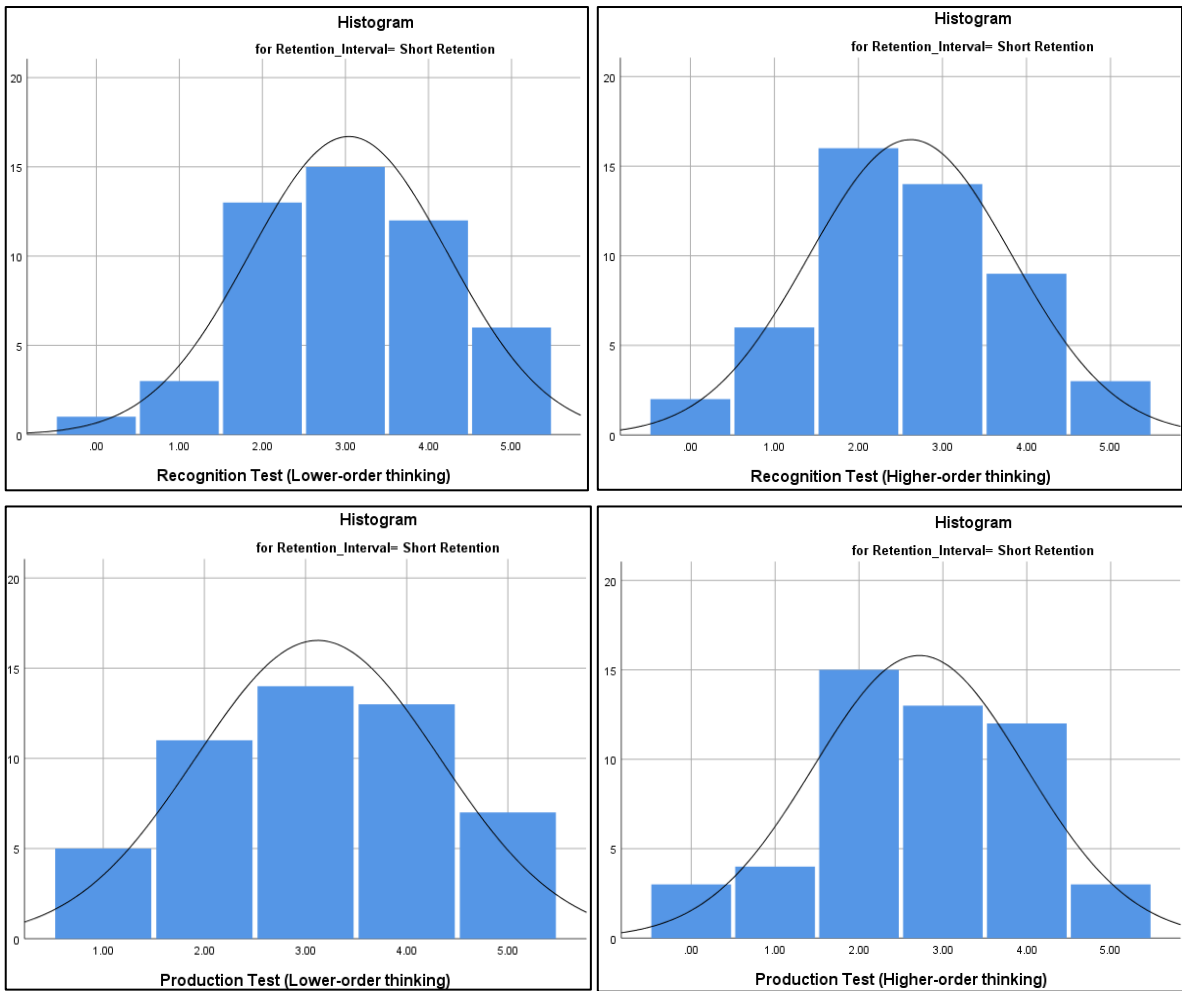


Figure 4.1: Histograms for the Short Retention Interval Group

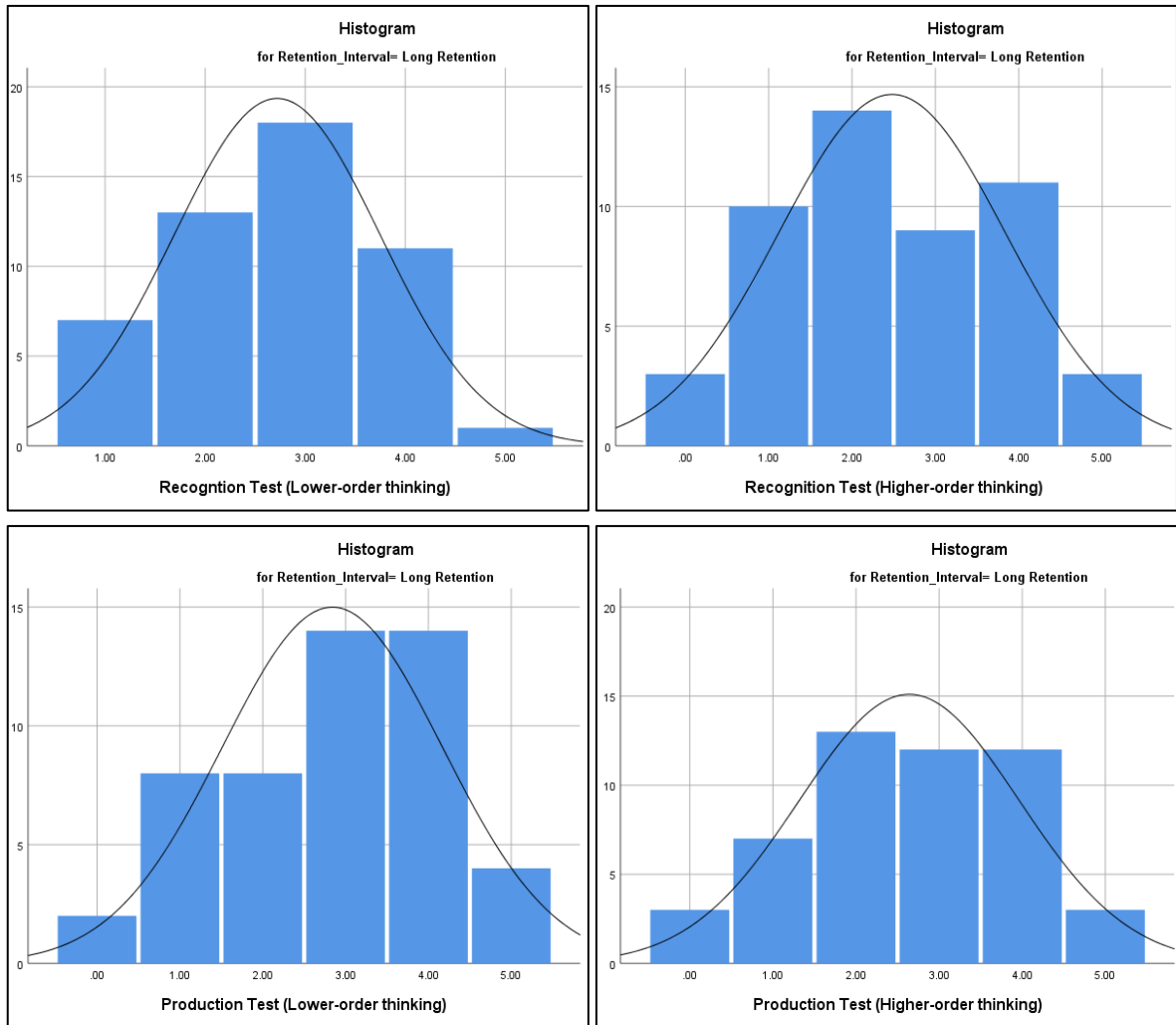


Figure 4.2: Histograms for the Long Retention Interval Group

Table 4.2: Test for Homogeneity of Variance

		Levene Statistic	df1	df2	Sig.
Recognition Test (Lower-order thinking)	Based on Mean	.330	1	98	.567
Recognition Test (Higher-order thinking)	Based on Mean	1.430	1	98	.235
Production Test (Lower-order thinking)	Based on Mean	.491	1	98	.485
Production Test (Higher-order thinking)	Based on Mean	.284	1	98	.595

4.4 Findings From the Three-Way Mixed ANOVA on Final Test Performance

Table 4.3: Three-Way Mixed ANOVA Results on Within-Subject Variables

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Retrieval Practice	Sphericity Assumed	1.322	1	1.322	1.311	.255
	Greenhouse-Geisser	1.322	1.000	1.322	1.311	.255
	Huynh-Feldt	1.322	1.000	1.322	1.311	.255
	Lower-bound	1.322	1.000	1.322	1.311	.255
Retrieval Practice * Retention Interval	Sphericity Assumed	.063	1	.063	.062	.804
	Greenhouse-Geisser	.063	1.000	.063	.062	.804
	Huynh-Feldt	.063	1.000	.063	.062	.804
	Lower-bound	.063	1.000	.063	.062	.804

