The benefits of memory control processes in working memory: comparing effects of self-reported and instructed strategy use

Lea M. Bartsch[1], Alessandra Souza[1,2] & Klaus Oberauer[1]

[1] University of Zurich, [2] Center for Psychology, Faculty of Psychology and Education Sciences, University of Porto

Author Note

We thank Atalia Adank, Natasa Milovanovic, and Dawid Strzelczyk for helping with data collection. The data and the analysis scripts can be accessed in the Open Science Framework (https://osf.io/m2hx3/). K. Oberauer (grant 100014_149193) and A. S. Souza (grant 100019_169302) were supported by grants from the Swiss National Science Foundation. A. S. Souza was additionally supported by national funds from the Portuguese Foundation for Science and Technology (UIB/00050/2020). Correspondence should be addressed to Lea M. Bartsch, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmuehlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: l.bartsch@psychologie.uzh.ch.

Abstract

Working memory performance is often assumed to benefit from different maintenance control strategies such as rehearsal, refreshing, elaboration, and grouping. In studies assessing strategy self-reports, some strategies were indeed associated with better recall. Nevertheless, experimental studies assessing the effect of instructing maintenance strategies compared to a no-instruction baseline lend no evidence for the effectiveness of these strategies for working memory. Explanations for this contradiction could be that instruction implementation engenders dual-task costs, or that strategy instructions reduce adaptive strategy switching. Across two experiments, we investigated the frequency and variability of strategy use with trial-wise self-reports in serial recall of word lists. Further, we examined potential instruction costs by comparing performance in trials with self-reported vs. instructed use of the same strategies. Self-reported strategy use varied from trial to trial, with elaboration and rehearsal being the most frequent. Self-reported elaboration was correlated with better performance than reading and rehearsal. For the most prevalent strategies – elaboration and rehearsal – there were no costs of instructed strategy-implementation. Our results speak against dual-task costs, and for an advantage of adaptively choosing one’s own strategy from trial to trial.

Keywords: Strategies, Working Memory, Long-term Memory, Elaboration, Rehearsal, Refreshing

The benefits of memory control processes in working memory: comparing effects of self-reported and instructed strategy use

Working memory (WM) is understood as a capacity-limited store that holds information available for ongoing processing. Maintenance strategies – such as rehearsal, refreshing, different forms of elaboration, as well as grouping – are assumed to control what is retained in WM, with different effects on immediate and delayed recall (Atkinson & Shiffrin, 1968). Rehearsal refers to the overt or covert repetition of the memoranda to oneself (Baddeley, 1986). Elaboration encompasses processes that enrich the memory representation of an item by activating many aspects of its meaning and by linking it into the pre-existing network of semantic associations (Craik & Tulving, 1975), as for example the creation of a sentence linking words presented for study, or the use of mental imagery. Refreshing is defined as a domain-general attentional process that is understood as briefly thinking of a stimulus just after it is no longer physically present but while its mental representation is still active (Johnson et al., 2002). Grouping is defined as organizing the memory list into shorter groups of pairs or triplets (Ryan, 1969a).

In some theories of WM, maintenance processes such as rehearsal and refreshing play a key role in explaining people's performance in immediate memory tasks (e.g., Baddeley, 1986; Barrouillet et al., 2011; Cowan, 2005). In other theories, maintenance processes play a more minor role (Nairne, 1990), or none at all (Lewandowsky & Farrell, 2008; Oberauer et al., 2012). Therefore, finding out which maintenance processes people engage in, and whether they are effective in maintaining information in WM, contributes to evaluating competing theories of WM. This is the aim of the present work.

Studies have shown that in WM tasks, subjects report to engage in these strategies to different degrees (Bailey et al., 2008, 2011; Belletier et al., 2023; Dunlosky & Kane, 2007; Kaakinen & Hyönä, 2007; Morrison et al., 2016; Richardson, 1998; Unsworth, 2016). Also, choice of strategy in these studies is correlated with retrieval success – some self-reported strategies such as mental imagery, sentence generation, and grouping are associated with better memory than others. Yet, this positive correlation between recall and self-reported strategy use is not sufficient to establish a causal effect of these strategies on memory storage. For instance, the relationship could equally well be explained by having a good representation of the current memory set being a prerequisite for carrying out strategies such as elaboration. Alternatively, participants could be using memory strength to infer the sorts of strategies that they may have used to encode this information. In order to draw causal inferences, we must turn to experimental evidence: Past research has compared performance under different experimental conditions in which the respective strategies were instructed, leading to inconclusive results or even evidence against their benefit (for a review see Oberauer, 2019).

Here we investigate possible explanations for why correlational studies on strategy use show consistent correlations with performance in typical WM tasks, such as complex or simple span (e.g., Bailey, Dunlosky, & Kane, 2008, 2011; Dunlosky & Kane, 2007; Richardson, 1998), whereas experimental attempts to improve performance by inducing strategies that were correlated with better performance in the aforementioned studies never succeeded (e.g., Bartsch, Singmann, & Oberauer, 2018; Jonker & Macleod, 2015; Souza & Oberauer, 2018). In the following, we briefly summarize both the correlational as well as experimental evidence for each strategy.

Self-Reported Use of Memory Control Processes

In past self-report studies, strategy use has been mostly assessed by asking participants at the end of the experiment to indicate for each trial – presented to them again - which strategy they used in that trial: reading, rehearsal, sentence generation, imagery, grouping, or something else. Rehearsal is generally reported most frequently, occurring in ca. 33-50% of trials of simple and complex span tasks; followed by reading (21-43%), mental imagery (6-14%), grouping (5-13%), and sentence generation (2-13%; see Bailey, Dunlosky, & Hertzog, 2009; Bailey et al., 2008, 2011 for details). Span performance is significantly higher in trials in which individuals report to engage in mental imagery and sentence generation compared to merely reading each word as it is presented, or engaging in rehearsal (Bailey et al., 2008, 2009, 2011; McNamara & Scott, 2001).

Given this correlational evidence on the effectiveness of some of these strategies, it seems plausible that either (a) increasing the time interleaving the memoranda in a WM task, or (additionally) (b) instructing a specific strategy in some trials, would allow any individual to engage in effective strategies *more often*, or to switch from a less effortful strategy (e.g., reading) to one that requires more processing, but which leads to a larger mnemonic benefit (e.g., sentence generation). In the following, the experimental evidence for this hypothesis and the role of each strategy is briefly summarized.

Experimental Evidence for a Beneficial Role of Maintenance Strategies in WM

To measure the occurrence of *rehearsal*, some previous studies have employed overt rehearsal protocols: participants are instructed to carry out their rehearsals overtly and the experimenter records them for analysis. This method permits the online evaluation of rehearsal as it unfolds in the WM task. Tan and Ward (2008) used an overt rehearsal protocol and observed a correlation between the length of cumulative rehearsal – i.e., looping through memory items in their order of presentation – and the benefits of increasing free time in a serial recall task. These results suggest that participants benefit from free time because they can rehearse the memoranda in the pauses between items.

Experimental work, however, has challenged the causal role of cumulative rehearsal for promoting memory benefits: One early study tested the effect of practicing cumulative articulatory rehearsal on performance in the complex-span paradigm (Turley-Ames & Whitfield, 2003). The trained group showed better performance when encoding was self-paced, yet this was because they took more time to encode the items. When the time to encode each item was controlled by the experimenter, the advantage of the trained group disappeared.

Using the overt rehearsal protocol, Souza and Oberauer (2018, 2020) showed that increasing the number and length of cumulative rehearsals during free time via instructions did not result in better performance compared to conditions without instruction (free rehearsal) in simple span as well as complex span tasks. These results cannot be explained by rehearsal being already maximally beneficial in the free rehearsal baseline. In one experiment, participants were instructed to read aloud during the free-time intervals either the spontaneous rehearsals performed by a yoked participant in the free-rehearsal condition, or irrelevant syllabi ("bababa"). Reading the rehearsals of another participant should facilitate rehearsal by even removing the necessity to retrieve memory items, whereas irrelevant articulation impedes rehearsal. Nevertheless, performance in these two conditions did not differ (Souza & Oberauer, 2018).

Elaboration (which encompasses sentence generation and imagery) is the second most commonly reported strategy. Recent studies investigating the effect of experimentally instructing elaboration in WM found no evidence for a WM benefit, although elaboration promoted long-term memory (LTM) (Bartsch et al., 2018, 2019; Bartsch & Oberauer, 2021; Jonker & Macleod, 2015). In the instructed elaboration condition, WM performance was never better, and in most cases worse, than in a baseline condition in which participants were free to do whatever they wanted.

Refreshing is a more recent addition to the list of maintenance strategies (see Camos et al., 2018 for a review). Participants hardly ever report spontaneously engaging in refreshing (Loaiza & Lavilla, 2021). Refreshing, however, has been argued to improve WM (Souza et al., 2015, 2018; Souza & Oberauer, 2017; Vergauwe & Langerock, 2017) as well as long-term memory (Camos & Portrat, 2015; Loaiza & Mccabe, 2012; McCabe, 2008). Studies explicitly guiding attention to individual WM elements via retro-cues during the retention interval improved performance for tests of the cued items, and performance increased with the number of guided refreshing steps (Souza et al., 2018). Nevertheless, instructing people to refresh a memory set did not lead to any WM or LTM benefits in comparison to a free baseline (Bartsch et al., 2018). Some authors have proposed a distinction between a quick and unconscious process of refreshing (e.g. in the TBRS model; Barrouillet & Camos, 2012) and a slower, more deliberate form of refreshing. The latter is defined as focusing one's conscious attention to an item after it is no longer physically present (Raye et al., 2007). Here, we are interested in the latter form of refreshing because only deliberate refreshing can be considered a strategy.

Participants also report *grouping* in a small proportion of trials (5-13%). Experimental evidence suggests that serial recall performance is higher in case the items are presented in groups of threes – operationalized through the insertion of temporal gaps at the group boundaries – compared to ungrouped presentation (e.g., Broadbent & Broadbent, 1981; Hartley, Hurlstone, & Hitch, 2016; Hitch, Burgess, Towse, & Culpin, 1996; Liu & Caplan, 2020; Ryan, 1969a, 1969b). Yet, imposing a temporal gap through the rhythm of the stimuli is not the same as participants trying to create groups of items presented at a constant pace. Early work indicated that performance in a condition with merely instructed grouping was numerically better than an ungrouped condition, but not significantly different (Ryan, 1969a, Exp 1). More recent work indicates that instructed grouping – when the list presentation is ungrouped – shows similar effects to those observed under temporal grouping on immediate memory performance (Farrell, 2008).

