



# A Replication of Waugh and Norman (1965) Primary Memory study

Marilou Poitras<sup>a</sup> , Lucie Péléja<sup>a</sup> , Gardy Lavertu<sup>a</sup> , Anouck Langlois<sup>a</sup> , Katia Boulerice<sup>a</sup> , Pauline Berthelot<sup>a</sup> , Philippe Vincent-Lamarre<sup>b</sup>, Sandy Beaulieu<sup>a</sup>, Vanessa Bournival<sup>a</sup>, Laurence Brault<sup>a</sup>, Jenna Charlebois<sup>a</sup>, Eve Camille Galloway<sup>a</sup>, Ariane Gauthier<sup>a</sup>, Rose-Marie Gibeau<sup>a</sup>, Nicolas Giroux<sup>a</sup>, Grace Jacob<sup>a</sup>, Myriam La Flèche<sup>a</sup>, Lély-Rose Laurin<sup>a</sup>, Véronique Legault<sup>a</sup>, Marie Lessard<sup>a</sup>, Chelsea McAlpine<sup>a</sup>, Camille Mercier<sup>a</sup>, Albino Nikolla<sup>a</sup>, Dominique Perras<sup>a</sup>, Bryanna Marie Pidgeon<sup>a</sup>, Sydney Running<sup>a</sup>, Kim Thériault<sup>a</sup>, Marissa Hailey Trudel<sup>a</sup> & Mikaela Rose Winder<sup>a</sup>

<sup>a</sup>Université d'Ottawa

<sup>b</sup>Université du Québec à Montréal

**Abstract** ■ Waugh & Norman's experiment (1965) is among the most influential studies in the field of cognitive psychology. Using a probe digit memory test, they proposed that the proportion of correctly remembered items, digits in this case, depends on the number of interfering items shown between the recall item and the signal, or probe, identifying the recall target. This indicates that interference alone accounts for forgetting in short-term memory. The following study aimed to replicate these results with greater statistical power as the original study used a small sample of 4 participants. In a second study, we used shorter lists to examine potential effects of the relative difficulty of the task. Both studies' results partially support Waugh & Norman's claim, as participants were more likely to recall a digit that was followed by fewer interfering items. Additionally, we observed an interaction involving presentation rate and interference, as participants performed best with a low amount of interfering items presented at 1 item per second. However, lists with a higher number of interfering items (7 and above) had similar correct recall proportion regardless of the presentation rate. These findings further support the prevalence of interference theory over the decay theory but also call for a closer look at the possible interaction between the two. Future studies should examine the latter as well as the possible effect of cognitive fatigue due to the difficulty of the task, underlining the importance of replication studies.

**Keywords** ■ Short-term memory, Interference theory, Decay theory, Replication study.

[pbert033@uottawa.ca](mailto:pbert033@uottawa.ca)

[10.20982/tqmp.16.2.r001](https://doi.org/10.20982/tqmp.16.2.r001)

**Acting Editor** ■ Denis Cousineau (Université d'Ottawa)

**Reviewers**  
■ No authors of the original article reviewed this article

## Introduction

There are various types of memory systems, such as the long term, the short term and the perceptual (echoic and iconic) memories. In this article, we focus on the short-term memory, which is characterized by its very limited capacity as well as its ability to temporarily maintain cognitive information accessible (Cowan, 2008). Earlier studies have shown that the short-term memory has a capacity of 7 plus or minus 2 items (Miller, 1956). However, recent studies revised this capacity to roughly 4 (Cowan, 2001).