Table 4.3 continued

Error (Retrieval Practice)	Sphericity Assumed	98.865	98	1.009		
	Greenhouse- Geisser	98.865	98.000	1.009		
	Huynh-Feldt	98.865	98.000	1.009		
	Lower-bound	98.865	98.000	1.009		
Levels of Thinking	Sphericity Assumed	9.923	1	9.923	9.880	.002
	Greenhouse- Geisser	9.923	1.000	9.923	9.880	.002
	Huynh-Feldt	9.923	1.000	9.923	9.880	.002
	Lower-bound	9.923	1.000	9.923	9.880	.002
Levels of Thinking * Retention Interval	Sphericity Assumed	.903	1	.903	.899	.345
	Greenhouse- Geisser	.903	1.000	.903	.899	.345
	Huynh-Feldt	.903	1.000	.903	.899	.345
	Lower-bound	.903	1.000	.903	.899	.345
Error (Levels of Thinking)	Sphericity Assumed	98.425	98	1.004		
	Greenhouse- Geisser	98.425	98.000	1.004		
	Huynh-Feldt	98.425	98.000	1.004		
	Lower-bound	98.425	98.000	1.004		
Retrieval Practice * Levels of Thinking	Sphericity Assumed	.023	1	.023	.024	.877
	Greenhouse- Geisser	.023	1.000	.023	.024	.877
	Huynh-Feldt	.023	1.000	.023	.024	.877
	Lower-bound	.023	1.000	.023	.024	.877

Table 4.3 continued

Retrieval Practice *	Sphericity	.003	1	.003	.003	.959
Levels of Thinking *	Assumed					
Retention	Greenhouse-Geisser	.003	1.000	.003	.003	.959
Interval	Huynh-Feldt	.003	1.000	.003	.003	.959
	Lower-bound	.003	1.000	.003	.003	.959
Error (Retrieval Practice *	Sphericity	91.225	98	.931		
Levels of Thinking)	Assumed					
	Greenhouse-Geisser	91.225	98.000	.931		
	Huynh-Feldt	91.225	98.000	.931		
	Lower-bound	91.225	98.000	.931		

Table 4.4: Three-Way Mixed ANOVA Results on Between-Subject Variable

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	3074.702	1	3074.702	949.506	.000
Retention Interval	4.202	1	4.202	1.298	.257
Error	317.345	98	3.238		

4.4.1 Students' Reading Comprehension Across Different Types of Retrieval Practice

The results in Table 4.3 show that there was no significant difference in students' reading comprehension across the different types of retrieval practice (production test versus recognition test), $F(1, 98) = 1.31$, $p = .255$. This indicates that the students' performance in reading comprehension was almost similar regardless of the types of retrieval practice used during the learning phase. Although the difference in reading comprehension across different

types of retrieval practice was not significant, the mean scores of students using the production test ($M = 2.83$) were slightly higher than that of the recognition test ($M = 2.72$), indicating that students must have benefitted a bit more from the production test compared to the recognition test.

4.4.2 Students' Reading Comprehension Between Lower-Order Thinking and Higher-Order Thinking

The results from the statistical analysis shown in Table 4.3 reveal that there was a significant difference in students' reading comprehension between lower-order thinking and higher-order thinking, $F(1, 98) = 9.88, p = .002$. This shows that by using retrieval practice during the retrieval practice phase regardless of the formats or types used, students' performance in reading comprehension was significantly better for lower-order thinking questions ($M = 2.93$) relative to higher-order thinking questions ($M = 2.62$).

4.4.3 Students' Reading Comprehension Between Different Types of Retention Interval

The results showed that there was no significant difference in students' reading comprehension between the different types of retention interval (short retention interval versus long retention interval), $F(1, 98) = 1.30, p = .257$ (Table 4.4). This suggests that students' performances in reading comprehension between the two groups were no different from each other. Based on the mean scores however, the students' reading comprehension scores in the short retention interval were slightly higher ($M = 2.88$) as compared to the performance in the long retention interval ($M = 2.67$).

4.4.4 Interaction Effect Between Types of Retrieval Practice and Levels of Thinking

The results in Table 4.3 shows that there was no significant interaction between types of retrieval practice and levels of thinking, $F(1, 98) = 0.02, p = .877$, signifying that the effect

of different types of retrieval practice on reading comprehension was almost similar for both lower- and higher-order thinking questions. Despite no significant interaction between the two variables was observed, the mean scores showed that test performance for lower-order thinking questions were slightly better when using the production test during the retrieval practice phase ($M = 2.98$) compared to using recognition test ($M = 2.88$). Similarly, for the higher-order thinking questions, the mean score for the production test was slightly higher ($M = 2.68$) than that of the recognition test ($M = 2.55$). This finding suggests that for both lower- and higher-order thinking questions, students may have benefited a bit more when they used the production test as the retrieval practice format compared to using the recognition test.

4.4.5 Interaction Effect Between Types of Retrieval Practice and Retention Interval

The results reveals that there was no significant interaction between types of retrieval practice and retention interval, $F(1, 98) = 0.06$, $p = .804$ (Table 4.3), indicating that the effect of different types of retrieval practice on reading comprehension was almost similar for both short and long retention intervals. While there was no significant interaction effect between types of retrieval practice and retention interval, for those in the short retention interval group, the mean score for the production test was a tad higher ($M = 2.92$) than that of recognition test ($M = 2.83$). Likewise, for those in the long retention interval group, the mean score for production test was slightly higher ($M = 2.74$) as compared to recognition test ($M = 2.60$). Such finding indicates that using production test as the retrieval practice format may have provided an added advantage compared to using recognition test for both short- and long retention interval.

4.4.6 Interaction Effect Between Retention Interval and Levels of Thinking

As for the interaction effect between retention interval and levels of thinking, the results in Table 4.3 show that there was no significant retention interval and levels of thinking interaction, $F(1, 98) = 0.90, p = .345$, implying that the effect of different retention intervals on reading comprehension was equivalent for both lower- and higher-order thinking questions. Although no significant interaction effect was observed, the mean values revealed that for the lower-order thinking questions, those in the short retention interval group scored slightly higher ($M = 3.08$) than those who were in the long retention group ($M = 2.78$). The same applies for the higher-order thinking questions in which those in the short retention interval group fared better ($M = 2.67$) compared to those in the long retention interval group ($M = 2.56$). This indicates that the reading comprehension performance among students in both lower- and higher-order thinking was slightly better when the effect was observed under immediate testing compared to delayed testing.

4.4.7 Interaction Effect Between Types of Retrieval Practice, Levels of Thinking and Retention Interval

In terms of the interaction effect between types of retrieval practice, levels of thinking and retention interval, the results in Table 4.3 showed that there was no significant interaction between these three variables, $F(1, 98) = 0.003, p = .959$. This shows that the effects of the types of retrieval practice and levels of thinking on student's reading comprehension were similar in both short and long retention intervals. Even though no significant interaction was observed, the mean scores in Table 4.5 shows that regardless of the retrieval formats used, students in both the short and long retention interval group perform slightly better in lower-order thinking questions ($M = 3.04, M = 3.12, M = 2.72, M = 2.84$) compared to higher-order thinking questions ($M = 2.62, M = 2.72, M = 2.48, M =$

2.64). Also, with regard to the scores from the retrieval practice groups as shown in Table 4.5, it can be seen that the students seemed to collectively perform slightly better when engaging in the production test compared to the recognition test. In addition, by comparing the scores from the two groups (short retention interval and long retention interval), it can be seen that those in the short retention interval group scored slightly higher than those in the long retention interval group.

Table 4.5: Means and Standard Deviations of Interaction Between Types of Retrieval Practice, Levels of Thinking and Retention Interval

Retention Interval	Types of Retrieval Practice	Levels of Thinking	Mean	Standard Deviation
Short retention interval	Recognition test	Lower-order thinking	3.04	1.19
		Higher-order thinking	2.62	1.21
	Production test	Lower-order thinking	3.12	1.21
		Higher-order thinking	2.72	1.26
Long retention interval	Recognition test	Lower-order thinking	2.72	1.03
		Higher-order thinking	2.48	1.36
	Production test	Lower-order thinking	2.84	1.33
		Higher-order thinking	2.64	1.32

4.5 Chapter Summary

This chapter presents the results of the experiment which used three-way mixed ANOVA in analysing the data. The results revealed that there was no significant difference

in students' reading comprehension across the different types of retrieval practice and no observed difference in students' reading comprehension between the different types of retention interval. There was, however, a significant difference observed in students' reading comprehension between the different levels of thinking. In terms of the interaction effect on the other hand, no significant interaction effect was observed between the types of retrieval practice and levels of thinking, no significant interaction between the types of retrieval practice and retention interval as well as no significant interaction between retention interval and levels of thinking. There was also no significant interaction effect between the types of retrieval practice, levels of thinking and retention interval.