A recent study by Barrouillet et al. (Barrouillet et al., 2020) combined instructions of rehearsal and refreshing for different subsets of the memory list – a procedure they termed the *maxispan*. In the basic procedure, part of the memory list was presented in blue (3-5 letters) and the remaining items (1 to 6 letters) were presented in black font. Participants in the maxispan condition were instructed to rehearse the blue letters and mentally maintain the black letters; whereas participants in the control condition did not receive instructions regarding the colors. The maxispan procedure yielded higher spans when participants were instructed to rehearse 4 to 5 letters, especially when these were followed by large sets of black letters. This study suggests that instructing rehearsal could be beneficial particularly when large lists are presented (lists larger than six items) and when subsets of the list are clearly segregated (using different colors or presentation modalities). Yet, given that this study did not include control conditions in which subsets of the list were not grouped, or in which rehearsal (or refreshing) of all letters was instructed, it is difficult to separate what was the contribution of grouping, or whether this effect is due to rehearsal of only a subset of the list, refreshing of only a subset of the list, or both. Furthermore, the combination of rehearsal and refreshing is not something that participants report doing on their own (and notice that their baseline is a free-strategy condition). Hence, it is unlikely to account for the conflicting results between the correlational and experimental studies on strategy use.

Taken together, the maintenance strategies that some prominent theories of WM assume to be effective (i.e., rehearsal and refreshing), or that self-report studies suggest to be effective (i.e., elaboration), did not show a consistent benefit in experimental studies in which these strategies were instructed. This led to questioning their beneficial role for WM in general.

Hypotheses on Discrepancies Between Self-report and Experimental Instruction Studies

On the one hand, correlational studies indicate a relation between engaging in certain maintenance strategies and better recall from WM whereas, on the other hand, most experimental studies that manipulated the engagement in these strategies have observed no evidence that they improve memory. Why do these two sets of studies yield conflicting evidence?

We can think of two possible explanations. First, experimentally instructing a strategy of interest requires participants to maintain and carry out the strategy as instructed, which can function as a secondary task, thereby creating a dual-task cost on WM that counteracts any potential benefits of the instructed strategy relative to a baseline in which people were free to select their strategy. To uncover a potential dual-task cost of instructed strategies, one needs to compare performance in conditions with instructed and self-selected strategy use. One prediction from this hypothesis is that dual-task costs should be larger particularly for those strategies that are more cognitively demanding. Previous research has shown that elaboration is fairly demanding, rehearsal is modestly demanding, and reading is minimally demanding (Naveh-Benjamin & Jonides, 1984; Thalmann et al., 2019). Refreshing has also been conceptualized as a fairly demanding strategy (Camos et al., 2018). Accordingly, we should expect instructions to elaborate and refresh to be more costly to implement, followed by rehearsal, and lastly by reading. A second prediction of this hypothesis is that the dual-task costs should be greater for strategies that are less frequently used because these strategies are less practiced. Accordingly, we should expect instructions to refresh to be more costly to implement, followed by grouping, elaboration, and rehearsal.[1]

A second explanation starts from the observation from self-report studies that individuals spontaneously apply a mixture of various memory strategies over the course of an experiment (Morrison et al., 2016). This may indicate that they adaptively shift from one strategy to another depending on the memoranda, or based on their subjective success with a strategy. Studies in which participants are instructed to use one strategy throughout prevent such adaptive switches, thereby resulting in worse performance across all trials than in a baseline condition in which people can freely switch between strategies. If people choose their strategy based on how well it is suited for the memoranda of a trial, the difference between performance in instructed trials and in trials where people use the same strategy in

[1] Reading is also a rarely reported strategy, but reading a word is a very frequent activity, so that we can't expect it to impose a higher dual-task due to being less practiced than other strategies.

the free baseline condition would become most apparent for the strategies that are rarely chosen, because in those cases, the instructed strategy is not the best for the majority of trials. Alternatively, if adaptive switches occur based on subject's assessment of their performance, we should see switches more frequently following low performance trials compared to high performance trials.

Experiment 1

The goal of the present study was to investigate the effectiveness of WM strategies for maintenance. To resolve the ambiguity from previous research on this question, we asked whether instructions to engage in a certain strategy alter memory performance compared to when participants indicate to have engaged in the same strategy spontaneously. To that end we were interested in whether – within a person rather than across studies – performance in trials of self-reported use of a given strategy differs from performance of that same participant when this strategy was instructed. If performance is better when participants spontaneously engage in a strategy, this may indicate that instructed strategy-use does not produce a benefit in experimental studies because maintaining and implementing the instruction imposes a cost, or because it precludes natural variability and adaptive shifting between strategies.

We invited participants into the lab on two separate days to complete two sessions of a verbal serial-recall task in which the memoranda were interleaved with generous free time – providing ample opportunity to engage in any of the above candidate strategies.

In the first session, participants were free to engage in any strategy they wanted, and their chosen strategies were measured with strategy self-reports after recall in each trial. This procedure allowed us to get an insight into the trial-to-trial variability or stability of using the different strategies. Strategy reports after each trial are preferable over global strategy reports or retrospective reports collected at the end of the experiment because retrospective reports could be distorted by participants mis-remembering their own strategies. Importantly, strategy reports after each trial have been shown previously to have minimal reactive effects, with the reporting of a strategy on one trial having little influence on strategy selection or immediate serial recall performance on subsequent trials (Dunlosky & Kane, 2007).

In the second session, participants were instructed to use one of the memory strategies for mini-blocks of four trials. After recalling all items of each trial, they were to indicate whether they had complied with the instruction, and if not, what they had done instead. We kept the order of sessions constant to avoid any influence of the instructions of strategies on the pattern of strategies self-chosen by the participants. We discuss possible confounding effects of this choice such as fatigue and practice effects in the General Discussion. Both sessions further included a delayed free recall test for all the items presented during the session in order to measure the influence of the WM strategy applied for the formation of LTM representations.

Method

Participants

We recruited 30 students (20 females, 10 males; mean age = 23.47) from the University of Zurich. We chose the sample size because it is sufficient to detect medium to large effects in within-subjects designs. The use of Bayesian statistics means that the sample size could have been increased in case of ambiguous evidence (Rouder, 2014), but this was not necessary.

In both experiments reported here, participants were compensated with either 45 Swiss Francs (about 45 USD) or partial course credit for completing two sessions, with the first lasting one and the second lasting two hours. Participants signed an informed-consent form before the start of the experiment and were debriefed at the end. The studies were carried out in agreement with the rules of the Ethics Committee of the Faculty of Arts and Sciences of the University of Zurich and did not require special approval.

Materials and Procedure

Lists of six nouns were studied for a forward serial recall test. The stimuli were drawn from a pool of 1182 German abstract and concrete nouns (591 each). Concrete nouns had a mean imageability rating of 5.65 (range = 4.54 – 6.89; on a 7-point scale) and abstract nouns of 3.16 (range = 1.35 – 4-45). The nouns were between three and ten letters long and had a mean frequency of 58.58/million (Vo et al., 2009). In both sessions, stimuli were drawn without replacement from the large pool of abstract and concrete nouns, and within a given trial all items were either abstract or concrete.

The sequence of an experimental trial is illustrated in **Figure *1***. Prior to completing the WM task, subjects were informed about the possible memory strategies people typically use during WM tasks, specifically about mere reading, rehearsal, refreshing, mental imagery, sentence generation, and grouping. Across four practice trials, they were familiarized with the task.

A fixation cross cued the beginning of a trial for 1000 ms. Thereafter, six words were presented sequentially for 500 ms with a 4500 ms blank inter-item-interval. The memoranda (font size = 24pt) were presented in boxes (size = 200 pixels) equidistantly arranged on an invisible circle centered on the screen in clockwise order, starting at the top box. After the blank following the sixth word, the WM test followed: A box appeared in the middle of the screen prompting subjects to type in the words one-by-one in forward serial order. Participants did not receive feedback regarding memory accuracy after the WM trials.

We chose this task because serial recall is one of the most common measures of working memory, as it entails the critical binding component (binding words to serial positions) which has been shown to underly the capacity limit of working memory (Oberauer, 2019b; Wilhelm et al., 2013). We choose the unusually long inter-item time to create optimal conditions in which strategies could be effectively employed and their putative beneficial effects observed. Tan and Ward (2008) and recent replications (Souza & Oberauer, 2018,

2020) using overt rehearsal protocols showed that with the more common timing of 1 second per item participants hardly engaged in rehearsal. On about half the occasions, they did not say any words aloud, and in the remaining cases they merely repeated the last word rather than cumulatively rehearsing all previous words. This led us to choose a slow presentation rate that in previous studies (Tan & Ward, 2008, 1 item/5sec – slow presentation condition; and Souza and Oberauer, 2018; 1 item/5 seconds – slow condition) induced substantial rates of self-selected cumulative rehearsal. Using faster paces would severely limit the use of rehearsal, and probably also other demanding maintenance strategies such as sentence generation and imagery, thereby biasing this study to fail to observe their potential beneficial effects on memory.

In Session 1, after recalling all six items, subjects were prompted to indicate which strategies they had engaged in during the past trial by clicking on the respective option. The participants were allowed to choose only one strategy per trial. The options were (with the original label in German): (1) reading (*lesen)*, (2) rehearsal (*wiederholen*), (3) refreshing, described as directing attention to the memoranda (*Aufmerksamkeit auf Gedächtnisinhalte legen*), (4) mental imagery (*bildlich vorstellen*), (5) sentence generation (*Satz generieren*), (6) grouping (*gruppieren*), (7) other (*eine andere Strategie*), and (8) nothing (*nichts*). The strategy prompts and the strategy descriptions provided at the beginning of the experiment were adapted from Dunlosky and Kane (2007), except for the refreshing instruction, which was based on descriptions provided to us by 12 experts in the field of refreshing[2]. The (translated) general descriptions can be found in Table 1[3].

In Session 2, prior to the start of the sequential presentation of memoranda, subjects

---

[2] We thank the attendees of the workshop: “The Crossroads of Attention in Working Memory: Consolidation, Refreshing, & Removal” for their descriptions provided to us anonymously.

[3] Note that our descriptions of rehearsal and mental imagery were general, including any possible implementation of these strategies (e.g., cumulative rehearsal or other rehearsal schedules; static images or an interactive image).

were instructed to use one of the six memory control processes for the upcoming mini-block of four trials (see Table ***1*** for the mini-block instructions of the strategies). After recalling the six words, the trials in Session 2 ended with a screen asking subjects whether they had complied with the instructions, and if not, they were prompted to indicate which strategies they used instead, using the same eight-alternative questionnaire as in Session 1.

Figure 1

Panel A. Illustration of the Working Memory Paradigm. A) Participants were shown a list of six words sequentially across different boxes on the screen and tested with typed recall in forward serial order. B) In Session 1, after recall of each list subjects were prompted to indicate which of the memory strategies, they had engaged in. C) In Session 2, a prompt like the one depicted in the panel instructed subjects which strategy to use during the upcoming trial. Here one example condition (reading) is depicted; instructions of the other strategies in Table 1 used an analogous screen display. After recall participants were asked whether they had followed this instruction during the past trial.

