

**Partial replication of Baddeley et al. (1974) experiment on the word
length effect on short-term memory recall**

Psychology Standard Level Internal Assessment

IB Candidate Code: lsh546

Group Members Codes: kkk807 & kkk804

Date of Submission: March, 2025

Word Count: 2197

Table of Content

1. Introduction.....	2
1.1 Aim.....	3
1.2 Variables.....	3
1.3 Hypothesis (H1).....	3
1.4 Null Hypothesis (H0).....	3
2. Exploration.....	4
2.1 Design.....	4
2.2 Sampling.....	4
2.3 Controls.....	4
2.4 Materials.....	5
2.5 Procedure.....	5
3. Analysis.....	7
3.1 Descriptive Statistics.....	7
3.2 Inferential Statistics.....	7
4. Evaluation.....	10
4.1 Discussion of Findings.....	10
4.2 Design.....	10
4.3 Sample.....	11
4.4 Procedure.....	11
Bibliography.....	13
Appendices.....	14
Appendix I: Pools of Words.....	14
Appendix II: Bar graph.....	15
Appendix III: Calculations.....	15
Appendix IV: Raw Data.....	16
Appendix V: Wilcoxon Signed-Rank Test.....	19
Appendix VI: Consent Form.....	20
Appendix VII: Debriefing Form.....	21
Appendix VIII: Generalized Instructions.....	22

1. Introduction

Numerous models and theories have attempted to explain and understand how memory works. One significant theory in this field is the Working Memory Model (WMM). As proposed by Baddeley & Hitch (1974), short-term memory is described as a dynamic, multi-component system unlike its predecessor, the Multi-Store Model of Memory (MSMM). The WMM consists of four components: the central executive, which directs attention and controls processes; the phonological loop, which processes verbal and auditory information; the visuospatial sketchpad, which handles visual and spatial data; and finally, the episodic buffer, which integrates information from different sources into a coherent episode. The WMM highlights the role of the phonological loop in processing and retaining verbal information. According to the WMM, the phonological loop has a limited capacity, constrained by the time it takes to rehearse information. This implies that shorter words, requiring less time for rehearsal, fit more efficiently within this capacity, making them easier to recall than longer words (Jalbert et al., 2011). This principle directly informs the design of our experiment, as it predicts that word length will significantly impact recall performance.

In 1975, Baddeley et al. conducted a study to investigate how word length affects short-term memory. They aimed to understand whether the duration of words impacted recall ability, focusing on factors such as rehearsal and the modality of presentation (visual or auditory). The study consisted of experiments comparing participants' recall of short and long words with variations of the presentation methods. The finding showed that shorter words were generally easier to recall, and articulatory suppression, where the verbal rehearsal is interrupted by a concurrent task (van 't Wout & Jarrold, 2022), eliminated the word length effect in visual tasks but not auditory ones. This study illustrated the significant role that word length and rehearsal play in short-term memory performance (Baddeley et al., 1975).

1.1 Aim

The investigation described in this report is a partial replication of Baddeley et al. (1975). By focusing on the word length effect, the research aims to **investigate whether the immediate memory span varies with the word length to be recalled in a high school student population by testing the phonological loop's time-based processing limitation**. Understanding the word length effect is highly valuable as it could be used to develop new note-taking techniques in the education sector, thus helping students study more effectively.

1.2 Variables

1.2.2 Independent variable

A collection of both short words (one and two syllables) and long words four and five syllables) with similar frequencies of use in the English language.

1.2.3 Dependent variable

The memory capacity of the participants in the study

1.3 Hypothesis (H1)

There will be a significant difference in the number of words recalled between short and long words in terms of memory span by International Baccalaureate DP students from a list of words taken from Baddeley et al. (1975).

1.4 Null Hypothesis (H0)

There will be no significant difference in the number of words recalled between short and long words in terms of memory span by International Baccalaureate DP students from a list of words taken from Baddeley et al. (1975).