CHAPTER 5

DISCUSSION AND CONCLUSION

5.1 Introduction

Chapter Five presents the discussion based on the results obtained from the statistical analysis conducted on the data collected as well as the conclusion of the research. It begins by addressing the difference in students reading comprehension across the different types of retrieval practice, followed by the difference in students' reading comprehension between the different levels of thinking and then the difference in students' reading comprehension between the different types of retention interval. This chapter also discusses the interaction between the types of retrieval practice and levels of thinking, the interaction between the types of retrieval practice and retention interval, the interaction between retention interval and levels of thinking as well as the interaction between the types of retrieval practice, levels of thinking and retention interval. In addition to that, this chapter summarises the research by explaining the theoretical and practical implications of the study based on the findings obtained, followed by addressing the limitations faced in the present study. This chapter also includes the suggestions for future research whereby it proposes a few ideas on how to extend the present study

5.2 Effect of Different Types of Retrieval Practice on Students' Reading Comprehension

With regard to the effect of different types of retrieval practice on reading comprehension, the statistical analysis revealed that there was no significant difference in students' reading comprehension across the different types of retrieval practice (production test versus recognition test), indicating that the students' performance in reading

comprehension using the recognition test (MCQ) was almost similar to the performance when using the production test (short-answer questions) in the retrieval practice phase. This finding is in line with a few studies which found that short-answer and MCQs are equally effective in enhancing the retention of studied materials (McDermott et al., 2014; Smith & Karpicke, 2014). Specifically, Smith and Karpicke's (2014) study which investigated the effects of different question formats on learning (short-answer, MCQ or hybrid questions) found that there was little to no difference in students' performance across the three different formats. They stated that their short-answer questions were more difficult than MCQs however, participants in their study seemed to perform better on the MCQs most of the time compared to the short-answer questions. The researchers reasoned that although a more difficult test is said to be more advantageous, the success of retrieval also plays an important role in enhancing subsequent performance. In other words, engaging in a retrieval practice format which produces successful retrievals contribute to just many positive effects as using difficult and effortful ones and that, not only it has to be difficult, but it also has to produce successful retrievals for any beneficial effects to follow suit. Thus, following their explanation, the findings from the present study suggest that the successful retrievals obtained from the recognition test produced a retrieval practice effect equivalent to using the production test that requires more retrieval effort. This can be seen from the mean scores whereby the mean scores from the recognition test ($M = 2.72$) and the production test ($M = 2.83$) were not very much different from each other, hence showing that performances in students' reading comprehension were almost similar for both types of retrieval practice. Additionally, a possible reason for the absence of significant differences in the reading comprehension across the different types of retrieval practice could be due to the unfamiliarity with the topic of the reading comprehension passage among participants. The

learned topic which was new to the participants might have introduced some difficulty to them and it seems that when they are unfamiliar with the reading topic, neither production nor recognition test contributed any significant effect in enhancing their reading comprehension.

Another plausible reason for why no significant difference was observed in students' reading comprehension across different types of retrieval practice could be that the recognition test questions in the present study included competitive alternatives. According to Little et al.'s (2012) study, in order for benefits to occur when taking MCQ tests, it must be properly constructed, that is, they must incorporate competitive incorrect alternatives which can trigger the retrieval processes. It triggers retrieval processes by encouraging students to recall why the correct alternatives were correct and why the incorrect alternatives were incorrect, which in turn will facilitate future recall of the information if it is tested later, such as during final exams. Following Little et al.'s (2012) study which showed that well-constructed MCQs (i.e., questions which include competitive incorrect choices of answers) are also capable of producing productive retrieval processes; therefore, it was likely that the reading comprehension performance when using the recognition test in the present study was on par with using the production test because the recognition test questions were almost as difficult as those used in the production test. An example of competitive multiple-choice items can be seen in Figure 5.1—they consist of distractors that were plausible and share important information with the correct answer. All of the choices shared one similar information in which all of the items are descriptions of the different types of problems, however option "C" particularly refers to insight problems.

Recognition Test (Question 3)

All of the following describe ill-defined problems **EXCEPT**

- A Usually occurs in our daily life.
- B The path leading to the solution may vary.
- C *Solutions come to the individuals all of a sudden.*
- D Does not necessarily have one right answer.

Figure 5.1: An Example of the Recognition Test Question

5.3 Effect of Levels of Thinking on Students' Reading Comprehension

Regarding the effect of levels of thinking on reading comprehension, the present study found a significant difference in students' reading comprehension between lower-order thinking and higher-order thinking, in which the students' performance in reading comprehension was significantly better for lower-order thinking questions compared to higher-order thinking questions. One plausible explanation for why the mean scores for lower-order thinking questions appear to be higher ($M = 2.93$) than the mean scores for higher-order thinking questions ($M = 2.62$) could be due to the fact that the students have not yet mastered the elements of higher-order levels in the particular topic, given that they were unfamiliar with the topic and had yet to study the topic thoroughly in class. Thus, some information in the reading passage may have seemed new to them. According to Bloom's taxonomy, in order to reach the higher category, the individual has to acquire the cognitive processes at the lower category first (Agarwal, 2019). In other words, the levels of thinking according to Bloom's taxonomy are said to follow a hierarchy (Bloom et al., 1956; Dobson et al., 2018). Hence, it was possible that the students were only able to master the lower-

level elements during the retrieval practice phase, resulting in marginally better performance for lower-order thinking questions in comparison to higher-order thinking questions.

5.4 Effect of Different Types of Retention Interval on Students' Reading Comprehension

In terms of the effect of retention interval on reading comprehension, the results showed that the difference in students' reading comprehension between the different types of retention interval (short retention interval versus long retention interval) was not statistically significant, implying that students' performances in reading comprehension between the two groups were equivalent to each other regardless of the types of retrieval practice format and levels of thinking. This finding is in contrast with some previous studies which found significant differences between different retention intervals—immediate testing and delayed testing (Wiklund-Hörnqvist et al., 2014; Ferreira et al., 2019). In Wiklund-Hörnqvist et al.'s (2014) study, they observed the effect by comparing a five-minute time delay (immediate test) and an 18-day delay (delayed test) whereas for Ferreira et al.'s (2019) study, they observed the effect by comparing an immediate test (immediately after the second learning session) and a delayed test (one day after the second learning session). As for the present study, the immediate test was done five minutes after the distractor task whereas the delayed test was done after a delay of three weeks (21 days). Perhaps when there is a large time gap between the short retention interval and long retention interval, the difference between both retention intervals was not so evident. Based on the mean scores however, it can be seen that students' reading comprehension mean scores in the short retention interval group were slightly higher ($M = 2.88$) than the scores from those in the long retention interval group ($M = 2.67$), indicating that the students in the short retention interval group (immediate testing) were able to recall slightly more information than those

in the long retention interval group (delayed testing). This seems to provide some support for the findings by Ferreira et al. (2019) which revealed that those who took an immediate test outperformed those who took a delayed test. One possible reason for such a finding in the present study to occur was likely due to the fact that some students in the long retention interval group may have forgotten what they had learnt prior to the final test which took place after a delay of three weeks.

5.5 Interaction Effect Between Types of Retrieval Practice and Levels of Thinking

In terms of interaction level, the present study found no significant interaction between types of retrieval practice and levels of thinking, implying that the effect of different types of retrieval practice on reading comprehension was almost similar for both lower- and higher-order thinking questions. This is possibly because of the learned topic of the reading comprehension passage which was new and unfamiliar to the participants. It seems that when the participants were unacquainted with the reading passage topic, neither production nor recognition test contributed a significant effect in facilitating their reading comprehension performance for both lower- and higher-order thinking questions. However, the mean scores showed that for lower-order thinking questions, students scored slightly higher when using the production test during the retrieval practice phase ($M = 2.98$) compared to when using the recognition test ($M = 2.88$). Likewise, for the higher-order thinking questions, the mean score for the production test was slightly higher ($M = 2.68$) than that of the recognition test ($M = 2.55$). This therefore suggests that production test (short-answer questions) may have provided an added advantage over the recognition test (MCQ) on different levels of thinking. A growing number of studies have found that short-answer tests elicit a greater retrieval practice effect than MCQs (Butler & Roediger, 2007; Pyc & Rawson, 2009; Greiving & Richter, 2018) because the former requires more retrieval effort compared to the latter. Such

finding is supported by the retrieval effort hypothesis that is derived from the desirable difficulty framework which asserts that when an individual engages in a more difficult and effortful retrieval process, it produces greater benefits in memory retention of learned materials (Gardiner et al., 1973; Pyc & Rawson, 2009) thus resulting in better performance when engaging in short-answer questions relative to engaging in MCQs as a learning strategy.

5.6 Interaction Effect Between Types of Retrieval Practice and Retention Interval

Furthermore, there was also no significant interaction observed between types of retrieval practice and retention interval. This indicates that the effect of different types of retrieval practice on reading comprehension was almost similar in both short and long retention intervals which was possibly due to the similarities in the difficulty of both test formats. Based on the study by Little et al. (2012), it was revealed that when a well-constructed recognition test is used—one which includes plausible, competitive alternatives, it can trigger the retrieval process by encouraging learners to recall why the correct alternatives were correct and why the incorrect alternatives were incorrect. Given that in the present study the answer choices for the recognition test questions included items that were competitive, hence it was likely that the recognition test was almost as difficult and required similar retrieval effort to the production test. Looking at the mean scores on the other hand, for both the short and long retention intervals groups, the students seemed to perform slightly better when they used the production test ($M = 2.92$ and $M = 2.74$ respectively) instead of using the recognition test ($M = 2.83$ and $M = 2.60$ respectively) in the retrieval practice phase. In the present study, the production test involved the production of answers which required more effort than the recognition test. This finding seems to provide some support to a few other studies which found that both short- and long-term retrieval practice effects

are present when the short-answer test format is used (Kang et al., 2007; McDaniel, Anderson, et al., 2007; Stenlund et al., 2014). Similar to the reasoning mentioned previously, the production test, particularly the short-answer test involves the production of answers in which more retrieval effort is required than in the recognition test (MCQ). Therefore, following the desirable difficulty framework, when learners engage in a more difficult but successful retrieval process, it leads to more beneficial effects on memory compared to easier, successful retrievals (Pyc & Rawson, 2009).