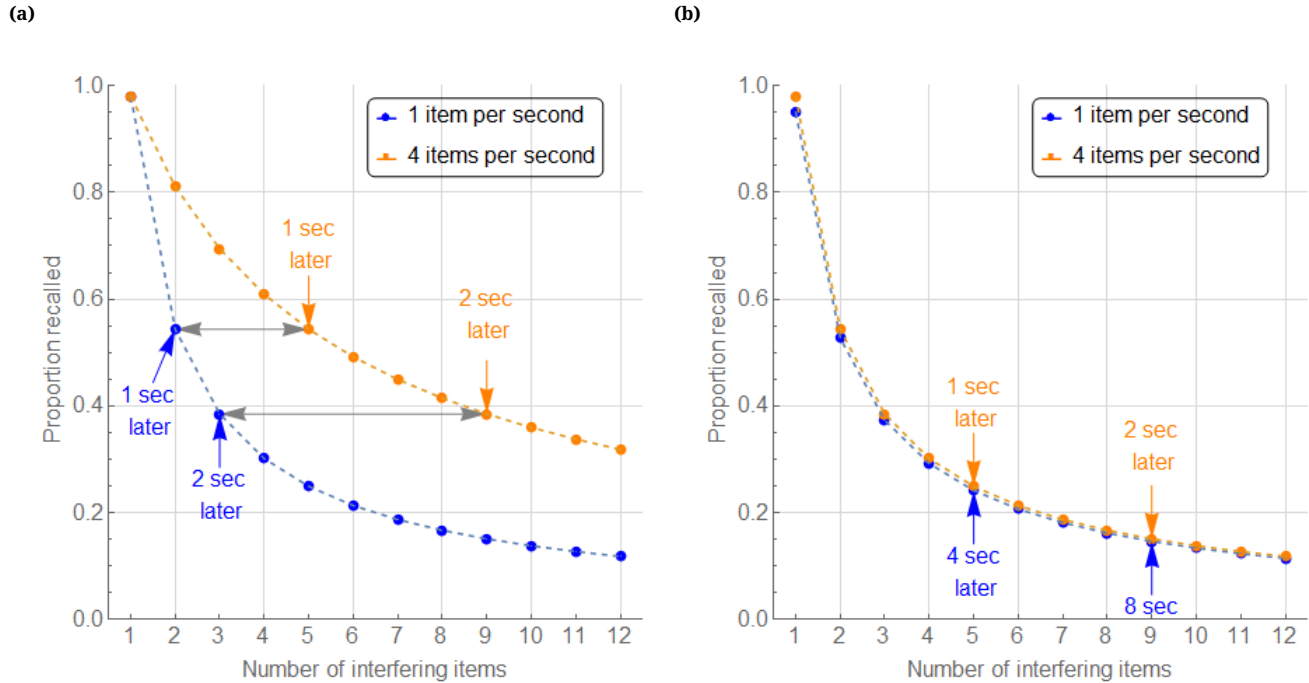
Many hypotheses have been considered to determine

what causes forgetting in short-term memory, the decay and interference theories being the most prominent (Jonides et al., 2008). Waugh and Norman (1965) argued that interference alone causes forgetting in short-term memory. It stipulates that the proportion of correctly remembered items depends on the number of interfering items shown between the to-be-remembered target and the cue indicating how to identify the target (a probe).

As shown in Figure 1, left panel, the interference theory predicts that the relevant factor is the number of interfering items. Thus, varying the rate of presentation –the main manipulation in Waugh and Norman (1965)– should make



**Figure 1** ■ Visual representation of the results expected based on the theory of interference (left) and the theory of decay (right). The curves were generated using a power curve,  $0.98 \times n^{-0.85}$ , where  $n$  is the number of interfering items, 0.98, the initial performance, and 0.85, the forgetting rate (Cousineau, Hélie, & Lefebvre, 2003).



no difference. Therefore, we should expect two identical curves regardless of the rate of presentation of items.

On the other hand, the right panel of Figure 1 shows the curve expected according to the decay theory, which states that the sole factor of forgetting is the amount of time elapsed. Contingent to this theory, varying the rate of presentation predicts different curves. However, the recall is identical for a given amount of time elapsed between the target and the probe.

This well-known study has been very influential and is found in nearly all textbooks on cognitive psychology. However, it has a major representativity limitation as it used a sample size of only 4 participants, constituted exclusively of Harvard undergraduate students. The APA (VandenBos, 2010) recommends that there are always at least 20 participants per group. It is meant to minimize the risk that a homogeneous—but unrepresentative—sample biases the results under cluster sampling (Cousineau & Laurencelle, 2015).

The purpose of this replication is to verify the results obtained by Waugh and Norman (1965) because their participants do not constitute a statistically representative sample due to its lack of size and diversity (Faber & Fonseca, 2014). Two studies were conducted. Study 1 is a direct replication of the original whereas Study 2 used a shorter

list of stimuli to examine the potential effect of the difficulty of the probe digit task.