2. Exploration

2.1 Design

The research design selected for this study is a repeated measures design. The participants were exposed to short and long words, experiencing all conditions within the experiment. A repeated measures design was chosen to control individual differences as each participant acted as their control. Using a repeated measures design minimized external interferences and enabled reliable attribution to word length.

Additionally, using a standardized presentation method, where words were shown at fixed intervals, ensured consistency across all participants, enhancing the study's internal validity.

2.2 Sampling

To collect our participants, we used convenience sampling, which was the only method available. This was done by approaching students who had not chosen psychology as a subject, ensuring they were unfamiliar with our experiment, and reducing possible demand characteristics. We chose these participants thanks to their availability as they were in the same math class as us. We used our classmates similarly to Baddeley that used his colleagues for his study.

The sample consisted of 16 IB students from our international school in Brussels, aged 16-18. They were a mix of males and females from diverse cultural backgrounds and native languages but shared the same fluency in English, so the language level did not affect the participant's recall ability.

2.3 Controls

Directives to participants

Directives were written in advance to ensure consistency of understanding between participants (*see Appendix VII*).

Standardized word set

The word lengths within the sets (One set comprised eight monosyllables, namely, *sum, hate, harm, wit, bond, yield, worst, and twice*. The other set comprised eight five-syllable words, namely *association, opportunity, representative, organization, considerable, immediately, university, and individual*.

Presentation consistency

A custom app that I've programmed displayed each word at 2 second intervals, ensuring consistency of exposure time.

Quiet environment

The experiment took place in a quiet room to minimize distractions and enhance data reliability.

2.4 Materials

Two phones were used: one to show words via a specially developed app and the other to record participant answers. The app was designed to display text at specified intervals, maintaining consistency and avoiding bias. This solution enabled a standardized presentation while preserving internal validity by managing the time a word was displayed.

The pool of words was selected based on the original experiment by Baddeley et al. (1975) (*see Appendix I*). Using a similar set allowed for a more accurate partial replication and comparison of results. The words were categorized by syllable count (short vs. long) to examine the word length effect clearly.

2.5 Procedure

- Quiet room
- We sat at a table with researchers on one side and participants on the other.
- Participants filled out the consent form (*see Appendix VII*).
- The researchers read the generalized introductory text (*see Appendix IX*)

- One researcher recorded the participants.
- One researcher showed the app to the participants.
- The app consisted of 5 lists of 5 words randomly selected from the pool of words from Baddeley et al. (1975) in uppercase letters.
- Each word was presented at 2-second intervals.
- After a list presentation was ended, the participants had to recall the words from that list.
- After the recall, the following list started until the 5 lists were presented.
- The researchers then debriefed the participants (*see Appendix VII*).

3. Analysis

3.1 Descriptive Statistics

Conditions	Short words (one and two variables)	Long words (Four and five variables)
Mean	5.89	4.50
Standard deviation	1.81	2.48

Table 1: Table shows the calculation of Appendix III for the mean and standard deviation with outliers