5.7 Interaction Effect Between Retention Interval and Levels of Thinking

The present study also observed no significant interaction effect between retention interval and levels of thinking, suggesting that the effect of different retention interval on reading comprehension was equivalent for both lower- and higher-order thinking. This finding could probably be due to the large time difference between the short retention interval and long retention interval (i.e., five minutes and three weeks respectively) as well as the participants' unfamiliarity with the learned items whereby they had to read about a topic which they had not explicitly learned in class yet. These possible reasons therefore explain why the effect of short and long retention intervals on students' reading comprehension was similar for both lower- and higher-order thinking questions. Despite no significant interaction effect between the two variables, the mean scores showed that for lower-order thinking questions, the students scored slightly higher in their reading comprehension when the effect was observed under the short retention interval ($M = 3.08$) compared to when it was observed under the long retention interval ($M = 2.78$). Similarly for the higher-order thinking questions, those in the short retention interval group slightly outperformed ($M = 2.67$) those in the long retention interval group ($M = 2.56$). These findings which revealed that those in the short retention interval group scored slightly better

than those in the long retention interval group were likely because the students in the long retention interval group had forgotten some of the information learnt before they took the final test. The reason an individual forgets is because their newly formed memories have yet to have the opportunity to be strengthened, therefore these memories are susceptible to the interruption of mental activity and additional mental information (Baddeley & Hitch, 1993; Ferreira et al., 2019). Moreover, the slight difference between the mean scores of those in the short and long retention interval groups can be explained by the recency effect which posits that when one takes a test and applies the information learnt immediately, there is a high chance of remembering that information than in the situation where the information is presented the day before (Ferreira et al., 2019). Also, while it is true that retrieval practice can improve long-term memory retention in comparison to restudying or rereading (Carpenter et al., 2006; Karpicke & Roediger, 2007a; McDaniel, Roediger, et al., 2007), when the retrieval practice effect is concerned, it yields more positive effects when it is observed in the short-term period (Roediger & Karpicke, 2006; Karpicke, 2017; Ferreira et al., 2019). Future research perhaps can further explore the interaction between retention interval and levels of thinking on reading comprehension in the context of using retrieval practice.

5.8 Interaction Effect Between Types of Retrieval Practice, Levels of Thinking and Retention Interval

With regard to the interaction between the types of retrieval practice, levels of thinking and retention interval, the present study found no significant interaction between these three variables. This finding indicates that the effects of the types of retrieval practice and levels of thinking on student's reading comprehension was no different in both short and long retention intervals. Despite no significant interaction observed between the types of

retrieval practice, levels of thinking and retention interval, the mean scores reported in Table 4.5 show that regardless of the retrieval practice formats used, students in both short and long retention interval groups seemed to perform slightly better in lower-order thinking questions than in higher-order thinking questions. This finding can be attributed to the popular belief that factual knowledge must go before skills (Bruner, 1977; Willingham, 2009). Given the fact that they had not yet learnt the topic explicitly in class and that some information in the reading passage may seemed novel to them, the students only managed to acquire the cognitive processes in the lower category during the course of the experiment, which resulted in slightly better performance for lower-order thinking questions. Additionally, by comparing the scores between different types of retrieval practice group for both levels of thinking and retention interval groups, students seemed to fare a bit better when using the production test in the retrieval practice phase compared to using the recognition test, suggesting that using short-answer tests leads to slightly more successful retrievals than when using the multiple-choice test. Several studies in the past have found that the production test, particularly short-answer questions have more benefits than MCQs because it requires a more effortful retrieval than the latter (McDaniel, Anderson, et al., 2007; Stenlund et al., 2016; Greving & Richter, 2018). Among the existing theories that are related to explaining the retrieval practice effect, the retrieval effort hypothesis that is driven by the desirable difficulty framework seems to be more suitable for justifying the present finding because it compares the retrieval effort required for each type of test format. Therefore, following the desirable difficulty framework which states that more challenging but successful retrievals are more beneficial for memory retention as opposed to easier, successful retrievals (Pyc & Rawson, 2009), it was of no surprise that the students seemed to gain slightly more benefits when they used the production test during the retrieval practice

phase which led to slightly better performance in their reading comprehension. Apart from that, by comparing the scores from the two groups (short retention interval and long retention interval), it can also be seen that those in the short retention interval group scored slightly higher than those in the long retention interval group, signifying that there was a slightly greater immediate testing effect compared to delayed testing effect in both different types of retrieval practice and levels of thinking. As mentioned earlier, one possible explanation behind such finding could be due to the recency effect in a situation when a student takes a test immediately after learning, there is a high possibility of them remembering the recently learnt information more than the information presented a day earlier (Ferreira et al., 2019).

5.9 Theoretical and Practical Implications

In conclusion, the present study extends previous findings on retrieval practice in the classroom setting, particularly examining the effects of different types of retrieval practice across levels of thinking and retention interval on reading comprehension. The present study found no significant difference in reading comprehension across the different types of retrieval practice, indicating that using the recognition test was equally as effective as using the production test in enhancing memory retention of learned items. This finding also contributed to the body of knowledge in which it is suggested that learner's reading comprehension performance when using recognition test as a learning strategy, particularly a properly constructed one with competitive alternatives was equivalent to when using the production test. Besides that, the difference in reading comprehension between the different types of retention interval also was not observed, implying that students in the short retention interval group retained just as much information as those in the long retention interval group. In addition, the present study observed a significant difference in students' reading comprehension between the different levels of thinking, in which the students' performance

in reading comprehension was significantly better for lower-order thinking questions than higher-order thinking questions. This suggests that even by using retrieval practice, it was not surprising to find that when students learn new items, they master the lower levels of thinking in Bloom's Taxonomy before mastering those of higher levels.

5.10 Limitations of the Present Research

The present study aims to examine the effects of different types of retrieval practice on reading comprehension across levels of thinking and retention interval among undergraduates. The initial research plan for data collection was to conduct the normal laboratory experiment where recruited participants from a Cognitive Psychology class arrive in a small group of 4 during their respective slots to do the experiment. However, due to the Covid-19 pandemic and subsequent restrictions introduced, all class lectures were held online, and hence, the researcher also had to change the initial plan to conducting the experiment online. When it comes to using the online method in conducting an experiment, the difficulty is in monitoring the participants particularly in ensuring that they will not cheat or refer to any other materials when answering the questions. Therefore, to address this matter, a method that was used in examinations was adopted in the present study. This entailed an Integrity Declaration Form (see Appendix E) which was issued to each participant before the experiment began to ensure that they were well-versed in what they should and should not do during the course of the experiment. This form was also given to make sure that all participants acknowledged that the present study required their honest participation and that they agreed to abide by the rules during the experiment.

Another limitation of the present research was the sample size. Larger sample size is more likely to yield results that can be extended to the larger population and should be used

in future research. The initial plan for this study was to recruit more participants who would be tested using a traditional experimental approach, however, due to the pandemic constraints, the researcher had to resolve to conduct this experiment using an online approach, which had to be limited to two classes due to class scheduling issues and had to select further only 100 participants out of the 201 who met MUET requirements. Furthermore, because this study used participants from two existing Cognitive Psychology classes, randomisation was not used, which limits the study's ability to conclude a causal relationship between the independent variable and dependent variable.

5.11 Suggestions for Future Research

Future work can extend the present study by particularly observing the transfer of knowledge between different levels of thinking using different types of retrieval practice format and examining the effect on different retention intervals. The research on transfer of knowledge has been the focus of many recent studies (Dobson et al., 2018; Agarwal, 2019) to investigate whether transfer of knowledge from one level to another level (e.g., lower-level to higher-level learning or vice versa) would occur if students use retrieval practice as the learning strategy. Therefore, it will be noteworthy to know if the transfer of knowledge is reliant on the type of retrieval practice format used and to see whether the effect can be retained for longer period of time. As mentioned in the discussion section, future studies can further examine the interaction between retention interval and levels of thinking on reading comprehension particularly in the context of using retrieval practice. There have been few studies that have particularly observed the effect of different retention intervals and not much is known pertaining to this matter. Therefore, it will be interesting to further explore this issue. Likewise, future work can also extend the present study by including a student reading anxiety scale in the observation. A recent study by Rahmat et al. (2020) which examined the

influence of students' fear and perceived difficulties in academic reading found that teaching method was one of the reasons undergraduate students fear reading and that they are most anxious when they have difficulty in understanding the content of the text they read. The study highlighted the possibility of overcoming the fear of reading by using good teaching methods of reading strategies. Thus, it would be interesting to know whether employing retrieval practice as a teaching and learning strategy in the classroom will contribute to any difference in students' reading anxiety.

REFERENCES

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87(3), 659–701. <https://doi.org/10.3102%2F0034654316689306>
- Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology*, 111(2), 189–209. <https://doi.org/10.1037/edu0000282>
- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology*, 22(7), 861–876. <https://doi.org/10.1002/acp.1391>
- Amin, H., & Malik, A. S. (2013). Human memory retention and recall processes. *Neurosciences*, 18(4), 330–344.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (abridged ed.). Addison Wesley Longman.
- American Psychological Association. (n.d.). *APA Dictionary of Psychology*. American Psychological Association. Retrieved November 20, 2021, from <https://dictionary.apa.org/retention-interval>
- Baddeley, A. D. & Hitch, G. (1993). The recency effect: Implicit learning with explicit retrieval? *Memory & Cognition*, 21(2), 146–155. <https://doi.org/10.3758/BF03202726>
- Bae, C. L., Theriault, D. J., & Redifer, J. L. (2019). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction*, 60, 206–214.

<https://doi.org/10.1016/j.learninstruc.2017.12.008>

Bekkink, M. O., Donders, R., van Muijen, G. N., & Ruiter, D. J. (2012). Challenging medical students with an interim assessment: A positive effect on formal examination score in a randomized controlled study. *Advances in Health Sciences Education*, 17(1), 27–37.

<https://doi.org/10.1007/s10459-011-9291-6>

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123–144). Lawrence Erlbaum Associates.