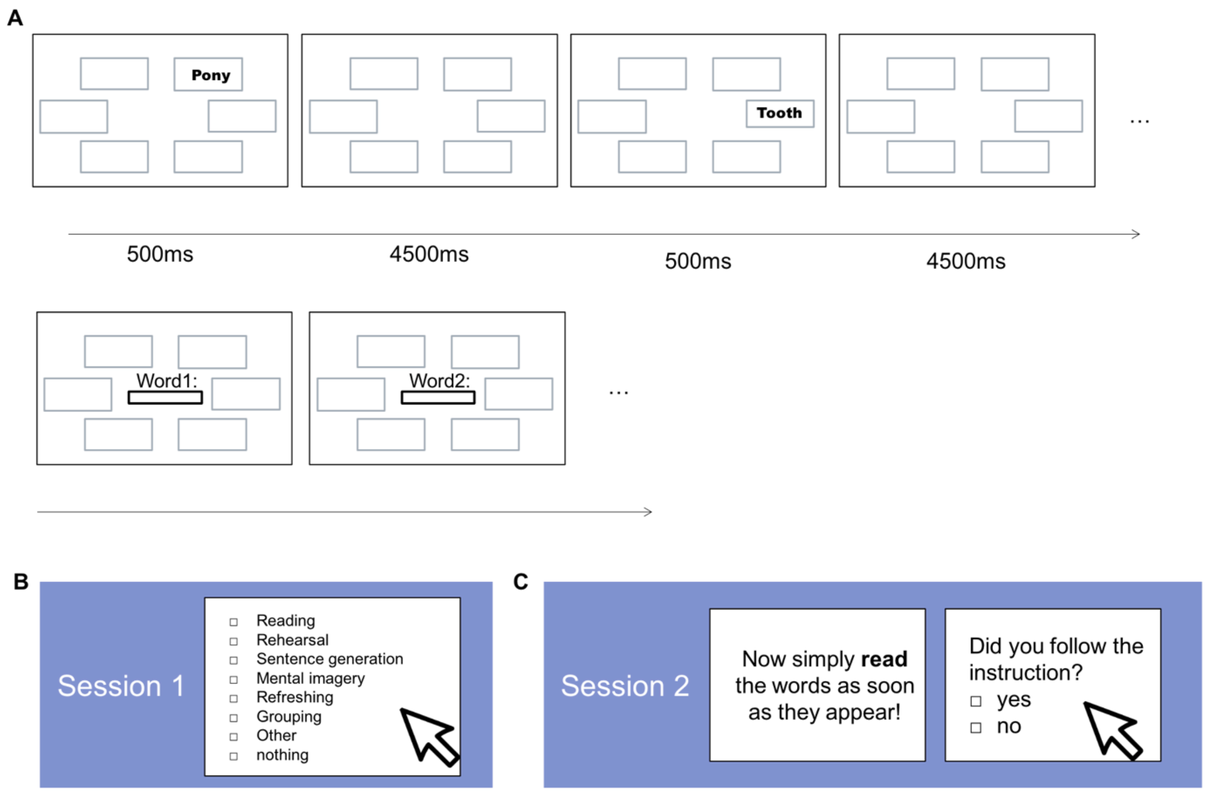


Table 1

Strategy Descriptions and Instructions used in Session 1 and 2 of Experiment 1 and 2.

Strategy	Description provided in Session 1	Mini-block instruction in Session 2
reading	You just read the words as they appeared on the screen.	Simply read the words as soon as they appear on the screen.
rehearsal	You repeat the words as often as possible in your head or out loud.	Repeat the words in your head or aloud as many times as possible.
refreshing	You direct your attention back to all the words presented so far. You do *not* articulate the words, but only think quickly and briefly about each word.	Direct your attention back to all the words presented so far. You should *not* articulate the words, but just think quickly and briefly about each word.
mental imagery	You mentally imagine the words in one picture.	Imagine the words in one mental picture.
sentence generation	You form a sentence to link the words together.	Form a sentence using the words in order to link the words together.
grouping	You memorize the words in groups, for example, always in groups of 2 or memorize the first half of the list together and the second half together.	Memorize the words in groups, for example always in groups of 2 or the first half of the list together and the second half together.

In Session 1 there were 60 trials of the WM task, and Session 2 comprised 24 mini-blocks of 4 trials per strategy for a total of 96 trials. The order of strategies across mini-blocks was randomized for each participant. Half of the trials of each session consisted of concrete and the other half of abstract lists.

After the end of the WM task in each session, an unrelated distracter task followed in which participants had to indicate the correctness of visually presented math equations (e.g., 9 x 8 = 72) for 2 minutes. After that, there was a typed delayed free recall memory test, wherein the participants were asked to recall as many memory items from the WM task as possible. This test served to assess the effect of each WM strategy on episodic LTM. Participants were made aware of the delayed memory test before the start of the experiment.

Data Analysis

The dependent variables of interest were: (a) frequency of strategy self-reports in Session 1, and (b) serial recall accuracy (i.e., the proportion of words correctly recalled using strict correct in-position scoring) as a function of self-reported and instructed WM strategy. Further, we also investigated the trial-to-trial variability in self-reported strategy use by examining the transition probabilities of reporting one strategy given the strategy reported in the previous trial. We analyzed the data using Bayesian mixed effect models (LME) using the *lmBF* function implemented in the *BayesFactor* package (Morey et al., 2015) in R (R Core Team, 2017). With this we calculated Bayes Factors (BFs) which represent the strength of evidence for a specified model (M1) against a Null or reduced model (M0). For instance, we can calculate the evidence for the effect of *WM strategy* (BF_{10}) by comparing the evidence for a model including this factor against an intercept-only model that serves as the null model. Additionally, we can calculate evidence against an effect of WM strategy (BF_{01}), where $BF_{01} = 1/BF_{10}$. A BF_{10} larger than 1 gives evidence for an effect, a BF_{10} lower than 1 provides evidence against an effect and hence evidence for the null hypothesis. A BF_{10} of 10 indicates that the data are 10 times more likely under the alternative hypothesis than under the null hypothesis. Usually, BFs below 3 are regarded as *anecdotal* evidence and BFs > 3 are regarded as providing *substantial* evidence for one hypothesis over the other (Kass & Raftery, 1995). All models included a random intercept for participant, and random slopes over participants for the effects of the variables manipulated within-subjects (Barr et al., 2013; Oberauer, 2022a).

Results

We first report the results on two questions about strategy use: (1) Which strategies do subjects report to engage in spontaneously? (2) Do subjects comply with strategy instructions? Then we turn to the performance in the immediate memory task to answer our main research question: (3) Do strategies impact WM? By answering these questions, we test our two main hypotheses: (a) Strategy instruction produce a general dual-task cost in WM that prevents their benefits from being measured; and (b) strategy instruction prevents adaptive switching between strategies. Finally, we then turn to the question (4): Do the strategies also impact LTM performance?

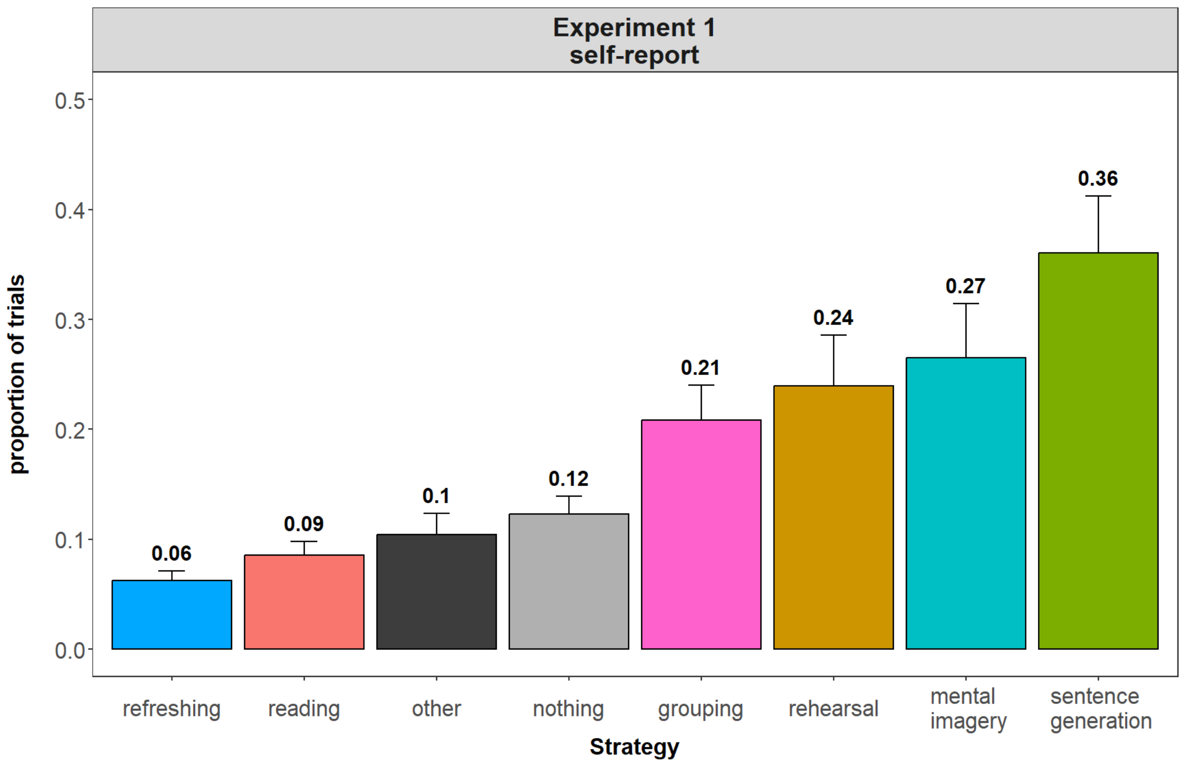
All data and analysis scripts can be accessed on the Open Science Framework (https://osf.io/m2hx3).

What Strategies do Subjects Spontaneously Use During Free Time?

Strategy Self-Reports. The mean proportion of trials in which each strategy was reported in Session 1 is shown in *Figure 2*. This variable was calculated by first computing the mean proportion of trials in which each strategy was reported by each subject, and then averaging across all subjects. As a first step, we were interested in the overall strategy occurrence and how they rank from most to least frequent. Here, the most reported strategies were sentence generation, mental imagery, rehearsal, and grouping, in that order. Other strategies, such as refreshing or reading, were very rarely reported.

Figure 2

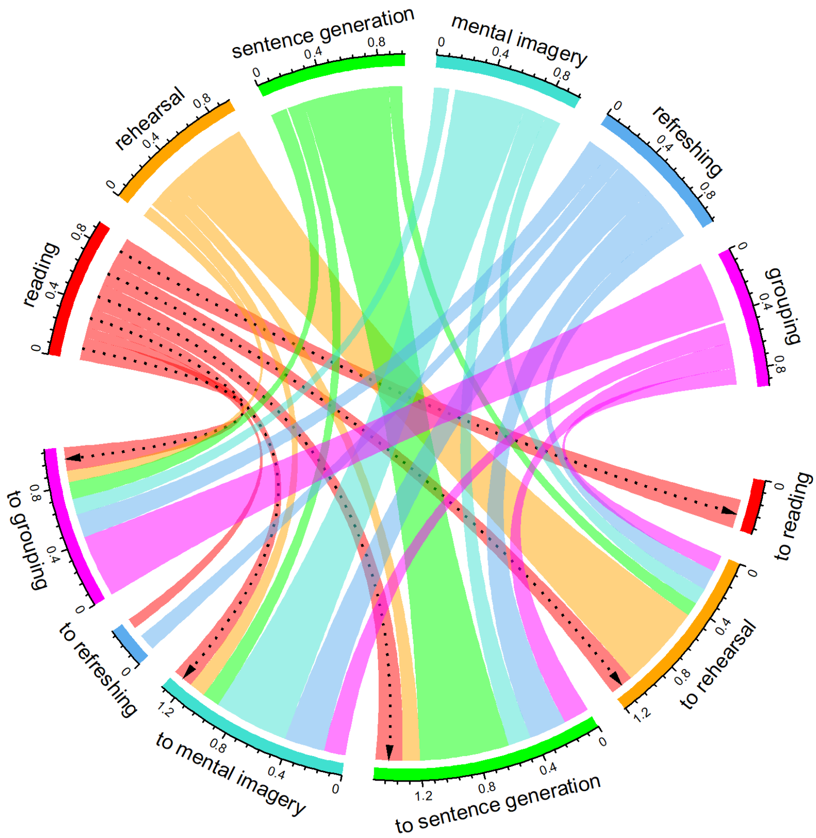
Mean Proportion of Self-Reported Strategies in Session 1 of Experiment 1. Error Bars Reflect the Standard Error of the Mean.