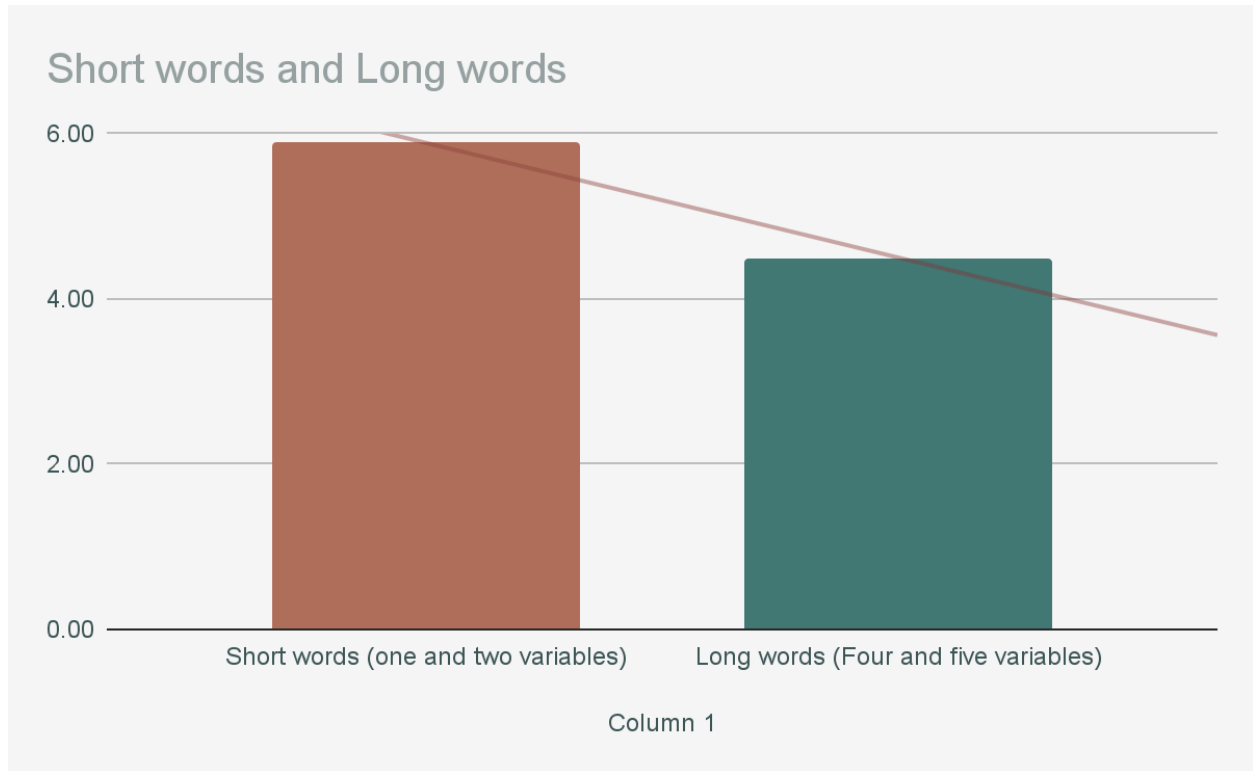
From the results obtained from 16 participants (*see Appendix IV*), the mean of both condition and standard deviation were calculated (*see Appendix III*) using Google Sheets.

We chose to calculate the mean as a measure of central tendency and the standard deviation as a measure of dispersion because the mean summarizes typical performance, and the standard deviation highlights the variability in recall consistency. Thus, the data synopsis and the graph effectively show that the mean between short words ($\bar{x} = 5.89$) and long words ($\bar{x} = 4.50$) were different, with the short words having the highest mean. This shows that the correlation between the mean number of correctly recalled items and the number of syllables in a word is inversely proportional. Furthermore, this suggests that short words are easier to recall than long words, which aligns with H_1 and that the memory span is inversely proportional to word length. Additionally, long words have the highest standard deviation ($\sigma = 2.48$), indicating a wide range of recall performance for these words. In contrast, short words have the lowest standard deviation ($\sigma = 1.81$), suggesting more consistent recall performance among participants.

In our raw data (see Appendix IV), we identified outliers such as the recall scores for “Puma” and “Mexico”. These outliers reflect either abnormal recall performance or external factors that influenced certain trials. Outliers were noted but did not significantly skew our overall findings.

3.2 Inferential Statistics

Investigate whether the immediate memory span varies with the word length to be recalled in a high school student population.



To analyze whether there was a statistically significant difference in recall performance between short and long words, a Wilcoxon Signed-Rank Test (see Appendix V) was conducted using raw data (see Appendix IV). This non-parametric test is suitable for comparing two related samples, especially when the data does not meet the assumptions of normality. In our study that used a repeated-measure design, each participant acted as their own control, with both short and long word recall scores being collected from the same group.

The test yielded a z -value of -1.6547 with a p -value of 0.04947 . Since $p < 0.05$, this result is statistically significant at the 0.05 level for a one-tailed hypothesis, supporting H_1 . However, the W -value of 36 did not meet the critical threshold of 35 , making the W -based interpretation non-significant.