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

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). David McKay Company.

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106(3), 849–858. <https://doi.org/10.1037/a0035934>

Brame, C. J., & Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science course. *CBE—Life Sciences Education*, 14(2), 1–12. <https://doi.org/10.1187/cbe.14-11-0208>

Bruner, J. S. (1977). *The process of education*. Harvard University Press.

Burdo, J., & O'Dwyer, L. (2015). The effectiveness of concept mapping and retrieval practice as learning strategies in an undergraduate physiology course. *Advances in Physiology Education*, 39(4), 335–340. <https://doi.org/10.1152/advan.00041.2015>

- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1133. <https://psycnet.apa.org/doi/10.1037/a0019902>
- Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4-5), 514–527. <https://doi.org/10.1080/09541440701326097>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569. <https://doi.org/10.1037/a0017021>
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*, 19, 619–636.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34, 268–276. <https://doi.org/10.3758/BF03193405>
- Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory and Language*, 92, 128–141. <https://doi.org/10.1016/j.jml.2016.06.008>
- Carpenter, S. K., Lund, T. J., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review*, 28(2), 353–375. <https://doi.org/10.1007/s10648-015-9311-9>
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of US history facts. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 23(6), 760–771.

<https://doi.org/10.1002/acp.1507>

- Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review*, 13(5), 826–830. <https://doi.org/10.3758/BF03194004>
- Carpenter, S. K., Pashler, H., Wixted, J. T., & Vul, E. (2008). The effects of tests on learning and forgetting. *Memory & Cognition*, 36(2), 438–448. <https://doi.org/10.3758/MC.36.2.438>
- Carrier, M., & Pashler, H. (1992). The influenced of retrieval on retention. *Memory and Cognition*, 20, 632–642. <https://doi.org/10.3758/BF03202713>
- Cranney, J., Ahn, M., McKinnon, R., Morris, S., & Watts, K. (2009). The testing effect, collaborative learning, and retrieval-induced facilitation in a classroom setting. *European Journal of Cognitive Psychology*, 21(6), 919–940. <https://doi.org/10.1080/09541440802413505>
- Dobson, J. L., & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Sciences Education*, 20, 149–161. <https://doi.org/10.1007/s10459-014-9514-8>
- Dobson, J., Linderholm, T., & Perez, J. (2018). Retrieval practice enhances the ability to evaluate complex physiology information. *Medical Education*, 52(5), 513–525. <https://doi.org/10.1111/medu.13503>
- Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01054>
- Ferreira, R., Sierra, V. S., & Vega, S. (2019). Immediate testing is more beneficial than delayed testing when learning novel words in a foreign language. *Revista signos*:

- estudios de lingüística*, 52(100), 290–305. <http://dx.doi.org/10.4067/S0718-09342019000200290>
- Gardiner, F. M., Craik, F. I., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory & Cognition*, 1(3), 213–216. <https://doi.org/10.3758/BF03198098>
- Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02412>
- Goldstein, E. B. (2008). *Cognitive psychology: Connecting mind, research and everyday experience* (2nd ed.). Cengage Learning.
- Hinze, S. R., & Wiley, J. (2011). Testing the limits of testing effects using completion tests. *Memory*, 19(3), 290–304. <http://dx.doi.org/10.1080/09658211.2011.560121>
- Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2015). Test-enhanced learning in third grade children. *Educational Psychology*, 35(4), 513–521. <https://doi.org/10.1080/01443410.2014.963030>
- Johnson, C. I., & Mayer, R. E. (2009). A testing effect with multimedia learning. *Journal of Educational Psychology*, 101(3), 621–629. <https://doi.org/10.1037/a0015183>
- Kang, S. H., Gollan, T. H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, 20(6), 1259–1265. <https://doi.org/10.3758/s13423-013-0450-z>
- Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp 487–509). Academic Press.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.

<https://doi.org/10.1126/science.1199327>

Karpicke, J. D., & Roediger, H. L. (2007a). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33(4), 704–719.

<https://doi.org/10.1037/0278-7393.33.4.704>

Karpicke, J. D., & Roediger, H. L. (2007b). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, 57(2), 151–162.

<https://doi.org/10.1016/j.jml.2006.09.004>

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968. <https://doi.org/10.1126/science.1152408>

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, 32(3), 401–410.

<https://doi.org/10.1002/acp.3410>

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, 23(4), 425–449. [https://doi.org/10.1016/S0022-5371\(84\)90282-](https://doi.org/10.1016/S0022-5371(84)90282-2)

2

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). Academic Press.

Larsen, D. P., & Butler, A. C. (2013). Test-enhanced learning. In Walsh, K. (Ed.), *Oxford textbook of medical education* (pp. 443–452). Oxford University Press.

Little, J. L., & Bjork, E. L. (2012). The persisting benefits of using multiple-choice tests as learning events. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 683–688). Cognitive

Science Society.

- Little, J. L., Bjork, E. L., Bjork, R. A., & Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: Fostering test-induced learning and avoiding test-induced forgetting. *Psychological Science*, 23, 1337–1344. <http://dx.doi.org/10.1177/0956797612443370>
- Lyle, K. B., & Crawford, N. A. (2011). Retrieving essential material at the end of lectures improves performance on statistics exams. *Teaching of Psychology*, 38(2), 94–97. <https://doi.org/10.1177/0098628311401587>
- McConnell, M., Hou, C., Panju, M., Panju, A., & Azzam, K. (2018). Does testing enhance learning in continuing medical education? *Canadian Medical Education Journal*, 9(3), e83–e88.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representation through retrieval. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 11, 371–385. <https://doi.org/10.1037/0278-7393.11.2.371>
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494–513. <https://doi.org/10.1080/09541440701326154>
- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14, 200–206. <https://doi.org/10.3758/BF03194052>
- McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, 27, 360–372. <http://dx.doi.org/10.1002/acp.2914>

- McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology*, 72, 609-633. <https://doi.org/10.1146/annurev-psych-010419-051019>
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, 20(1), 3–21. <https://doi.org/10.1037/xap0000004>
- Moreira, B. F. T., Pinto, T. S. S., Starling, D. S. V., & Jaeger, A. (2019). Retrieval practice in classroom settings: a review of applied research. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00005>
- 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. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9)
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Rahmat, N. H., Arepin, M., & Sulaiman, S. (2020). The cycle of academic reading fear among undergraduates. *Asian Journal of University Education*, 16(3), 265–274. <https://doi.org/10.24191/ajue.v16i3.9730>
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>

- Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, 17, 382–395. doi: 10.1037/a0026252
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 36(1), 233–239. <https://doi.org/10.1037/a0017678>
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140(6), 1432–1463. <https://doi.org/10.1037/a0037559>
- Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, 22(7), 784–802. <https://doi.org/10.1080/09658211.2013.831454>
- Smith, M. A., Blunt, J. R., Whiffen, J. W., & Karpicke, J. D. (2016). Does providing prompts during retrieval practice improve learning? *Applied Cognitive Psychology*, 30(4), 544–553. <https://doi.org/10.1002/acp.3227>
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation.
- Stenlund, T., Sundström, A., & Jonsson, B. (2016). Effects of repeated testing on short-and long-term memory performance across different test formats. *Educational Psychology*, 36(10), 1710–1727. <https://doi.org/10.1080/01443410.2014.953037>
- Sternberg, R. J. & Sternberg, K. (2012). *Cognitive psychology* (6th ed.). Cengage Learning.
- Thomas, A. K., & McDaniel, M. A. (2007). The negative cascade of incongruent generative study-test processing in memory and metacomprehension. *Memory & Cognition*, 35(4), 668–678. <https://doi.org/10.3758/BF03193305>

- Thomas, R. C., Weywadt, C. R., Anderson, J. L., Martinez-Papponi, B., & McDaniel, M. A. (2018). Testing encourages transfer between factual and application questions in an online learning environment. *Journal of Applied Research in Memory and Cognition*, 7(2), 252–260. <https://doi.org/10.1016/j.jarmac.2018.03.007>
- Wiklund-Hörnqvist, C., Jonsson, B., & Nyberg, L. (2014). Strengthening concept learning by repeated testing. *Scandinavian Journal of Psychology*, 55(1), 10–16. <https://doi.org/10.1111/sjop.12093>
- Willingham, D. T. (2009). *Why don't students like school? A cognitive scientist answers questions about how the mind works and what it means for the classroom*. John Wiley & Sons.
- Yong, P. Z., & Lim, S. W. H. (2016). Observing the testing effect using Coursera video-recorded lectures: A preliminary study. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.02064>

APPENDICES

APPENDIX A

Reading Passage on Problem Solving

A problem occurs when there is an obstacle between a present state and a goal, and it is not immediately obvious how to get around the obstacle. Thus problem, as defined by psychologists, is difficult, and the solution is not immediately obvious. There are two different types of problems. One type, such as solving a math or physics problem, is called a well-defined problem. Well-defined problems usually have a correct answer, and there are certain procedures that, when applied correctly, will lead to a solution. Another type of problem, like dealing with relationships or picking a career, is called an ill-defined problem. Ill-defined problems, which occur frequently in everyday life, do not necessarily have one “correct” answer, and the path to their solution is often unclear.