Regarding strategy differences between lists, concreteness increased the probability of choosing mental imagery and grouping (see Figure S9 in the online supplementary material for details). Rehearsal and reading were more likely to be chosen in abstract than concrete trials. We turn to the question of whether concreteness also affected the effectiveness of the instructed strategies below.

Figure 3

Chord Diagram of the Proportion of Transitions Between Self-Reported Strategies Across Trials of Experiment 1. The Links Represent the Transitions from the Strategy Reported in Trial n-1 at the Top to the Strategy Reported in Trial n at the Bottom. The Arrows Illustrate the Directionality of Transitions (i.e., from Top to Bottom) for the Strategy of Reading as an Example. The Link Width Represents Transition Frequency. The Total Length of the Strategy Sector at the Bottom Represents the Total Strength of the Connections to this Strategy (e.g., Fewer Connections to Refreshing and Reading).

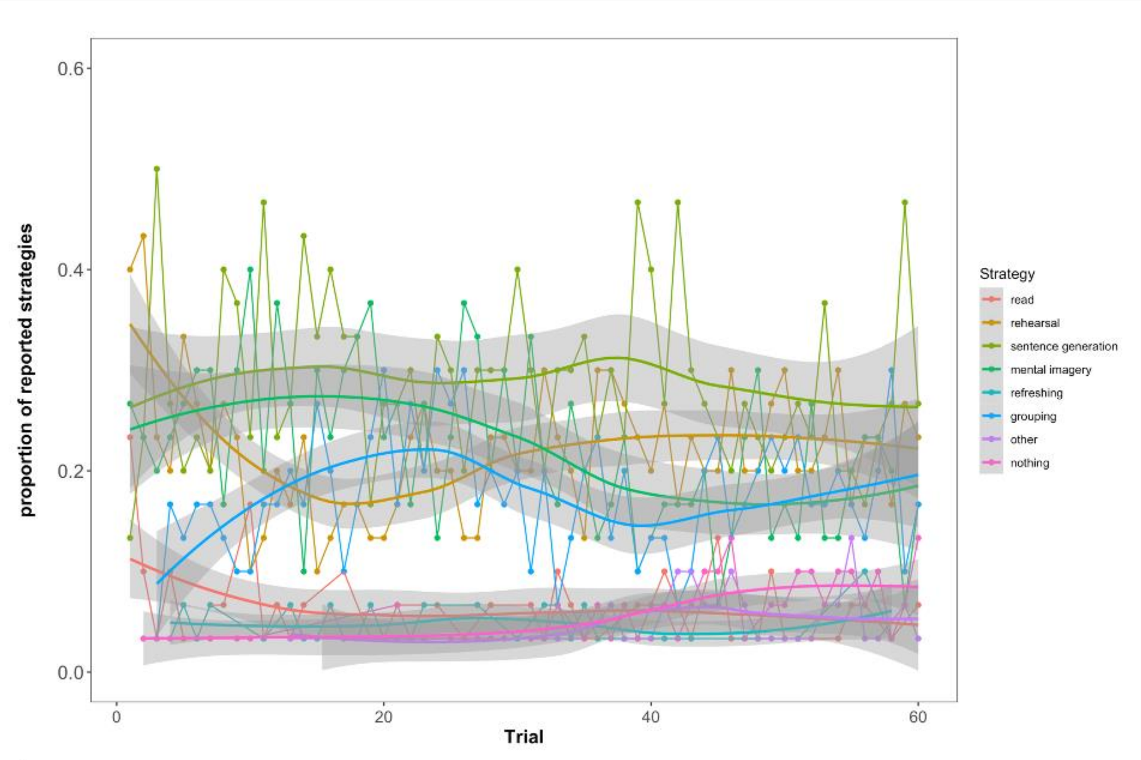


Strategy Switches Between Trials. Figure 3 shows the transition probabilities of reporting each strategy, given the strategy reported in the previous trial. This figure illustrates the trial-to-trial variability in strategy use. Subjects were generally more likely to switch than to stay with a certain strategy, with a mean switching frequency of 77.95% (SD = 11.72). The probability of switching is hereby calculated as the proportion of transitions (from trial n to n+1) with a switch out of all transitions. We also investigated whether trial-to-trial variability in strategy use might reflect changes induced by task exposition such as learning or fatigue.

Figure *4* shows the proportion of self-reported strategies over trials. The pattern indicates that the subjects showed neither patterns of fatigue nor learning.

Figure 4

Proportion of Reported Strategies as a Function of Trial in Experiment 1. The Grey Bands Represent the Smoothed Conditional Means Derived from Local Polynomial Regression Fitting (loess function).



Do Subjects Comply with Strategy Instructions?

One critical issue in instructing a strategy is to measure compliance with strategy use. In Session 2, we asked participants after each instructed-strategy trial whether they did in fact apply the memory control process instructed to them, and if not, what they were doing instead. Subjects indicated to comply with the instructions in 84.39% of the trials (SD = 24.22). The Bayesian linear mixed effects model revealed evidence against a difference in compliance across instructed strategies (BF10= 0.0002). In sum, self-reported compliance was very high; in the recall performance analyses, only compliance trials were included.

The Effect of Strategies on Immediate Serial Recall

Do strategies impact WM performance? Now we will turn to the functionality of maintenance strategies for WM. We examined whether there was evidence for an overall WM cost to implementing a strategy instruction by comparing WM performance in conditions in which a strategy was self-reported vs. instructed[4]. If there is a dual-task cost of implementing instructed processes, this should lead to a cost particularly for those strategies that are cognitively demanding (i.e., refreshing and elaboration), those strategies used rarely (i.e., reading, refreshing), or both.

We present analyses with mean serial recall accuracy, defined as proportion of words recalled in their correct serial position, as dependent variable. We had no a-priori hypothesis about the interaction of serial position with the effect of instructed compared to self-chosen strategies on immediate recall performance. Therefore, we have included those visualizations and analyses for the interested reader in the Online Supplementary Material.

Are There Dual-Task Costs of Implementing Instructed Strategies? ***Figure 5*** shows the immediate serial recall performance over strategies and sessions. The Bayesian LMEs revealed evidence for an interaction effect of session with strategy ($BF_{10} = 5.63$), as well as decisive evidence for both main effects (strategy: $BF_{10} = 1.85 \times 10^{24}$ and session: $BF_{10} = 1.42 \times 10^{6}$). There was overall worse performance in the instructed (M = 0.61; SD = 0.49) compared to the self-report session (M = 0.78; SD = 0.41). The main effect of strategy was driven by better performance for grouping, mental imagery and sentence generation compared to reading, rehearsal, and refreshing (see Table 2 for all pairwise comparisons).

---

[4] Our Experiments entailed more trials for the instructed strategy session (96 trials) compared to the self-report session (60 trials). To investigate whether that differences might have influenced our results, we ran all critical analyses also on data including only the first 60 trials of the instructed session. As shown in the Supplementary Material, the pattern of results remained the same.

To answer whether there are indeed dual-tasks costs of implementing instructed strategies, we turned to the interaction effect: Post-hoc comparisons of the interaction effect revealed that there was no difference between self-reported and instructed WM performance in case of rehearsal and sentence generation (BF_{01} = 3.31 and BF_{01} = 3.05, respectively). As apparent in ***Figure 5*** by the large error bars, reading and refreshing were very rarely self-reported, leading to more uncertain performance estimation. Still, choosing these low frequency strategies resulted in better performance than when they were instructed (BF_{10} = 35.09 and BF_{10} = 15.54, respectively). Likewise, both mental imagery and grouping resulted in better WM performance when these strategies were self-reported than instructed (BF_{10} = 56.98 and BF_{10} = 11.99, respectively).

Table 2

Bayes Factors of the Pairwise comparisons of the Main Effect of Strategy on WM performance (Collapsed across the Self-Reported and Instructed Sessions) in Experiments 1 and 2.

	refreshing		rehearsal		grouping		mental imagery		sentence generation	
	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2
read	0.14	0.02	9.42×10^{6}	2.25×10^{5}	5.05×10^{10}	1.42×10^{7}	2.86×10^{11}	6010	2.79×10^{12}	3.91×10^{15}
refreshing			1.35	7102	589	1.88×10^{6}	4490	156	7736	1.72×10^{14}
rehearsal					12.81	0.01	78.54	0.026	1.62×10^{4}	94.96
grouping							0.12	0.02	0.02	31.06
mental imagery									0.28	105

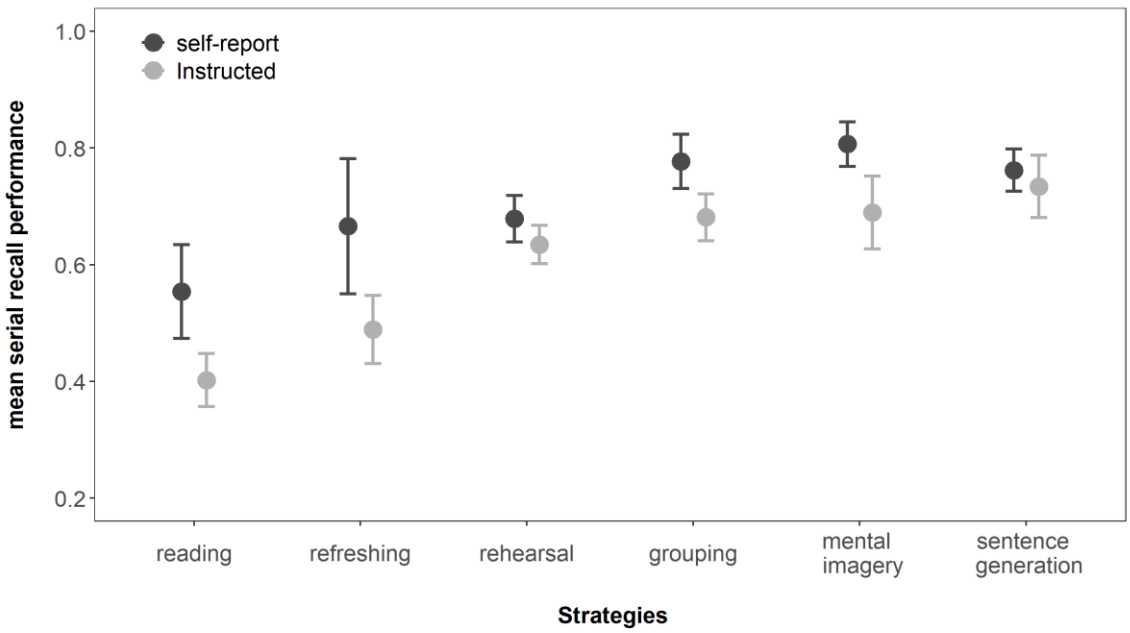
Note: Bayes Factors > 3 represent substantial evidence for credible differences in performance. As the strategies are ordered from least to most successful, in case of a credible difference, performance is better for the strategies listed in the columns compared to the strategy in the Rows. BF < 1/3 reflect substantial evidence against a difference between both strategies.