### Study 1

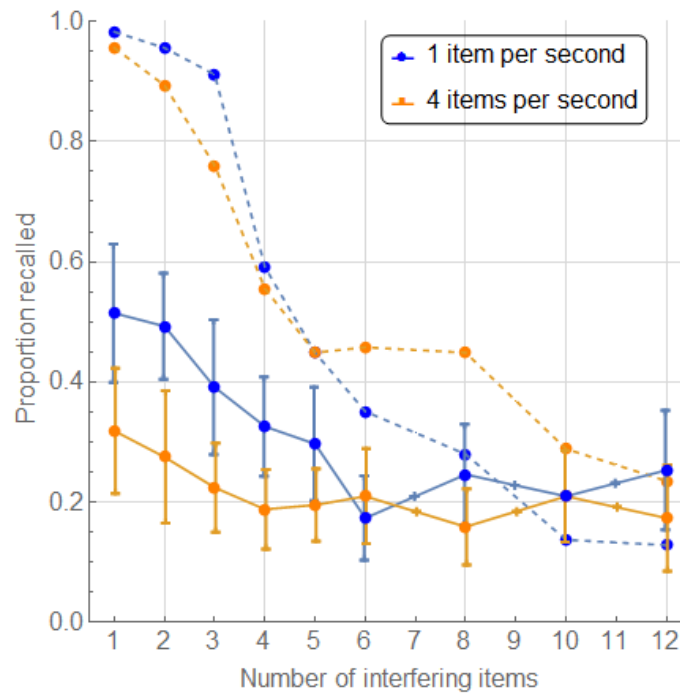
The goal of Study 1 is to validate the mechanisms underlying short-term memory forgetting using lists of unrehearsed digits, by modifying the presentation rate and the number of interfering items. Following the original study (Waugh & Norman, 1965), this paradigm tests the theory of decay against the theory of interference. The procedure is an exact replication of the classic study done by Waugh and Norman (1965). The replication study was pre-registered prior to beginning the data collection and the stimuli as well as the raw data for the experiment have been made available online (<https://osf.io/hgfyq/>).

### Methodology

The experimental session was divided in 4 blocks, with breaks in between each block of a duration determined by the participant. Each block contained 27 trials. One trial is composed of a list of 16 randomly generated single digits, in which the same digit could not appear more than twice in a row. The lists were presented at a rate of either one digit per second (slow list) or four digits per second (fast list). Each participant did both lists, the first 2 blocks were



**Figure 2** ■ Results obtained in the present replication (full lines) and in the original Waugh and Norman (1965) study (dashed line). The proportion of correct answers is presented as a function of the number of interfering items; error bars shows correlation and difference-adjusted 95% confidence intervals of the mean (Cousineau, 2017). There are no error bars on the original study because the raw data are not available.



presented at one randomly determined presentation rate and the following 2 blocks at the other rate. Prior to the presentation of each list, a white fixation plus sign would appear on the black screen for 4 seconds, after which it would disappear and the list would begin, presented auditorily.

The objective for the participants was to identify the number having immediately followed a probe digit, which could be presented only once before in the list, in positions 3, 5, 7, 9, 10, 11, 12, 13 or 14. The probe was the last digit presented and was followed by a 250 ms tone to indicate the end of the list. Each probe position was tested 6 times within each presentation rate. The participants indicated the digit having immediately followed the probe in the list by pressing the correct number key on the keyboard, guessing if unsure. Since the probe had only been presented once before in the list, there was only one possible correct answer. Nine practice lists were presented before the first and the third blocks to allow participants to accommodate to the task. During these practice trials, the participants could see if their answer was correct; these trials were not included in the analyses. During the actual trials, participants did not receive any feedback regarding

their response.

Instructions were given verbally to the participants by the experimenter and were also displayed on the computer screen prior to beginning the task. The instructions and the audio stream were presented through earphones, in either French or English according to the participant's primary language. The digits were read aloud by a monotone male voice synthesized using *Text2speech* (<https://www.text2speech.org/>). The recordings were modified using the software Audacity to ensure that the sound fit within 250 ms. The digit zero was excluded from the stimuli in both languages and the digit seven was excluded from the English stimuli because two-syllable digits would not fit within 250 ms.

Participants were recruited directly by the experimenters, who were students enrolled in the University of Ottawa's PSY3777 course of autumn 2019. There were no eligibility criteria other than for participants to not have any memory impairment. The participants were asked to bring their own headphones for the experiment. The software used for the experiment was E-Prime. The experience took place in silence to avoid any distractions. Each participant in the study was entered into a draw for a



chance to win one of three 25\$ Amazon gift cards.