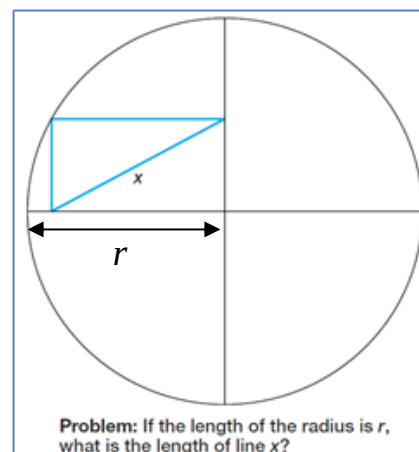


Figure 1

The Gestalt psychologists realised that one way to approach problem solving was to ask how problems are represented in a person’s mind. Problem solving, for the Gestalt psychologists, was about (1) how people represent a problem in their mind, and (2) how solving a problem involves a reorganisation or restructuring of this representation. The Gestalt idea of representation and restructuring in problem solving is illustrated by considering Figure 1. This problem, which was posed by Gestalt psychologist Wolfgang Kohler, asks us to determine the length of the segment marked x , if the radius of the circle has a length r . The key to solving this problem is to create the mental representation of x as being a diagonal of the small rectangle. Representing x as

the diagonal enables us to reorganize the representation by creating the rectangle's other diagonal. Once we realise that this diagonal is the radius of the circle, and that both diagonals are the same length, we can conclude that the length of x equals the length of the radius, r . What is important about this solution is that it does not require mathematical equations. Instead, the solution is obtained by first perceiving the object and then representing it in a different way. The Gestalt psychologists called the process of changing the problem's representation restructuring.

The Gestalt psychologists also introduced the idea that restructuring is associated with insight—a sudden realisation of a problem's solution. Gestalt psychologists' emphasis on insight is reflected in the types of problems they posed. The solution to most of their problems involves discovering a crucial element that leads to solution of the problem. The Gestalt psychologists assumed that people solving their problems were experiencing insight because the solutions usually seemed to come to them all of a sudden. Modern researchers have debated whether insight actually exists, with some pointing to the fact that people often experience problem solving as an “Aha!” experience—at one point they don't have the answer, and then next minute they have solved the problem—which is one of the characteristics associated with insight problems. Therefore, besides the two types of problems mentioned earlier (well-defined and ill-defined problems), the problems can also be classified as either insight problems or non-insight problems. Janet Metcalfe and David Wiebe (1987) did an experiment designed to distinguish between insight problems and non-insight problems. Their starting point was the idea that there should be a basic difference between how participants feel they are progressing toward a solution as they are working on an insight problem, and how they feel as they are working on a non-insight problem.

They predicted that participants working on an insight problem, in which the answer appears suddenly, should not be very good at predicting how near they are to a solution, but that participants working on a non-insight problem, which involves a more methodical process, would have some knowledge that they are getting closer to the solution. To test this hypothesis, Metcalfe and Wiebe gave participants insight problems as shown in Figure 2 (the solution to this problem is in Figure 3), and non-insight problems (Algebra problems taken from high school mathematics test) and had them make “warmth” judgments (nearness to solution) every 15 seconds, as they were working on the problems. Ratings closer to “hot” (7 on a 7-point scale) were used when they felt they were getting close to a solution, and ratings closer to “cold” (1 on the scale) were used when they felt that they were far from a solution. Results of the experiment showed that for the algebra problems, the ratings gradually increased until the problem was solved and as for insight problems, the solution does in fact, occurred suddenly.

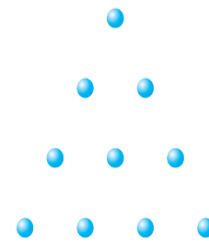


Figure 2: Show how you can move three of the circles to get the triangle to point to the bottom of the

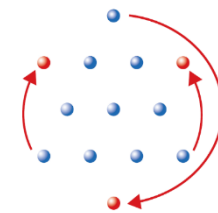


Figure 3: Solution to the triangle problem. Arrows indicate movement; coloured circles indicate new positions.

One of the major obstacles to problem solving, according to the Gestalt psychologists, is fixation (mental set)—people’s tendency to focus on a specific characteristic of the problem that keeps them from arriving at a solution. One type of fixation that can work against solving a problem is focusing on familiar uses of an object. Restricting the use of an object to its familiar functions is called functional fixedness. A simple scenario that demonstrates functional fixedness would be when a person wants to enter the house but

forgets to bring her house key. She only has her purse, a mobile phone and a hairpin. The person would be experiencing functional fixedness if she thinks the objects she has in hand would not be useful to help unlock her house door because she restricts the objects to its familiar use (purse = to keep money; mobile phone = to call someone; hairpin = to keep hair in place). To solve her problem, she can use the hairpin to unlock the padlock of her house. Hairpins are not only useful to clip and pin our hair in place but can also be used to unlock the house door.

Passage taken and modified from:

Goldstein, E. B. (2008). *Cognitive psychology: Connecting mind, research and everyday experience* (2nd ed.). Cengage Learning.

APPENDIX B

Retrieval Practice Phase Questions

MULTIPLE CHOICE QUESTIONS (10 marks)

Choose the best answer.

1. When does a problem occur?
 - A When the initial state is not clearly defined.
 - B When there is an obstacle between present state and goal.
 - C When it is immediately obvious on how to get around the obstacle.
 - D When there is a need to find easy solution to get to the goal state.

(LOT: Remember)

2. What is the psychological definition of a problem?
 - I Only involves initial state and goal state.
 - II Is difficult
 - III Solution is always obvious
 - IV Solution is not immediately obvious

- A I and II
 - B II and III
 - C II and IV
 - D III and IV

(LOT: Remember)

3. All of the following describe ill-defined problems EXCEPT

- A Usually occurs in our daily life.
- B The path leading to the solution may vary.
- C Solutions come to the individuals all of a sudden.
- D Does not necessarily have one right answer.

(LOT: Understand)

4. According to Gestalt psychologists, which one of the following is the first step in solving the problem in Figure 1?

- A Apply mathematical equation to find the length of x
- B Relate the problem to past solutions of finding the radius of the circle
- C Relate the problem to past solutions of finding the length of x
- D Create a mental representation of the object in the circle

(LOT: Understand)

5. What is involved when one is experiencing insight during the process of problem solving?

- A He applies a methodical process to find the solution to the problem.
- B He suddenly becomes aware of the solution to the problem.
- C He can find more than one solution.
- D He compares the current solution to the previous one.

(LOT: Understand)

6. The unemployment rate in Malaysia has gone from 3.3 percent in March 2019 to 3.9 percent in March 2020, signifying an increase in our country's unemployment rate. This is a concern for many especially among graduates. This has long been an issue, but it is now become a grave concern especially now we are now in the midst of a pandemic. How should we tackle this problem?

The above mentioned problem demonstrates

- A problem solving involving insight.
- B problem solving involving restructuring
- C a well-defined problem
- D an ill-defined problem

(HOT: Apply)

7. "A car starts from rest and accelerates uniformly over a time of 5.21 seconds for a distance of 110 m. Determine the acceleration of the car."

The above mentioned problem demonstrates

- A a well-defined problem
- B problem solving involving insight
- C an ill-defined problem
- D problem solving involving restructuring

(HOT: Apply)

8. Which one of the following aspects was used by Metcalfe and Wiebe (1987) as a way to differentiate between insight problems and non-insight problems?
- A How someone finds the fastest way to solve a problem.
 - B How one feels when he is progressing towards a solution
 - C How an individual represents a problem in their mind
 - D How one finds the easiest solution to a problem by applying certain procedures.

(HOT: Analyse)

9.

Lily was in her mother's kitchen preparing lunch for the family. When she was ready to serve the soup, she searched all the cupboards and drawers for a ladle but could not find one. She decided to wait until her mother to return to ask her where the ladle was, leaving the soup in the stove pot. Her mother later explained that the ladle had been broken, so she told Lily to use a coffee mug to spoon the soup into bowls.

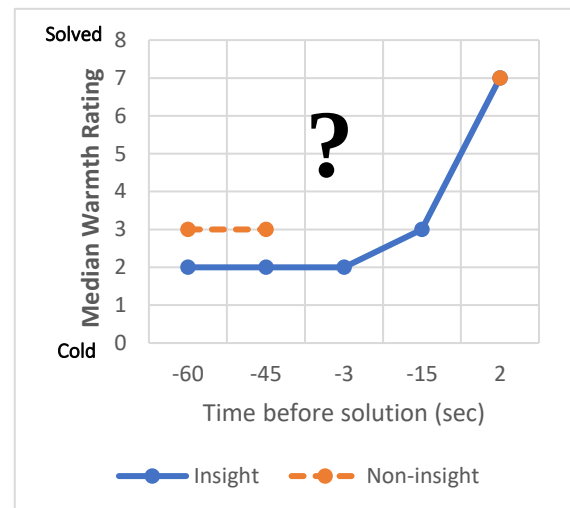
Which following description explains the obstacle Lily faced in solving the “serve the soup” problem?

- A It requires the application of the correct procedures in order to solve the problem.
- B She thinks that there could be more than one solution.
- C She restricts the use of the coffee mug to its familiar functions.
- D It requires her to create a mental representation of the problem.

(HOT: Evaluate)

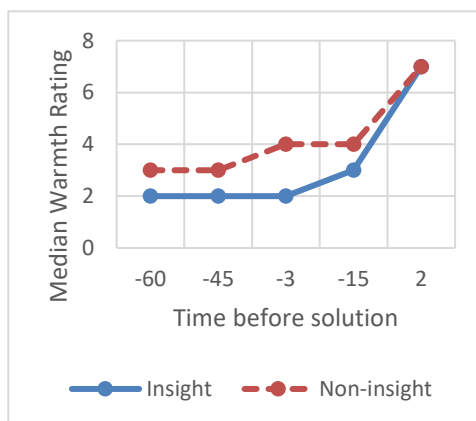
10. Janet Metcalfe and David Wiebe (1987) did an experiment to differentiate between insight problems and non-insight problems. They gave participants insight and non-insight problems and asked them to make “warmth” judgments (how closer they are to the solution) every 15 seconds as they were working on the problems. Graph 1 showed the results of the experiment.