In conclusion, the null hypothesis H_0 is cautiously rejected, indicating that word length affects recall, with shorter words being easier to remember.

4. Evaluation

4.1 Discussion of Findings

Even if the W -value did not meet the critical threshold of 35, we can see the difference between the mean of the number of short words recalled and the mean of the number of long words recalled on the bar graph (*see Appendix II*). The results obtained in the replicated study are consistent with findings from Baddeley et al. (1975), which support the phonological buffer theory within the Working Memory Model. The results indicate that participants recalled more short words ($\bar{x} = 5.89$) than long words ($\bar{x} = 4.50$). This trend suggests that word length impacts recall, with short words being more accessible to remember due to the time-based limitation of the phonological loop.

The Wilcoxon Signed-Rank Test confirmed this effect, yielding a significant p -value of 0.04947 (one-tailed), allowing us to reject the null hypothesis cautiously. However, the non-significant W -value suggests caution in interpreting the effect's strength. This discrepancy may reflect the small sample size or individual differences in recall consistency, particularly given the higher standard deviation observed for long words (2.48 for long words vs. 1.81 for short words).

These results imply that word length influences memory span, consistent with the time-dependent nature of the phonological loop. Thus, findings are valuable for educational contexts, as they suggest strategies focusing on short, more manageable chunks of information to aid memory retention.

4.2 Design

A strength of this study's design was using a repeated measures approach, where each participant completed both the short and long word recall tasks. This design helped reduce participant differences as it allowed us to use the same list of words among participants, making it easier to compare results across conditions and enhancing the reliability of the findings. Additionally, individual differences were minimized because participants were in both conditions, making comparisons across tasks more reliable.

However, a potential limitation of this design is that all participants did the tasks in the same order. This could have led to order effects, such as participants introducing fatigue or enhancing their recall performance for this exercise with training.

A potential modification would be to implement an independent sample design. By having participants experiencing only one condition, either short or long words, the risk of guessing the study's aim would be reduced, thereby minimizing demand characteristics. This approach would also prevent carryover effects from influencing results, strengthening the study's validity.

4.3 Sample

The participants were selected from our school, which heavily simplified logistics, as everyone had a similar schedule, ensuring consistent availability. While the sample comprised students from diverse cultural backgrounds and different mother tongues, all were fluent in English, enhancing the study's applicability to an international audience. However, these linguistic differences might have influenced how participants processed and recalled English words, introducing variability unrelated to word length. Despite this diversity, all participants were from the same age group and education environment, limiting the finding's broader generalizability. The use of convenience sampling created a relatively homogeneous group, given the shared social and academic backgrounds typical of International Baccalaureate students. Furthermore, with only 16 participants, the sample size was small, increasing the likelihood of Type II errors and limiting the statistical power of the analysis. A sample modification would be expanding the sample to include a broader age range and varied educational backgrounds, which would improve the result's generalizability.

4.4 Procedure

A key strength of the procedure was the use of a custom-developed app to standardize the timing and presentation of words across all participants. This ensured that each word was shown for precisely 2

seconds, enhancing internal validity by maintaining consistent experimental conditions. Additionally, conducting the experiment in a quiet environment minimized distractions, helping to ensure that participants were entirely focused on the recall and memorization tasks, which contributed to reliable data collection.

However, the procedure had some limitations. One issue was the potential for participants' stress due to the researchers being present during the task. This may have affected performance and introduced variability in the results. Additionally, interruptions during data recording, such as speaking to indicate what sequence was being analyzed, may have disrupted participants' short-term memory recall. These procedural elements could have negatively affected the reliability of the data collected.

To improve the procedure, it would be beneficial to use a pre-organized system for recording responses, reducing the need for interruptions that could disrupt participants' memory recall. Furthermore, using a fully automated system that records responses without researcher interference could further reduce stress and potential biases. Another modification would be to conduct the experiment in a more neutral setting, away from the presence of researchers, to create a more comfortable environment for participants. Finally, standardizing the word list across all participants by creating the lists in advance in a randomized manner, using, for example, "random.org/lists/," would reduce variability while avoiding bias. This approach strengthens the internal validity of the study by ensuring that differences in recall performance are related to word length rather than variations in word difficulty.