Taken together, the results speak against the hypothesis that instructing strategies incurs a general dual-task cost, as there was no sign of such a cost for one of the most demanding strategies, sentence generation, whereas there was evidence for a cost for the least demanding strategy, reading. We also did not find systematically larger dual-task costs for rarely chosen, and hence less practiced, strategies, as one of the most frequent strategies, mental imagery, showed poorer performance when instructed than when self-chosen.

Instead, our results are better explained by our second hypothesis, namely that instructing one specific strategy impedes adaptive strategy choice and switching. This could explain why experimental studies that manipulated the engagement in some of these strategies across all trials have observed no evidence that they improve memory above conditions in which subjects are free to choose their own strategy.

Figure 5

Mean Immediate Serial Recall Performance Across Strategies and Session (Self-report vs. Instructed) of the Data of Experiment 1. Error bars represent 95% Confidence Intervals.



Does the Choice of Strategy Make a Difference Within a Person? The analysis above revealed evidence for a main effect of strategy, indicating that some of them were more beneficial to immediate serial recall performance than others. This could be because participants who are better at serial recall tend to choose certain strategies (e.g., mental imagery) more often than participants whose memory is worse (a between-subjects effect), or it could be because some strategies are more effective for memory maintenance than others for the same person (a within-subjects effect). The latter effect is of primary interest for assessing whether strategies make a difference for memory performance. To separate within-subjects from between-subjects covariation, we calculated the effect of each strategy relative to the individuals' mean performance.

Pairwise comparisons showed better within-subject performance for self-reported mental imagery (BF_{10} = 8.30) and worse performance for self-reported reading and rehearsal (BF_{10} = 31.57 and BF_{10} = 5.85, respectively). For self-reported sentence generation, grouping, and refreshing there was ambiguous evidence for a difference compared to the individuals' mean performance (BF_{10} = 0.63, BF_{10} = 0.73, and BF_{10} = 0.37, respectively).

Taken together, the results of Experiment 1 hint at mental imagery being beneficial to WM performance compared to the other memory control processes – even when considering the individuals' mean performance. Processes like reading and rehearsal instead lead to worse performance within individuals. Therefore, based on the findings using an open item pool there seems to be some self-chosen strategies that yield benefits whereas others produce costs within a person.

Which Instructed Strategy is More Beneficial? Next, we were interested in the effect of instructed strategies on WM recall. There was decisive evidence for a main effect of instructed strategy ($BF_{10} = 3.16 \times 10^{24}$). Table 3 presents BFs of the follow-up pairwise comparisons of the instructed strategies in Session 2. These effects revealed that instructing reading led to worse performance than all other strategies, and instructed refreshing led to worse performance than rehearsal, grouping, mental imagery, and sentence generation. Instructed rehearsal, grouping, and mental imagery led to similar performance. Instructed sentence generation surpassed performance of instructed rehearsal, but yielded equivalent immediate recall scores as grouping and mental imagery. [5]

To compare these effects of instructed strategies to a "free baseline", as done in previous studies (Bartsch et al., 2018; Bartsch & Oberauer, 2021; Souza & Oberauer, 2018, 2020), we next compared performance with each instructed strategy to the mean serial-recall performance from the self-report session (Session 1). None of the instructed strategies surpassed the mean serial-recall performance of the session with free strategy choice, consistent with previous literature. Instructed reading ($BF_{10} = 2.94 \times 10^7$), refreshing ($BF_{10} = 2.21 \times 10^4$), and rehearsal ($BF_{10} = 133$) led to worse performance than in Session 1, and instructed sentence generation yielded equivalent performance to the mean performance in Session 1 ($BF_{01} = 4.65$). There was anecdotal evidence against a difference in case of instructed mental imagery and grouping ($BF_{01} = 2.35$, $BF_{01} = 1.42$ respectively).

Table 3

Bayes Factors of the pairwise comparisons of instructed-strategy effects on WM performance in Session 2 of Experiment 1 and 2.

	refreshing		rehearsal		grouping		mental imagery		sentence generation	
	E1	E2	E1	E2	E1	E2	E1	E2	E1	E2
read	2.48	4.62	5.44 × 10^7	2.52 × 10^7	1.63 × 10^8	2.15 × 10^8	7.26 × 10^6	7.85 × 10^8	1.46 × 10^8	5.99 × 10^{15}
refreshing			218	742	1.14 × 10^4	1.31 × 10^4	2719	2542	8.96 × 10^4	1.43 × 10^{10}
rehearsal					0.60	0.24	0.60	0.31	6.99	1317
grouping							0.35	0.15	0.54	2333
mental imagery									0.54	4718

Note: Bayes Factors > 3 represent substantial evidence for credible differences in performance. As the

[5] One could expect that the concreteness of the to-be-remembered words affected how beneficial a certain strategy would be for immediate memory performance. This analysis can be found in the Supplementary Material.

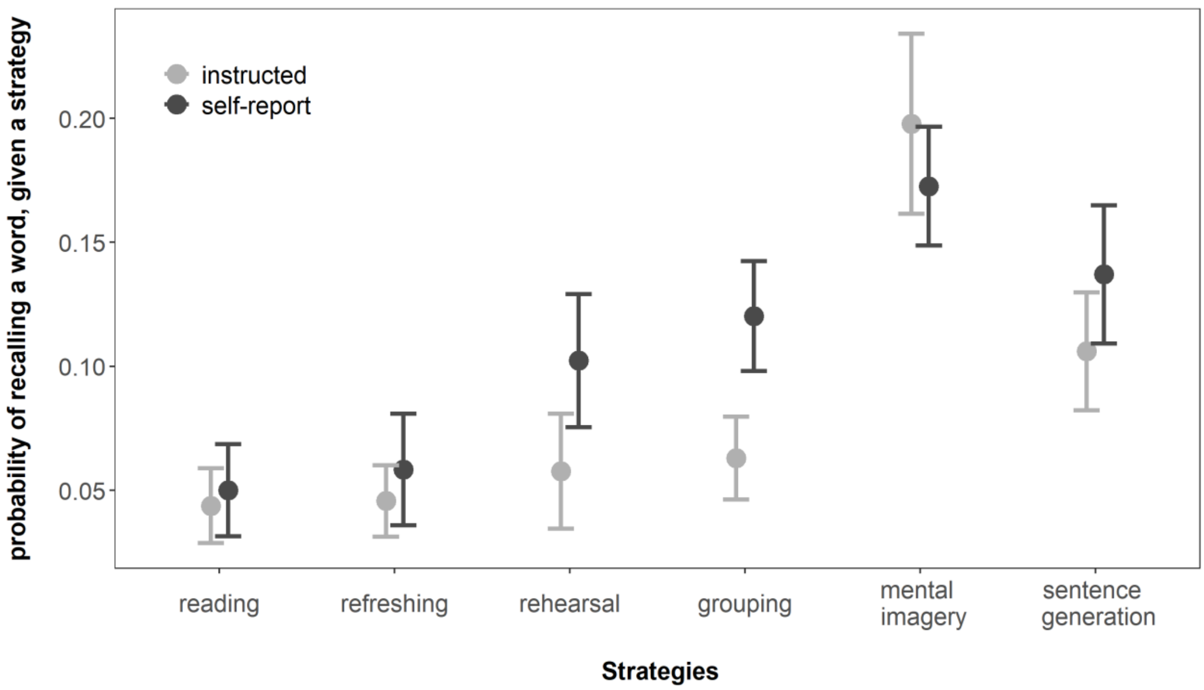
strategies are ordered from least to most successful, in case of a credible difference, performance is better for the strategies listed in the columns compared to the strategy in the Rows. BF < 1/3 reflect substantial evidence against a difference between both strategies.

Do WM Strategies Impact LTM Recall?

When a memory strategy is beneficial for immediate serial recall, this could be because it improves maintenance in WM, but also because it improves the accessibility of the memoranda in LTM. As we know from past literature, elaboration benefits LTM (Craik & Tulving, 1975), whereas other strategies – like rehearsal and refreshing – have been shown to have little impact on LTM (Bartsch et al., 2018; Greene, 1987). Therefore, it is possible that elaboration improves immediate recall indirectly through creating more accessible traces in episodic LTM.

Figure 6

Mean Probability of Correctly Recalling a Word in the Delayed Free Recall Test as a Function of Strategy and Session (Self-report vs. Instructed) relative to the Frequency of each Strategy in Experiment 1 – with an Open Pool of Items. The Order of the Strategies Represents the Effectiveness in the Immediate Serial Recall task (from least to most effective). Error bars Represent the Standard Error of the Mean.



Experiment 1 included a delayed free-recall test to assess how WM strategy use affected the formation of LTM representations. We calculated the probability of correctly recalling a word in the delayed test per session and strategy. For the Session 1 data, for each participant we calculated the sum of correctly recalled words per strategy and divided this by the number of words across the trials in which this participant chose this strategy in the immediate serial recall task. This gave us the proportion of correct delayed memory recall out of the number of words for which that strategy had been used by each participant, which estimates the probability of recalling a word, given that it has been encountered with a particular strategy. In Session 2, the number of trials per instructed strategy was constant, hence we simply calculated the proportion of correctly recalled words per strategy (see **Figure *6***).

The Bayesian LME revealed evidence for a main effect of strategy ($BF_{10} = 2.83 \times 10^9$). Reading and refreshing were ineffective in promoting long-term recall of the words. Rehearsal and grouping ended up in the middle. By contrast, the two forms of elaboration, that is sentence generation and mental imagery, led to better long-term learning than the other strategies (see **Table *4*** for all BFs of the pairwise comparisons). Furthermore, there was evidence against an effect of whether the process was instructed vs. self-reported (BF_{01} = 3.99), as well as strong evidence against an interaction (BF_{01} = 13.05). Although not critical for our research question, the evidence against a main effect of whether the process was instructed vs. self-reported needs to be taken with caution, as we are comparing across delayed free recall with different numbers of trials in the two sessions (60 vs. 96) and therefore, if there were any differences, they could be due to a list length effect.

We expected that the concreteness of the to-be-remembered words affected how beneficial some of the strategies – in particular, mental imagery – would be for delayed memory performance. Our analysis in the supplementary material shows that this was the case: Concrete words benefited more from imagery, and also from sentence generation and grouping.