Based on the above description, predict how the line for non-insight problem should look like in Graph 1.
(HOT: Create)

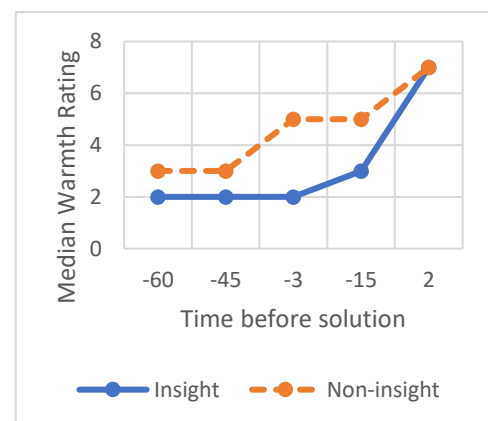


Graph 1

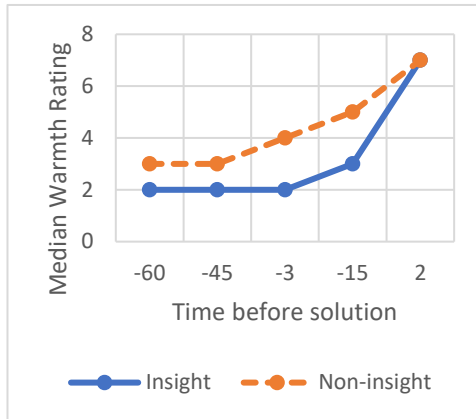
A



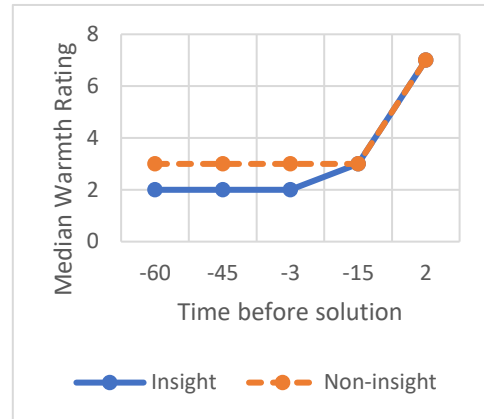
B



C



D



SHORT-ANSWER QUESTIONS (10 marks)

Write down the correct answer.

1. “Occurs when there is an obstacle between a present state and a goal, and the solution to get around the obstacle is not immediately obvious”.

The above description refers to ...

(LOT: Remember)

2. According to the Gestalt psychologists, one of the major obstacles to problem solving is called ...

(LOT: Remember)

3. How do you describe well-defined problems?

(LOT: Understand)

4. Based on the Gestalt psychologist's approach, what are the TWO (2) important processes involved?

(LOT: Understand)

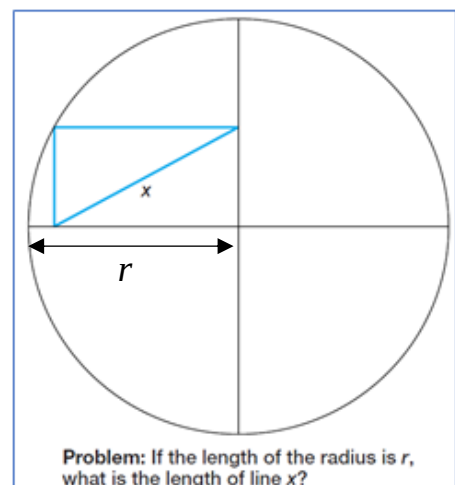


Figure 1

5. Give ONE (1) difference between insight problems and non-insight problems?

(LOT: Understand)

6. Provide ONE (1) example of an ill-defined problem (besides the one stated in the passage).

(HOT: Apply)

7. “A woman has four pieces of chain. Each piece is made up of three links. She wants to join the pieces into a single closed loop of chain. To open a link costs 2 cents and to close a link costs 3 cents. She only has 15 cents. How does she do it?”



What type of problem is demonstrated by the above description?

(HOT: Apply)

- 8.

Restructuring
(process of changing
problem's representation)

Insight

According to the passage, how are these two concepts related?

(HOT: Analyse)

9.

Find the derivative of $h(y) = y^{-4} - 9y^{-3} + 8y^{-2} + 12h$

What is the type of problem illustrated by the above example? Justify your answer.

(HOT: Evaluate)

10.

“Mental illness is expected to be the second biggest health problem affecting Malaysians after heart disease. Based on the latest National Health and Morbidity Survey, every three in 10 adults aged 16 years and above in Malaysia suffer from some form of mental health issues. Like any parts of the globe, mental illness is a rising issue in the country and should be taken seriously.”

What type of problem is this? Can you elaborate on the reason why you give that answer?

(HOT: Create)

APPENDIX C

Final Test Phase Questions

FINAL TEST: MULTIPLE CHOICE QUESTIONS (20 marks)

Choose the best answer.

1. When does a problem happen?
 - A When it is immediately obvious on how to get around the obstacle.
 - B When the initial state is not clearly defined.
 - C When there is an obstacle between present state and goal.
 - D When there is a need to find easy solution to get to the goal state.

(LOT: Remember, MCQ Q1)

2. “The solution to get around the obstacle is not immediately obvious”.

The above description refers to ...

- A Representation
- B Analogy
- C Restructuring
- D Problem

(LOT: Remember, SA Q1)

3. What is the psychological definition of a problem?

- I Is difficult
- II Only involves initial state and goal state.
- III Solution is not immediately obvious
- IV Solution is always obvious

- A I and II
- B I and III
- C II and III
- D III and IV

(LOT: Remember, MCQ Q2)

4. is defined as people's tendency to focus on a specific characteristic of the problem that keeps them from arriving at a solution.

- A Representation
- B Analogy
- C Fixation
- D Insight

(LOT: Remember, SA Q2)

5. All of the following describe ill-defined problems EXCEPT

- A The path leading to the solution may vary.
- B Solutions come to the individuals all of a sudden.
- C Does not certainly have one right answer.
- D Happens frequently in everyday life.

(LOT: Understand, MCQ Q3)

6. According to Gestalt psychologists, which one of the following is the first step in solving the problem in Figure 1?

- A Relate the problem to past solutions of finding the radius of the circle.
- B Construct a mental representation of the object in the circle.
- C Use mathematical equation to find the length of x .
- D Relate the problem to previous solutions of finding the length of x .

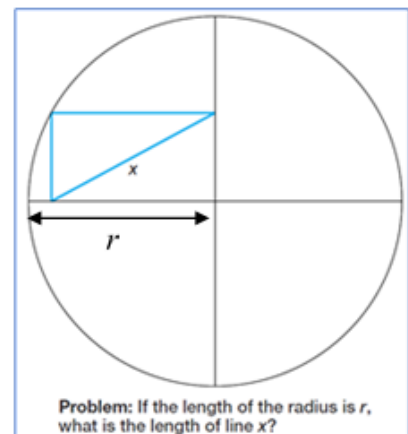


Figure 1

7. What is involved when one experiences insight during the process of problem solving?

- A The individual suddenly becomes aware of the solution to the problem.
- B The individual can find more than one solution.
- C The individual applies a methodical process to find the solution to the problem.
- D The individual compares the current solution to the previous one.

(LOT: Understand, MCQ Q5)

8. Which description best describes well-defined problems?
- A Typically have one correct answer
 - B The path leading to the solution is usually not obvious.
 - C Usually can have more than one solution.
 - D Difficult to specify a single correct answer for the problem.

(LOT: Understand, SA Q3)

9. Based on the Gestalt psychologists' approach, what are the TWO (2) important processes involved?
- I How people represent a problem in their mind.
 - II How the problem must have more than one right answer.
 - III How solving a problem involves restructuring of this representation.
 - IV How to ensure the path to the solution must always be clear.

- A I and II
- B I and III
- C II and III
- D I and IV

(LOT: Understand, SA Q4)

10. What is the difference between insight problems and non-insight problems?

A **Insight:** Answer come all of a sudden

Non-insight: Individuals are not good at predicting how near they are to the solution

B **Insight:** Have knowledge that they are closer to the solution

Non-insight: Involves more methodical process

C **Insight:** Individuals have the knowledge that they are getting closer to the solution.

Non-insight: Solution to the problem appears suddenly

D **Insight:** Answer to the problem appears suddenly.

Non-insight: Involves a more methodical process in solving the problem

(LOT: Understand, SA Q5)

11. Which of the examples below is NOT an ill-defined problem?

- A Deciding what to eat for lunch.
- B Solving trigonometry problem.
- C Overpopulation in a country.
- D Choosing what to wear to a dinner.

(HOT: Apply, SA Q6)

12. Covid-19 brings great impact not only to public health but also to business owners. It was reported that most business owners especially those running small business have suffered great loss due to the pandemic. The profits they gained may not be enough to sustain the business as well to pay their employees. How can they solve this problem?

The above mentioned problem demonstrates

- A problem solving involving restructuring.
- B problem solving involving insight.
- C an ill-defined problem.
- D a well-defined problem.