To conclude, the results demonstrate that short words are remembered more effectively than long words, which can be attributed to the word length effect. This suggests that short-term memory has a higher capacity for shorter words, as longer words require more time to articulate, thereby limiting the opportunity for rehearsal.

Bibliography

APA Dictionary of Psychology. (n.d.). Dictionary.apa.org.

<https://dictionary.apa.org/phonological-loop>

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word Length and the Structure of short-term Memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575–589.

[https://doi.org/10.1016/s0022-5371\(75\)80045-4](https://doi.org/10.1016/s0022-5371(75)80045-4)

Baddeley, A., & Hitch, G. (1974). *WORKING MEMORY*.

<https://app.nova.edu/toolbox/instructionalproducts/edd8124/fall11/1974-Baddeley-and-Hitch.pdf>

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 338–353. <https://doi.org/10.1037/a0021804>

van 't Wout, F., & Jarrold, C. (2022). Articulatory suppression during instruction encoding impedes performance in choice reaction time tasks. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-022-02100-5>

Appendices

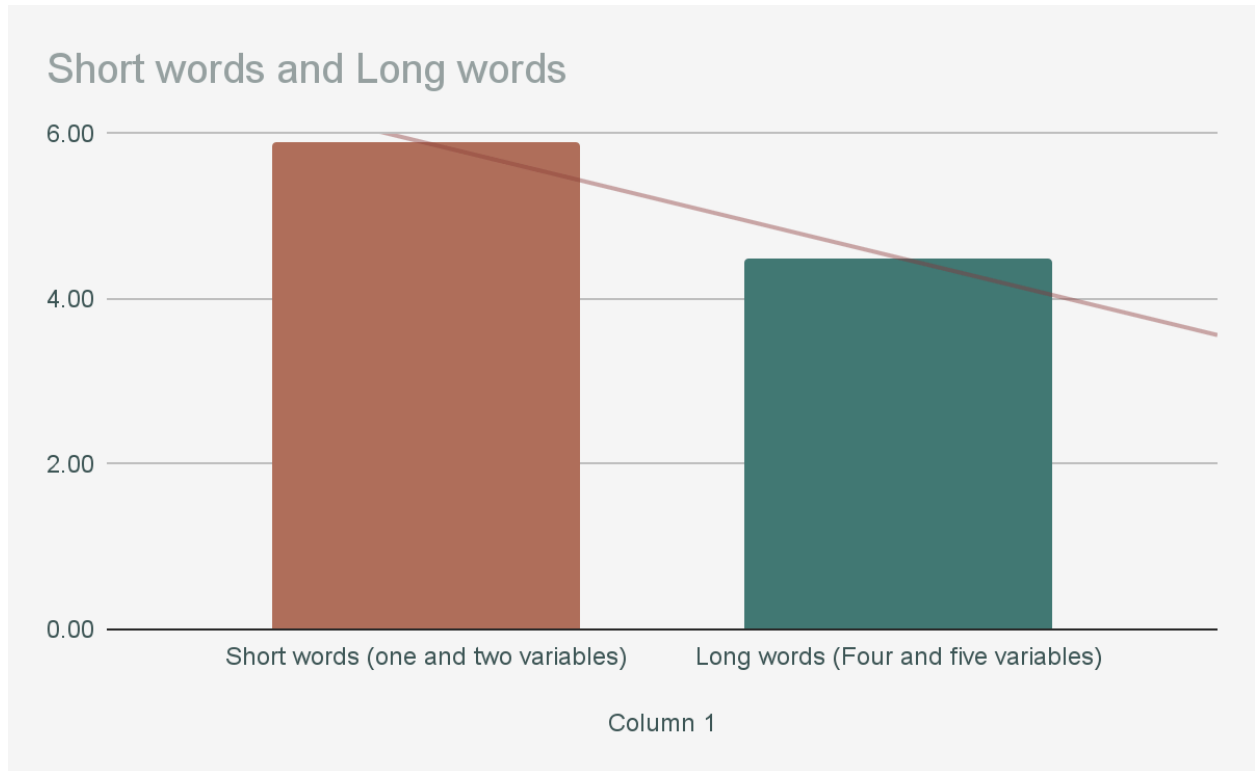
Appendix I: Pools of Words

Pools of words:

Column 1	Column 2	Column 3	Column 4	Column 5
Stoat	Puma	Gorilla	Rhinoceros	Hippopotamus
Mumps	Measles	Leprosy	Diphtheria	Tuberculosis
School	College	Nursery	Academy	University
Greece	Peru	Mexico	Australia	Yugoslavia
Crewe	Blackpool	Exeter	Wolverhampton	Weston-Super-Mare
Switch	Kettle	Radio	Television	Refrigerator
Maths	Physics	Botany	Biology	Physiology
Maine	Utah	Wyoming	Alabama	Louisiana
Scroll	Essay	Bulletin	Dictionary	Periodical
Zinc	Carbon	Calcium	Uranium	Aluminum

Appendix II: Bar graph

Bar graph showing the relationship between the mean number of items recalled correctly and the number of syllables - with outliers



Appendix III: Calculations

Calculations in Google Sheets to obtain the mean and standard deviation.

	Short words	Long words
Mean	=AVERAGE(C3:C20)	=AVERAGE(C30:C47)
Standard deviation	=STDEV(C3:C20)	=STDEV(C30:C47)

Appendix IV: Raw Data

Detailed Recall Data by Word and Syllable Count

Word	Number of syllables	Items recalled correctly
Stoat	1	4
Mumps		4
School		5
Greece		9
Crewe		9
Switch		6
Maths		7
Maine		4
Scroll		4
Puma (OUTLIER)	2	9
Measles		6
College		5
Peru		4
Blackpool		5
Kettle		6
Physics		5
Utah		6
Essay		8
Gorilla	3	7

Leprosy		3
Nursery		4
Mexico (OUTLIER)		8
Exeter		3
Radio		5
Botany		4
Wyoming		4
Bulletin		5
Rhinoceros		2
Diphtheria		2
Academy		7
Australia		3
Wolverhampton		2
Television		6
Biology		9
Alabama		7
Dictionary	4	2
Hippopotamus		4
Tuberculosis		8
University		6
Yugoslavia		6
Weston-Super-Mare (OUTLIER)	5	0

Refrigerator		6
Physiology		3
Louisiana		4
Periodical		4

Appendix V: Wilcoxon Signed-Rank Test

Wilcoxon Signed-Rank Test done with experiment data ([Appendix IV](#))

Wilcoxon Signed-Rank Test Calculator

Success!

Explanation of results

We have calculated both a W -value and z -value. If the size of N is at least 20 - see the Results Details box - then the distribution of the Wilcoxon W statistic tends to form a normal distribution. This means you can use the z -value to evaluate your hypothesis. If, on the other hand, the size of N is low, and particularly if it's below 10, you should use the W -value to evaluate your hypothesis.

You should also note that if a subject's difference score is zero - that is, if a subject has the same score in both treatment conditions - then the test discards the individual from the analysis and reduces the sample size. If you have a lot of ties, this procedure will undermine the reliability of the test (and also suggests that the requirement that the data is continuous has not been met).