## Results

The participants were 24 individuals from the greater Ottawa region, Ontario. The sample was composed of 63.8% females and most of the participants were undergraduate students at the University of Ottawa. The average age of the participants was 21.4 years ( $SD = 2.2$ ). The study was conducted in a computer room with 30 stations at the University of Ottawa.

The mean proportion of correct recall is given in Figure 2 as a function of the number of interfering items, separated for fast presentation rate and slow presentation rates. For comparison, we also added (dashed lines) the results of the original study.

Informal examination indicates no effect of the language in which the stimuli were presented and likewise no effects of gender.

A one-way within-subjects analysis of variance (ANOVA) comparing the effect of number of interfering items on short-term memory recall in both slow and fast presentation rate found a significant effect of presentation rate,  $F(1, 22) = 32.90, p < 0.001$ . Moreover, the effect of interfering items on memory recall also yielded a significant effect of interference,  $F(8, 176) = 7.84, p < 0.001$ . Finally, there was also a significant interaction between rate of presentation and number of interfering items,  $F(8, 176) = 2.47, p = 0.015$ .

## Discussion

The results obtained from this experiment partially support Waugh and Norman's (1965) conclusion that only interference affects retention in short-term memory. It can be noted that the average proportion of correct recall decreases with the number of interfering items, as predicted by Figure 2. However, the curves are distinct at first and then converge to an asymptotic performance at about 20% correct recall. Whereas the decay theory predicts that the two curves would diverge, we see the inverse effect.

When there were less interfering items, participants had overall a better performance when the listed numbers were presented at a rate of one digit per second than a rate of 4 digits per second. However, as the number of interfering items increased, performance for both presentation rates tend to be similar. Although this supports the interference theory, the fact that at a slower rate performances are higher indicates another possible variable causing this distinction.

A notable difference between our results and those obtained in the original experiment is the starting mean accuracy. As presented in Figure 2, at a minimum of three interfering items, Waugh and Norman (1965) reported a

proportion of correct answers almost equal to 1.0 for both presentation speeds. For the same amount of interfering items, our replication found a mean accuracy around 0.5 for the slower rate and around 0.3 for the faster one. The small sample size of the original study could partially account for these differences, as the original sample might have been less representative.

## Study 2

Waugh and Norman derived their experiment by taking into account a model of short term memory with two processes in which repetition serves both to maintain the items in primary memory and eventually to record them in a secondary memory; the modal model (Watkins, 1974). However, the original methodology with the initial lists of 16 digits does not take into consideration the memory span of 7 plus or minus 2 items of the short-term memory (Miller, 1956).

We therefore considered that it would be interesting to bring some modifications to Waugh and Norman's (1965) original study. The goal was to observe whether the results would remain the same with shorter lists and ascertain that the performances in the original study were not partly caused by the presentation of too many digits which would have made the task too difficult.

In the original study, there was an important effect of the number of interfering items, yet no effects were reported regarding the rate of presentation (1 item/sec and 4 items/sec). However, a study by Altmann and Schunn (2012), showed that the two factors interact more than what was initially believed. Indeed, when the rate of presentation is fast, the first items in the list benefit from the reduced effect of decay. Conversely, when presentation is slow, the later items benefit from the decay of previous items and the resulting reduction in proactive interference.

To examine this point, we tested these two presentation rates on shorter lists to observe if the influence would be the same on the modal model illustrated by Waugh and Norman as with the original lists of 16 digits.