(HOT: Apply, MCQ Q6)

13. “A man's speed with the current (rate of movement in the water) is 15 km/hr and the speed of the current is 2.5 km/hr. The man's speed against the current is?”

The above mentioned problem demonstrates

- A an ill-defined problem
- B problem solving involving insight
- C problem solving involving restructuring
- D a well-defined problem

(HOT: Apply, MCQ Q7)

14. “A woman has four pieces of chain. Each piece is made up of three links. She wants to join the pieces into a single closed loop of chain. To open a link costs 2 cents and to close a link costs 3 cents. She only has 15 cents. How does she do it?”



The above description demonstrates what type of problem?

- A An insight problem
- B A non-insight problem
- C An ill-defined problem
- D A functional fixedness

(HOT: Apply, SA Q7)

15. Which of the following characteristic was used by Metcalfe and Wiebe (1987) as a way to differentiate between insight problems and non-insight problems?
- A How someone finds the fastest solution to a problem by applying certain procedures.
 - B How someone represent a problem in their mind.
 - C How someone feels when he is making progress towards a solution.
 - D How someone finds the easiest way to solve a problem.

(HOT: Analyse, MCQ Q8)

16.

Restructuring

Insight

According to the passage you have read, how are these two components connected?

- A Both are discovered by Metcalfe and Wiebe (1987).
- B Both lead to the discovery of the solution
- C Restructuring does not lead one to experiencing insight.
- D As the number of possible restructuring increases, the chances to experience insight decreases.

(HOT: Analyse, SA Q8)

17.

Factorise $x^2 + 2x - 8$

Choose the suitable explanation that justifies why the above problem is a non-insight problem.

- A The person solving the problem would not have any knowledge that he is getting nearer to the solution.
- B Solution to the problem appears suddenly to the individual trying to solve it.
- C The individual would have some knowledge that he is getting closer to the solution .
- D Solution requires the person to create a mental representation of the problem in the mind.

(HOT: Evaluate, SA Q9)

18.

Jenny bought fresh flowers to decorate the living room of the house. When she wanted to put the flowers in a vase, she searched everywhere in the house for an empty vase but could not find one. She decided to wait until her mother to return to ask her if there is any empty vase, leaving the flowers still wrapped in newspaper. Her mother later told her that the empty vase in the storeroom had broken, so she told Jenny to use an empty wine bottle to keep those flowers.

Jenny's ability to solve the problem was hindered by which of the following obstacle?

- A Analogy

- B Reorganisation
- C Insight
- D Functional fixedness

(HOT: Evaluate, MCQ Q9)

19. “Mental illness is expected to be the second biggest health problem affecting Malaysians after heart disease. Based on the latest National Health and Morbidity Survey, every three in 10 adults aged 16 years and above in Malaysia suffer from some form of mental health issues. Like any parts of the globe, mental illness is a rising issue in the country and actions should be taken to solve this problem.”

Source: https://umsc.my/umsc_news/resolving-mental-illness-issues-in-malaysia/

Which of the following facts can you gather from the above problem description?

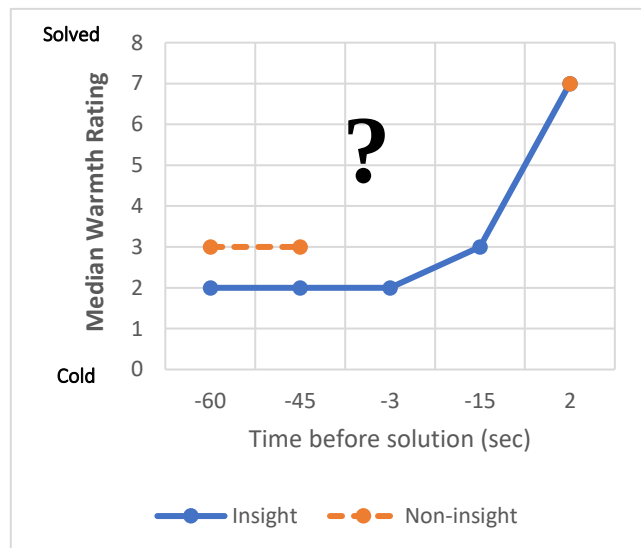
- A It is an insight problem and the solution to the problem appears all of a sudden.
- B It is an ill-defined problem and the path to the solution is usually not clear.
- C It is a well-defined problem and the path to the solution is usually clear.
- D It is a problem that requires restructuring to find a solution.

(HOT: Create, SA Q10)

20. Janet Metcalfe and David Wiebe (1987) did an experiment to differentiate between insight problems and non-insight problems. They gave participants insight and non-insight problems and asked them to make “warmth” judgments (how closer they are to the solution) every 15 seconds as they were working on the problems. Graph 1 showed the results of the experiment.

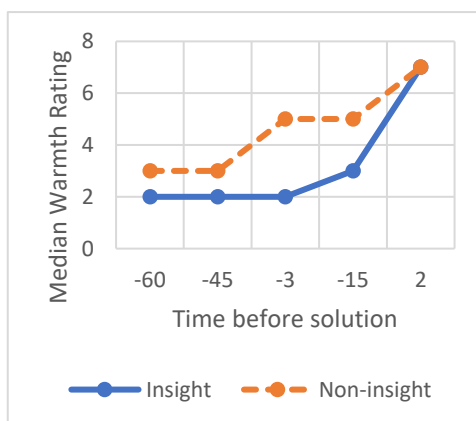
Based on the above description, predict how the line for non-insight problem should look like in Graph 1.

(HOT: Create, MCQ Q10)

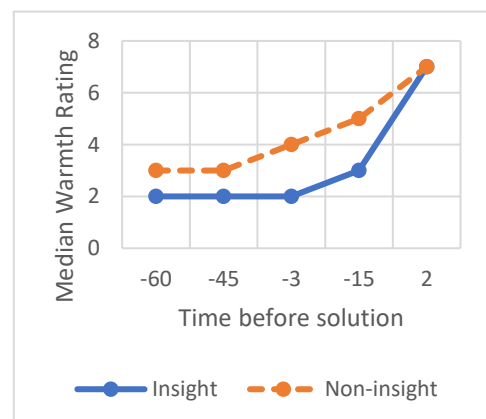


Graph 1

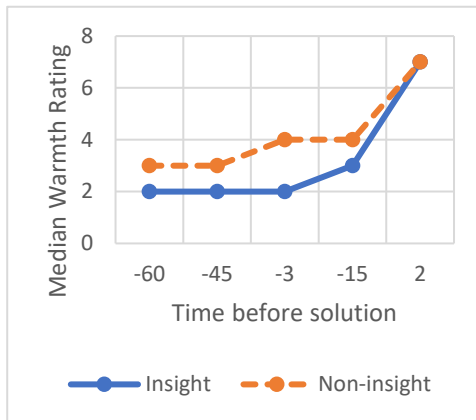
A



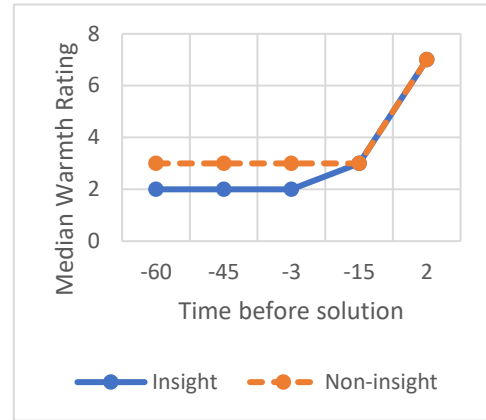
B



C



D



APPENDIX D

Research Information and Consent Form

Title of Study: The Effects of Retrieval Practice Across Levels of Thinking and Retention Interval on Reading Comprehension

The purpose of the study

The purpose of this study is to examine the effect of different types of retrieval practice on reading comprehension across levels of thinking and retention interval.

What is involved

In this experiment, you will be asked to read one reading passage on problem solving. Later, you will be asked to restudy the passage by answering some questions about it. Participation in this study will take approximately 60 minutes.

Participation and withdrawal

Participation in this study is voluntary and you are free to withdraw from this study at any time without prejudice or penalty.

Confidentiality and security of data

All data collected in this study will be confidential and solely be used for research purposes.

Risks

Participation in this study should involve no physical or mental discomfort, and no risks beyond those of everyday living.

Thank you for your participation in this study.

Corrin Alicia.

(MSc Cognitive Science student)

PARTICIPANT CONSENT

If you agree to participate in this experiment, please fill in your details.

Name:

Matric No.:

Email:

APPENDIX E

Integrity Declaration Form

I hereby acknowledge that research studies require honest participation from the participants to ensure the accuracy of the data collected. For the research experiment I am about to participate in, I make the following pledge:

1. I will not make a copy of the reading passage and questions shown to me.
2. All the answers given are from my own and I will not copy from any other sources (Internet/Books/Screenshots).
3. All other extra devices (e.g., handphone, tablet) should be in off-mode and kept away during the course of the experiment.
4. I shall not disturb other participants and contact other people during the course of the experiment.
5. I shall start the experiment under the instruction of the researcher.

I have read and understood the declaration and I certify that all information given or provided is correct, and nothing but the truth.

By affixing my NAME below, I further confirm that the answers I am submitting with this experiment solely my own. I have used NO notes, materials, or any other assistance.

NAME:

APPENDIX F

Journal Publication

1. Nero, C. A., & Zulkipli, N. (2021). The effects of retrieval practice across levels of thinking and retention interval on reading comprehension. *Asian Journal of University Education*, 17(4), 288–301. <https://doi.org/10.24191/ajue.v17i4.16222>