Table 4

Bayes Factors of the pairwise comparisons of strategy effects on LTM performance collapsed across sessions of Experiment 1. Bayes Factors > 3 represent substantial evidence for better performance in the strategies listed on the top compared to the ones listed on the left. Bayes Factors below 0.33 represent substantial evidence against a difference.

	refreshing	rehearsal	grouping	sentence generation	mental imagery
read	0.03	0.06	0.72	229	2.66×10^6
refreshing		0.01	0.17	62.03	4.00×10^5
rehearsal			0.02	0.28	1484
grouping				0.14	2512
sentence generation					10.94

Discussion

Strategy Selection

As a first step, we were interested in which strategies subjects spontaneously report using during a long free time interval between individual memory items. We have shown that subjects engage in some strategies more frequently than others: sentence generation and rehearsal appear as frequent choices, replicating previous findings from self-report studies (Bailey et al., 2011; Dunlosky & Kane, 2007) - although most previous studies showed more rehearsal and less elaboration. This difference can easily be explained by previous studies giving less time in-between items than we did here. Furthermore, mental imagery and grouping were also frequent choices whereas others, like refreshing were rarely or never reported (see also AuBuchon & Wagner, 2023; Loaiza & Lavilla, 2021). Further, concreteness affected the choice of strategies, with subjects choosing to engage in mental imagery and grouping more often in case of concrete words, whereas rehearsal and reading were more frequent in abstract lists. The increase on the usage of imagery was expected, because concrete words lend themselves more to mental imagery. For the effect on choosing grouping, we have no explanation.

Furthermore, we assessed the variability in strategy selection from trial-to-trial. Overall, subjects were more likely to switch than to stay within the same strategy. This hints at an adaptive choice of strategies, potentially in reaction to the material of a certain trial being easier to form a sentence, mentally imagine in a joint picture, or easier to group than to rehearse. These results suggest that reductions in adaptive strategy switching are a viable explanation for the lack of benefits in instructed strategy studies.

Finally, we observed that subjects complied to a high degree with the strategy instructions in both experiments. This observation converges with previous studies finding that young adults can implement an instructed rehearsal schedule (Souza & Oberauer, 2018, 2020) and can implement the instruction to elaborate the memoranda (Bartsch et al., 2019; 2021). These results show that the choice of a memory control process is under volitional control. Therefore, a lack of compliance with instructions cannot explain the past lack of evidence for any of the candidate processes benefitting WM in comparison to a free baseline.

The Effect of Strategies on WM Performance

Previous studies assessed the causal effect of instructing maintenance strategies compared to a free baseline in which subjects were free to use whatever strategy they preferred, and observed no beneficial effects of instructing a strategy (Bartsch & Oberauer, 2021; Souza & Oberauer, 2018, 2020). Here we replicated this observation when we compared the performance of each instructed strategy to the individuals' mean performance across all self-report trials: None of the instructed strategies surpassed the performance in a free baseline condition. We investigated whether this could be explained by (1) secondary task demands that the implementation of a strategy instruction potentially entails, or (2) instructions preventing participants from adaptively choosing different strategies from trial to trial.

We examined this question by comparing serial recall performance in trials in which a WM strategy was self-reported vs. instructed. We observed that for some strategies (i.e., reading, refreshing, mental imagery, and grouping), there was a detrimental effect of instruction implementation. For the remaining two strategies, namely sentence generation and rehearsal, performance was similar between self-reported and instructed conditions.

Taken together, these results do not support the hypothesis that instructed strategies impose a dual-task cost. Such dual-task cost would either be equal for all strategies – if it simply reflected the cost of maintaining and implementing any instruction – or it would be highest for the cognitively most demanding strategies, namely refreshing, imagery, and sentence generation, and negligible for undemanding strategies such as reading. The observed pattern of costs does not match either of these predictions. A third possibility was that the costs would be lowest for the most familiar strategies. Whereas there was no instruction cost for rehearsal and sentence generation – which are common self-reported strategies, instructed reading produced costs although this is probably the most familiar activity for our participants. Hence, all in all, the dual task hypothesis cannot explain our data.

Instead, the results are better explained by instructions preventing participants from adaptively choosing different strategies from trial to trial. Sentence generation and rehearsal could be applied more successfully than other strategies to a relatively large number of trials – as indicated by their higher frequency in self-reports. Therefore, instructing participants to use these strategies consistently does not force them to use an unsuitable strategy on many trials, and therefore does not lead to a performance decrement. The other strategies are not helpful on most of the trials in which participants were instructed to use them. Therefore, using them only on suitable lists is more beneficial than using them consistently on a fixed number of lists, as demanded in the instruction condition. Comparing performance with different instructed strategies, we found that mental imagery and sentence generation yielded benefits for immediate serial recall performance compared to instructed reading, refreshing and, partly, rehearsal. Yet, these two processes were also the ones which led to the most long-term learning. Therefore, it could be that mental imagery and sentence generation did not *directly* benefit WM, and instead these processes are helpful because they allow people to rely on episodic LTM traces, which drives both the immediate serial-recall performance benefit and the higher LTM performance. In order to independently test this possibility, we conducted Experiment 2. This experiment further allowed us to test the replicability of our results.

Experiment 2

We designed Experiment 2 with two purposes in mind. First, it served as a replication to test our main hypotheses regarding the impact of strategies on WM. Second, it served to test the possibility that elaborative strategies benefited WM in Experiment 1 via the use of episodic LTM. Experiment 2 implemented the same design and task as in Experiment 1, with one difference: Whereas in Experiment 1 we used a large open set of words as memory items, in Experiment 2 we used a small, closed set of memory items which were repeated in random order from trial to trial. The repeated use of the same small set of words makes it harder to distinguish successive trials from each other, and increases the demand to remember in which particular order the words have been presented in the current trial. This increases proactive interference between trials in episodic LTM, and therefore, participants in Experiment 2 were less able to rely on episodic memory for immediate serial recall, potentially curtailing the benefits of elaborative strategies. If the beneficial effect of elaboration is mediated through improved episodic LTM, then that effect should largely disappear in Experiment 2. Further, we assessed whether participants adapted their strategy use to the fact that they could rely less on episodic LTM to complete the WM task. We expected that an adaptive strategy choice should lead to a reduction of elaborative strategies in Experiment 2, because elaboration is a particularly effective strategy for improving episodic memory.

Method

Participants

Participants were 30 students of the University of Zurich (26 females, 4 males; mean age = 22.13).

Materials and Procedure

The materials and procedure were the same as for Experiment 1, with two differences. First, for each session of Experiment 2, six abstract and six concrete nouns were drawn from the large pool once for each subject, and stimuli for each trial in that session were randomly drawn only from this small pool of 12 words irrespective of concreteness. Second, there was no final free recall test because it would be rather uninformative to ask whether people could recall the 12 words that they had seen throughout all serial-recall trials.

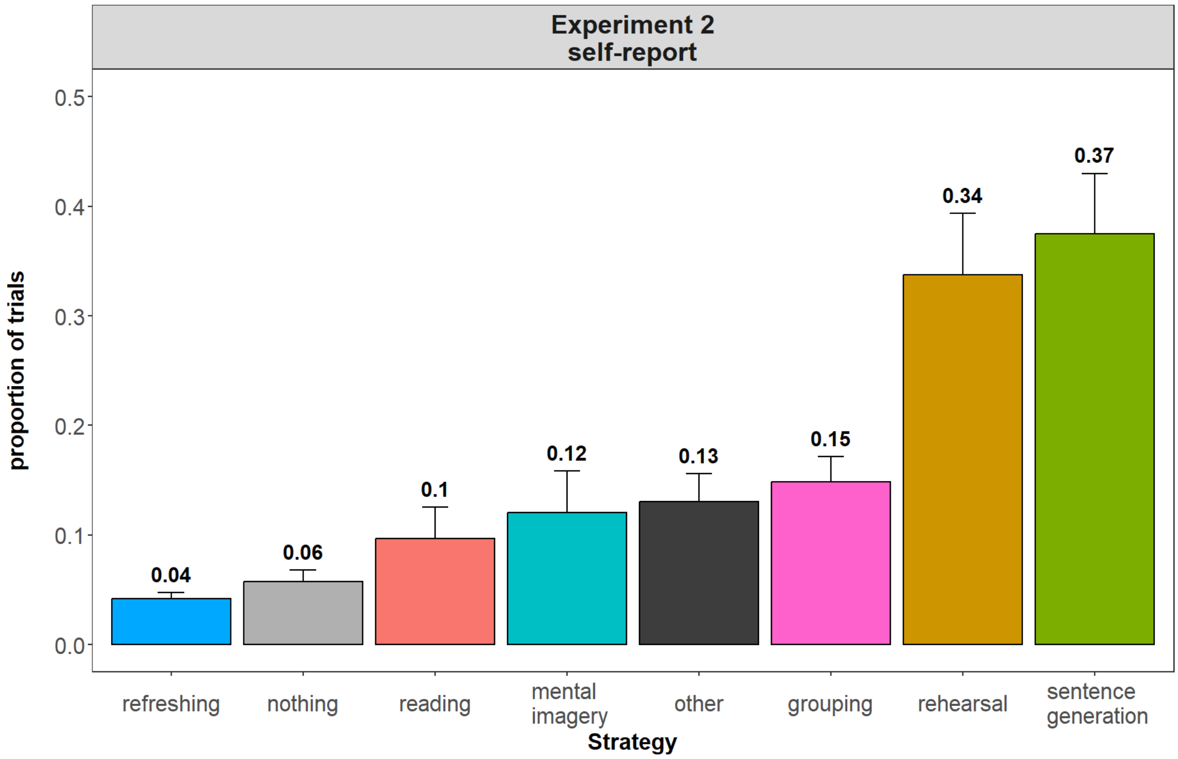
Results

What Strategies do Subjects Spontaneously Use During Free Time?

Strategy Self-Reports. The mean proportion of trials in which each strategy was reported is shown in Figure 7. Compared to Experiment 1, there was a strong drop in the use of mental imagery and grouping, as well as an increase in rehearsal. This partially confirms our prediction that with larger proactive interference between trials, elaborative strategies become less useful, and are used less often. Yet, sentence generation remained the most used strategy, contrary to our prediction. The rare reporting of refreshing and reading we saw in Experiment 1 persisted in Experiment 2.