Treatment 1	Treatment 2	Sign	Abs	R	Sign R
4	2	1	2	6	6
4	2	1	2	6	6
5	7	-1	2	6	-6
9	3	1	6	15	15
9	2	1	7	16	16
6	6	n/a	0	n/a	n/a
7	9	-1	2	6	-6
4	7	-1	3	11	-11
4	2	1	2	6	6
9	4	1	5	13.5	13.5
6	8	-1	2	6	-6
5	6	-1	1	1	-1
4	6	-1	2	6	-6
5	0	1	5	13.5	13.5
6	6	n/a	0	n/a	n/a
5	3	1	2	6	6
6	4	1	2	6	6
8	4	1	4	12	12

Significance Level:

☐ .01

☒ .05

1 or 2-tailed hypothesis?:

☒ One-tailed

☐ Two-tailed

Result Details

W -value: 36
Mean Difference: 4.38
Sum of pos. ranks: 100
Sum of neg. ranks: 36

Z -value: -1.6547
Mean (W): 68
Standard Deviation (W): 19.34

Sample Size (N): 16

Result 1 - Z -value

The value of z is -1.6547. The p -value is .04947.

The result is significant at $p < .05$.

Result 2 - W -value

The value of W is 36. The critical value for W at $N = 16$ ($p < .05$) is 35.

The result is *not* significant at $p < .05$.

Calculate

Reset

Appendix VI: Consent Form

Consent form

The Department of Psychology at [REDACTED] supports the practice of the protection of human participants in research. The following will provide information about the experiment that will help you decide whether or not you wish to participate.

If you agree to participate, please be aware that you are free to withdraw at any point throughout the duration of the experiment without any penalty.

In this study we will ask you to memorize words from lists, if you have any discomfort or don't feel at ease during the experiment, please inform the experimenter and the study will end now. All information you provide will remain confidential and will not be associated with your name. If for any reason during this study you do not feel comfortable, you may stop the study and leave and receive credit for the time you participated, and your information will be discarded. Your participation in this study will require approximately 20 minutes. When this study is complete, you will be provided with the results of the experiment if you request them, and you will be free to ask any questions. If you have any further questions concerning this study, please feel free to contact us through email:

[REDACTED]

Please indicate with your signature on the space below that you understand your rights and agree to participate in the experiment. Your participation is solicited, yet strictly voluntary. All information will be kept confidential, and your name will not be associated with any research findings.

Full name: _____

Date: ____ / ____ / ____

Signature:

Appendix VII: Debriefing Form

Debriefing Form

Thank you for your participation in our study! Your participation is greatly appreciated.

Purpose of the Study:

Earlier in our consent form, we informed you that the purpose of the study was to memorize words from lists. In actuality, our study is about looking if there is a difference in serial recall and memory performance between participants and the length of words they had to remember from the sets.

Unfortunately, in order to properly test our hypothesis, we could not say that we were looking to see if you memorized shorter words easier prior to your participation as you could unconsciously develop strategies to take more effort to remember longer words which would reduce the validity of our data. This ensures that your reactions in this study were spontaneous and not influenced by prior knowledge about the purpose of the study, we didn't tell you that the length of words mattered. If we had told you the actual purposes of our study, your ability to memorize longer or shorter words could have been affected. We regret the deception, but we hope you understand the reason for it.

Confidentiality:

Please note that although the purpose of this study has changed from the originally stated purpose, everything else on the consent form is correct. This includes the ways in which we will keep your data confidential.

Now that you know the true purpose of our study and are fully informed, you may decide that you do not want your data used in this research. If you would like your data removed from the study and permanently deleted, please tick the following box.

☐ *I do not agree to have my data used from the study*

Please do not disclose research procedures and/or hypotheses to anyone who might participate in this study in the future as this could affect the results of the study.

Once again, thank you for your participation in this study

Appendix VIII: Generalized Instructions

Welcome to our study. Your participation is crucial for advancing our understanding of the memory of the brain. Please carefully follow our instructions to ensure a good collection of data. Before we begin, ensure that you have read and understood the informed consent form. By proceeding, you are confirming your voluntary participation in this study. You will be engaging in a series of tasks designed to assess your attention and memory abilities.

Please follow the instructions:

We will begin by showing you a set of 5 words each with a two-second delay to memorize the word and once the set is over, you will have time to recall all the words remembered. This process will be repeated for a total of 5 sets. The study will last around 5 minutes.

If you understood the instructions correctly, let us begin!