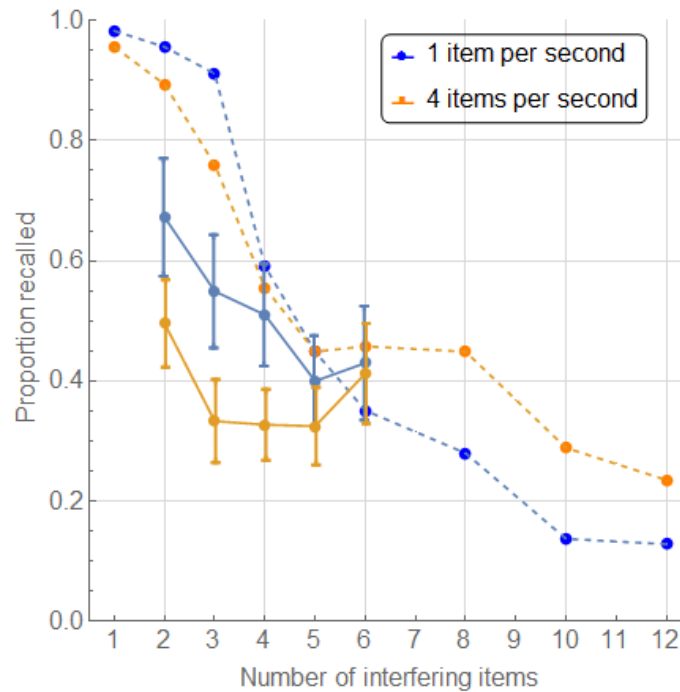
The theories of decay and interference were contrasted by using two different presentation rates. Also, having less interfering items in this study aimed to test the theory of interference and how much it would influence the participants' results. This study is not an exact replication of Waugh and Norman's (1965) original study as the lists included fewer items.

## Methodology

Study 2 aimed at evaluating once again the mechanisms of short-term memory, this time using lists of eight unrehearsed digits, instead of sixteen. Each block included 25



**Figure 3** ■ Mean proportion recalled across participants as a function of the number of interfering items for both presentation rates in Study 2 (full lines) and the original experiment (dashed lines).



trials of eight randomly generated single digits, with the constraint that the same digit could not appear more than twice in a row. The probe was the 8th digit presented, followed by a 250 ms tone to indicate the end of the list.

As in Study 1, the objective of the task was to identify the number having immediately followed the probe digit, which was presented only once before in the list, in positions 2, 3, 4, 5 or 6. The position of the probe digit and the length of the lists are the only differences in methodology between the two studies. Participants were entered in the same draw as those of Study 1 for a chance to win one of three 25\$ Amazon gift cards. For further details, refer to the methodology section of Study 1.

### Results

Study 2 had a total of 29 participants of which 69.1% ( $N = 20$ ) were women. Most subjects were undergraduate students from University of Ottawa and their age ranged from 17 to 25 years old ( $M = 20.56$ ,  $SD = 1.9$ ). Most participants completed the test in a University of Ottawa computer laboratory with 30 stations; 5 did it alone in an individual booth.

Figure 3 shows the proportion of correctly recalled items as a function of the number of interfering items. When there are fewer interfering items, the performance

for the slower presentation rate is superior to the faster presentation rate. A weaker performance is noted when the number of interfering items increased (3 to 6).

A two-way repeated measure ANOVA revealed a significant main effect of presentation rates on proportion of correct recall ( $F(1, 29) = 25.85$ ,  $p < 0.001$ ). Participants showed an increased performance in the 1 item per second compared to the 4 items per second condition. It also revealed a significant main effect of the number of interfering items on proportion of correct recall ( $F(4, 116) = 13.12$ ,  $p < 0.001$ ). Participants' performance improved when there were less interfering items following the probe digit. Finally, there was a significant interaction between presentation rate and number of interfering items ( $F(4, 116) = 4.01$ ,  $p = 0.004$ ) such that participants scored the highest when they were in the 1 item per second condition with a small number of interfering items following the probe digit. However, this effect seems to disappear when the probe digit is at the very beginning of the series (interfering items = 7). Participants scored similarly with 7 interfering items regardless of the presentation rate.

### Discussion

When comparing our results with those obtained by Waugh and Norman (1965), we noticed that for the larger





number of interfering items (6), there seems to be no difference between the two. However, a faster presentation is associated with a reduced recall performance for a smaller number of interfering items. This is contrary to the results that would be predicted by the decay theory. The implications and suspected causes of this phenomenon will be further discussed in the next section. It is also possible to graphically observe that the data of the last interfering item of both presentation rates are almost layered on top of each other which could potentially be the result of a primary effect. The participants' average performance in study 2 is lower than the average obtained by Waugh & Norman but higher than the one obtained in study 1.

### General discussion

In Study 1, performance declines, reaching a floor, regardless of presentation speed, after 6 interfering items. Average performance thereafter is just slightly above chance level, indicating that accurate recall was highly unlikely. This coincides with the maximal capacity of short term memory of 7 plus or minus 2 items reported in the literature (Miller, 1956). With the results found in study 2, we can see that performance for both presentation speeds also rejoins after 5 or 6 interfering items. This, in combination with the higher average recall when the length of the list is reduced such as in study 2, accentuates the concerns that presenting more than 8 digits surpasses the capacities of human short term memory.