Figure 7

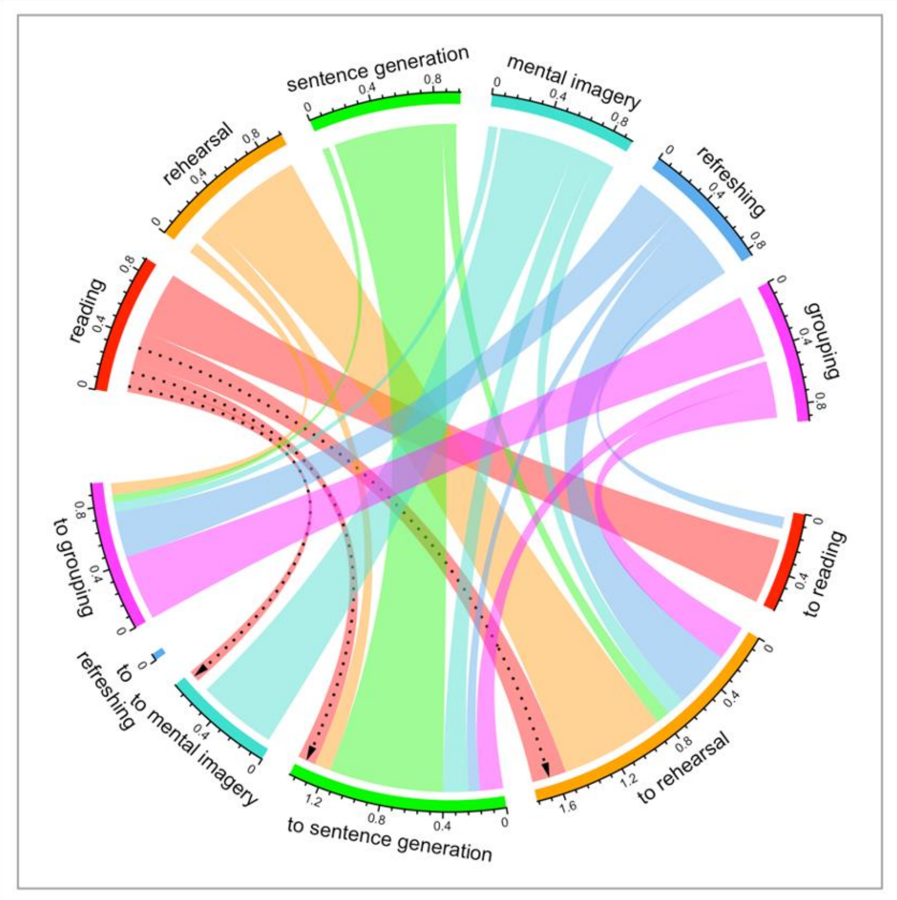
Mean Proportion of Self-Reported Strategies in Session 1 of Experiment 2. Error Bars Reflect the Standard Error of the Mean.



Strategy Switches Between Trials. ***Figure 8*** shows the transition probabilities of reporting one strategy given the strategy reported in the previous trial. Replicating the findings of Experiment 1, participants were generally more likely to switch than to stay with a certain strategy, with a mean switching frequency of 72.54 % (SD = 17.00). We investigated whether trial-to-trial variability in strategy use might reflect changes induced by task exposition such as learning or fatigue. Figure 9 shows the proportion of self-reported strategies over trials. The pattern indicates that the subjects seem to learn that sentence generation is a beneficial strategy, leading to a trade off in choosing it over rehearsal as a strategy.

Figure 8

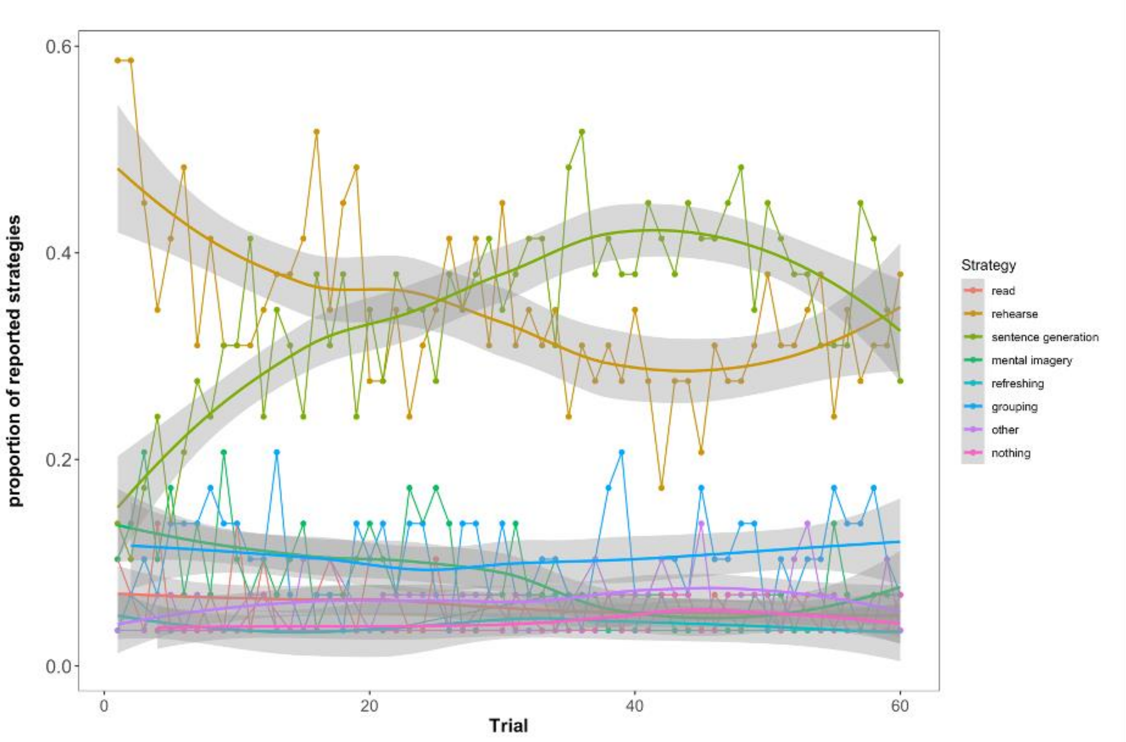
Chord Diagram of the Proportion of Transitions Between Self-Reported Strategies Across Trials of Experiment 2. The Links Represent the Transitions from the Strategy Reported in Trial n-1 at the Top to the Strategy Reported in Trial n at the Bottom. The Arrows Illustrate the Directionality of Transitions for the Strategy of Reading as an Example. The Link Width Represents Transition Frequency. The Total Length of the Strategy Sector at the Bottom Represents the Total Strength of the Connections to this Strategy (e.g., Fewer Connections to Refreshing and Reading).



Do Subjects Comply with Strategy Instructions? Similar to Experiment 1, self-reported compliance to strategy instructions in Session 2 was very high (94.37 %; SD = 18.16). The Bayesian linear mixed effects model revealed anecdotal evidence for a main effect of instructed strategy (BF_{10}= 2.64). Follow-up analyses revealed that compliance was lower when the instruction was to implement refreshing (M = 89.58 %; SD = 15.58) compared to rehearsal (M = 96.04 %; SD = 12.44; BF_{10}= 1.88), sentence generation (M = 95.20%; SD = 13.31; BF_{10}= 2.01), and grouping (M = 96.66%; SD = 12.39; BF_{10}= 3.61). In the recall performance analyses, only compliance trials were included.

Figure 9

Proportion of Reported Strategies as a Function of Trial in Experiment 2. The Grey Bands Represent the Smoothed Conditional Means Derived from Local Polynomial Regression Fitting (Loess Function).



Effects of Strategies on Immediate Serial Recall

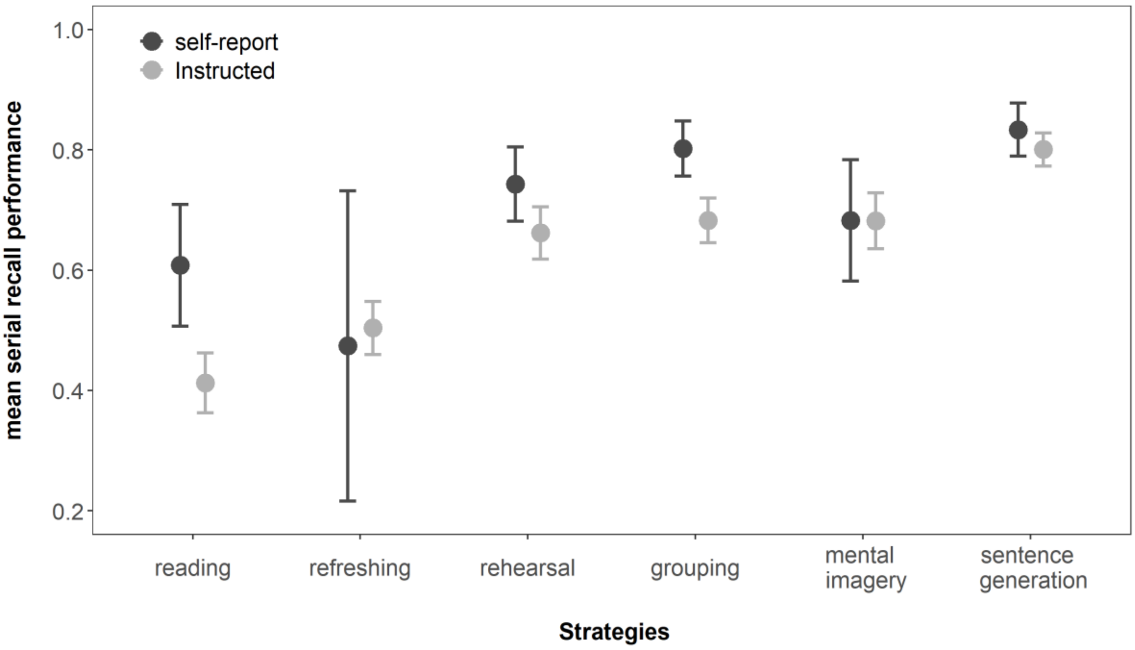
Are There Dual-Task Costs of Implementing Instructed Strategies? ***Figure 10*** shows the immediate serial recall performance over strategies and sessions. The Bayesian LMEs revealed anecdotal evidence for an interaction of session with strategy (BF_{10} = 2.87), as well as decisive evidence for both main effects (strategy: BF_{10} = 5.19 x 10^{23}; session: BF_{10} = 432). Replicating Experiment 1, there was overall worse performance in the instructed (M = 0.64; SD = 0.48) compared to the self-report session (M = 0.78; SD = 0.41). Further replicating Experiment 1, the main effect of strategy was driven by better performance for grouping, mental imagery and sentence generation compared to reading and refreshing (see Table 4 for all pairwise comparisons).

There was evidence against a recall difference between self-reported vs. instructed mental imagery and sentence generation (BF_{10}= 3.00 and BF_{10}= 4.24, respectively). Instruction resulted in worse performance than self-report for reading (BF_{10}= 245) and grouping (BF_{10}= 10.33), whereas evidence was ambiguous for rehearsal (BF_{10}= 1.54).

Taken together, results of Experiment 2 replicate our main findings of Experiment 1. Thus, Experiment 2 yields further evidence against the notion that instructing strategies incurs a general dual-task cost. Again, our results are better explained by our second hypothesis, namely that instructing one specific strategy impedes adaptive strategy choice and switching.

Figure 10

Immediate Serial Recall Performance Across Strategies and Session (Self-report vs. Instructed) of Experiment 2. Error bars represent 95% Confidence Intervals.



Does the Choice of Strategy Make a Difference Within a Person? Next, we turned to separating within-subjects from between-subjects covariation, and we compared performance with each strategy to the individual's mean performance. Pairwise comparisons showed worse performance in reading trials (BF_{10}= 11.97). For self-reported refreshing, sentence generation, and mental imagery, evidence was inconclusive regarding differences from an individual's mean performance (BF_{10}= 0.88, BF_{10} = 1.15, and BF_{10}= 0.81, respectively). For rehearsal and grouping, there was some evidence against a difference from an individual's mean performance (BF_{10}= 3.33, and BF_{10}= 3.22).