Although this replication addressed the main limitation of the original study by having a statistically more valid sample size, other limitations still apply. The sample size, albeit much larger than the original, was still mainly composed of young adults pursuing post-secondary education. Additionally, while the original sample was composed mainly of men, our replication sample was mainly composed of women. These factors may impact the representativeness of our sample and affect the generalizability of our results.

As shown by the reduced performance in the replication, it is possible that the task was too difficult. Due to the challenge of presenting 4 digits per second, it is possible that the quality of the audio was affected, which negatively impacted performance. This could partially account for the observed differences between the fast and slow presentation rates, as the slower presentation rate might have been clearer. Due to the difficulty of the task, participants might also have been prone to becoming discouraged, which could have reduced the amount of effort made to correctly answer each list.

The experimental setting, which presents notable distinctions from the original study, might also account for the deviation from the expected results. Due to the higher

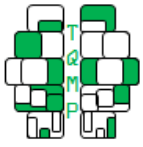
number of participants as well as the limited amount of time given to replicate the original study, multiple participants were tested in one location at the same time. This decreased the amount of control the experimenters had on the experimental setting and increased the amount of biases. For instance, the amount of distractions both in the surroundings, such as movement or conversations, and in the participants' own vicinity, such as their phone, could have decreased concentration on the task. Over 30 experimenters also took part in the making of this study which increases the variability of the experimental flow from one participant to the other. Furthermore, because some participants were given shorter lists than others, they finished the task sooner and were free to leave. This could have exerted a social pressure on the remaining participants with longer lists who were unaware of the time difference between the two experimental conditions. It is possible that this additional pressure further decreased performance.

### Authors' note

The first six authors have contributed equally to the article.

### References

- Altmann, E. M., & Schunn, C. D. (2012). Decay versus interference: A new look at an old interaction. *Psychological Science*, 23(11), 1435–1437. doi:10.1177/0956797612446027
- Cousineau, D. (2017). Varieties of confidence intervals. *Advances in Cognitive Psychology*, 13, 140–155. doi:10.5709/acp-0214-z
- Cousineau, D., Hélie, S., & Lefebvre, C. (2003). Testing curvatures of learning function on individual trial and block average data. *Behavior Research Methods, Instruments, & Computers*, 35, 493–503. doi:10.3758/bf03195528
- Cousineau, D., & Laurencelle, L. (2015). A correction factor for the impact of cluster randomized sampling and its applications. *Psychological Methods*, 21, 121–135. doi:10.1037/met0000055
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114. doi:10.1017/S0140525X01003922
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169, 323–338. doi:10.1016/S0079-6123(07)00020-9
- Faber, J., & Fonseca, L. M. (2014). How sample size influences research outcomes. *Dental press journal of orthodontics*, 19(4), 27–29. doi:10.1590/2176-9451.19.4.027-029.ebo



- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual review of psychology*, 59(1), 193–224. doi:[10.1146/annurev.psych.59.103006.093615](https://doi.org/10.1146/annurev.psych.59.103006.093615)
- Miller, G. A. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97. Retrieved from <https://doi.org/10.1037/h0043158>
- VandenBos, G. R. (2010). *Publication manual of the American psychological association (6th ed.)* Washington, DC: American Psychological Association.
- Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin*, 81(10), 695–711. Retrieved from <https://doi.org/10.1037/h0036952>
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104. doi:[10.1037/h0021797](https://doi.org/10.1037/h0021797)

### Open practices

- 👉 The *Open Material* badge was earned because supplementary material(s) are available on <https://osf.io/hgfgq/>.
- 👍 The *Preregistered* badge was earned because the experiment(s) was preregistered at <https://osf.io/hgfgq/>.

### Citation

Poitras, M., Péléja, L., Lavertu, G., Langlois, A., Boulerice, K., Berthelot, P., . . . Winder, M. R. (2020). A replication of Waugh and Norman (1965) Primary Memory study. *The Quantitative Methods for Psychology*, 16(2), r1–r7. doi:[10.20982/tqmp.16.2.r001](https://doi.org/10.20982/tqmp.16.2.r001)

Copyright © 2020, Poitras et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 27/03/2020 ~ Accepted: 13/04/2020