In summary, the results of Experiment 2 provide no evidence for any strategy being beneficial to WM performance compared to an individuals' mean performance. Reading instead leads to worse performance within individuals. Further, in Experiment 2 the benefit of mental imagery vanished, as we predicted from the assumption that the benefit of mental imagery is mediated through episodic memory. In a situation that generates a high degree of proactive interference between trials in episodic memory, and therefore affords little reliance on episodic LTM, there appears to be no strategy that participants can make use of to boost their memory relative to what they can accomplish with another strategy.

Which Instructed Strategy is More Beneficial? Next, we were interested in the effect of instructed strategies on WM recall. There was decisive evidence for a main effect of instructed strategy ($BF_{10} = 1.42 \times 10^{22}$). Table 3 presents BFs of the follow-up pairwise comparisons of the instructed strategies in Session 2. These effects fully replicated the findings of Experiment 1, with the exception that instructed mental imagery here yielded worse performance than sentence generation, and that instructed sentence generation now led to better performance than all the other strategies.

To compare these effects of instructed strategies to a "free baseline", as done in previous studies (Bartsch et al., 2018; Bartsch & Oberauer, 2021; Souza & Oberauer, 2018, 2020), we next compared performance with each instructed strategy to the mean serial-recall performance from the self-report sessions (Session 1). Here, instructed reading ($BF_{10} = 2.42 \times 10^9$), refreshing ($BF_{10} = 4.09 \times 10^6$), rehearsal ($BF_{10} = 428$), mental imagery ($BF_{10} = 106$), and grouping ($BF_{10} = 33.79$) produced worse performance compared to Session 1, whereas instructed sentence generation yielded anecdotal evidence for equivalent performance to the mean performance in Session 1 ($BF_{01}= 2.71$).

Discussion

We replicated the main findings of Experiment 1 with regards to strategy use.

Specifically, rehearsal and sentence generation appeared as frequent choices. One notable exception is that people reported using mental imagery less frequently than rehearsal or sentence generation. This observation provides partial support for the prediction we made at the end of Experiment 1: The beneficial effect of elaborative strategies such as mental imagery in Experiment 1 could be explained by the formation of stronger episodic LTM traces, which improved both immediate as well as delayed recall performance. With unique memoranda in every trial, mental imagery has a clear advantage because distinct LTM representations of the self-created images can be created for every episode, benefiting item memory. As unique episodes cannot be created with a closed pool of repeating items, subjects engage in this strategy less frequently in Experiment 2.

Yet why did sentence generation remain both popular and effective in the face of strong proactive interference? One explanation is that sentences – in contrast to mental images – preserve the order of the individual items, which is especially crucial in case the same items are repeated over the course of different trials (yet at varying serial positions). Forming a mental image does not preserve this order.

Further, by forming sentences of the memoranda, the level of proactive interference between trials can actually be reduced. Specifically, including a word in a specific role in a new sentence can increase the uniqueness of this episode, distinguishing this instance of the word from its many previous occurrences. For example, for remembering the words *Pony – Tooth – Coffee – Ball*, one could form the sentence: *The pony loses its tooth in the coffee after being hit by a white ball*. In the next trial, the memoranda could be *Coffee – Tooth – Ball – Pony*. Now, one can form the sentence: *Drinking my hot coffee hurts my tooth, but later I can play ball more energetically with my pony*. From the first example sentence to the second, the context of each of the words has changed: For instance, the coffee is either made undrinkable by a tooth or the person is drinking it in the morning. Therefore, sentence generation seems to facilitate both item memory and order memory, even under conditions of proactive interference, which makes it still effective for immediate serial recall.

To conclude, the results of Experiment 2 suggest that different forms of elaboration are differently susceptible to manipulations of proactive interference: Mental imagery becomes less effective, but sentence generation remains effective. The fact that proactive interference between trials modulates the effectiveness of elaboration strategies for immediate serial recall confirms our assumption that elaboration is helpful for performance on a WM test indirectly through boosting episodic LTM, which contributes to immediate recall.

These conclusions remain tentative because we cannot unambiguously attribute differences between Experiments 1 and 2 to the difference in proactive interference between these experiments. These experiments also differ in other aspects, such as the time of the year when participants were recruited, and the composition of trials (pure concrete or pure abstract word lists in Experiment 1, mixed concrete and abstract word lists in Experiment 2).

General Discussion

Based on theoretical assumptions about a causal role of memory control processes for maintenance in WM, as well as correlational evidence on the effectiveness of maintenance strategies on WM performance, previous experimental research has attempted to provide *causal* evidence that these processes benefit memory. These approaches were based on the hypotheses that instructing a specific strategy would induce individuals to engage in successful strategies *more* than they would spontaneously, or to switch from a less effortful strategy (e.g., reading) to one that requires more processing, but which leads to a larger mnemonic benefit (e.g., sentence generation). Yet, both approaches failed to provide evidence in favor of a beneficial effect of any of these strategies.

Our main aim for the present work was to examine reasons for the discrepancies between, on the one hand, correlational studies pointing to benefits of some maintenance strategies for WM performance and, on the other hand, experimental studies indicating no causal effect of these strategies on memory performance.

Dual-Task Costs Cannot Explain Why Strategy Instructions Don't Improve Memory

Previous experimental studies have compared performance under instructed strategy conditions to a no-instruction baseline, generally failing to find performance differences between them (Bartsch et al., 2018; Souza & Oberauer, 2018, 2020). One possible explanation for this outcome is that strategy instructions create dual-task costs. So far, no previous studies directly compared self-reported to instructed strategy use to test this possibility.

Across two experiments, comparisons of an instructed strategy condition to a free baseline including trial-to-trial strategy reports revealed no costs of instructing two commonly reported maintenance strategies, namely rehearsal and sentence generation. Other reported strategies such as reading, refreshing, mental imagery, and grouping showed instruction implementation-costs. In order to uphold the dual-task-cost explanation for the finding that experimental investigations using instructions never reached the performance of a free baseline, one would have to assume that mental imagery, reading and refreshing impose a dual-task cost, whereas sentence generation and rehearsal do not. In the introduction we discussed two variables that could moderate dual-task costs: The cognitive demand of a strategy, and the person's lack of experience with it.

These variables could explain why refreshing incurs a dual-task cost whereas rehearsal does not (Vergauwe et al., 2014) However, such an explanation is very implausible for the contrast between sentence generation and mental imagery. It is also very implausible that implementing the instruction to merely read each word as it is presented imposes a dual-task cost through increased cognitive demand because that instruction is the only one that does not even add a secondary task: Participants have to read the words to do the memory task anyway. Furthermore, participants should be rather experienced with reading, therefore a larger dual-task cost is unexpected.

Therefore, the finding that experimental investigations using instructions not showing a benefit for WM performance can't be plausibly explained by the instruction imposing a secondary task cost.

Adaptive Strategy Switching is Beneficial for Memory

The second possibility we addressed in the present study is that instructed strategies failed to benefit memory performance because they prevent adaptive switching between strategies. The recall performance, as well as the high frequency of switches between strategies reported in Session 1 of both Experiments 1 and 2, indicate that in a free-baseline, participants choose their strategy adaptively on a trial-by-trial basis. According to this explanation, for each strategy, the trials on which they self-reported a strategy are the ones for which this strategy is relatively successful – more successful than when the same strategy is applied to all trials. The difference to instructed trials becomes most apparent for the strategies that are rarely chosen, because in those cases, the majority of trials in the instructed-strategy condition are not particularly suitable for that strategy.

What drives the choice of strategies from trial to trial? By definition, a strategy is a process under the person's volitional control, and participants' ability to comply with our strategy instructions testifies to that. On what basis do they make their choices? A first question to ask in this context is whether they decide upon the strategy for the next trial already before presentation of the memory list (a form of proactive control; Braver, 2012), or only later, after having seen at least the first few items? In the latter case, it is plausible that the strategy is selected in part to be suited for the content of the memory list. For instance, some words, or sets of words, are easier to elaborate than others. This assumption motivated our variation of concreteness of list words in Experiment 1. Indeed, participants chose mental imagery and grouping more often on lists with concrete words than lists with abstract words (Supplementary Figure S9). Despite these differences in choice behavior, the effect of concreteness did not interact with the type of strategy chosen when it came to immediate memory performance. This fits with our observation that strategy switches also did not seem to be motivated by their recall performance in the preceding trial (see Online Supplementary Materials for details).[6] Future experiments could investigate the effects of other aspects of list items, or list composition, on strategy choice by varying these features of the memoranda systematically between trials.

As we kept the order of sessions constant to avoid any influence of the strategy instructions on the pattern of self-chosen strategies by the participants, one might wonder whether our results on the direct comparison between self-reported and instructed strategy use are compromised by this design choice. We argue that our findings cannot be explained by the constant session order (self-report first, instructed strategies second). Given that the two sessions were completed on different days, we can rule out fatigue effects across sessions. To rule out fatigue effects within a session – especially as the second session included more trials than the first – we analyzed performance as a function of trial, which yielded evidence *against* an interaction effect of session with trial number (Exp.1: BF_{01} = 4.24 and Exp. 2: BF_{01} = 649.65; see Online Supplementary Materials for details). Finally, the only remaining plausible effect of session order is a practice effect, which would have led to an advantage for performance in the instructed strategy condition in Session 2. In contrast, our results go in the opposite direction, showing that performance with instructed strategies (Session 2) was *worse* than when participants were to freely chose the strategy to use (Session 1).

Previous experiments instructing individual control processes show that using one strategy for all subjects and trials does not yield an overall benefit. The results across both experiments here revealed that choosing a strategy for each trial by each person leads to a

[6] Still, it could be that participants choose their strategy based on a more general meta-cognitive evaluation beyond their performance in trial n-1: They could choose strategies based on which of them they experienced as generally helpful or easy to implement, given the details of the task. For instance, the serial presentation of words around a circle could have facilitated the formation of a sentence out of the presented words, rendering that form of elaboration attractive.

benefit, relative to the choice of a single strategy by everyone on every trial. Subjects seem to flexibility switch strategies from trial to trial. One interpretation would be that subjects choose an optimal process based on the material and their own capacity. Strategy choice depending on one's own capacity could explain why people with higher simple and complex span performance tend to engage in cognitively demanding strategies like mental imagery or sentence generation (e.g., Bailey et al., 2008, 2011); They choose these strategies because they have enough capacity to implement them.

In conclusion, dual-task costs are unlikely to explain the lack of improvements for the mostly common assessed strategies. The ineffectiveness of strategy instructions seems more related to them impeding adaptive strategy choices.

Do Strategies Explain Why Free Time Benefits Serial Recall?

Presenting memory items at a slower rate – thus adding free time between them – leads to better performance in simple span (Oberauer, 2022b; Souza & Oberauer, 2018; Tan & Ward, 2008) and complex span tasks (Barrouillet et al., 2011; Souza & Oberauer, 2020), and some researchers have explained this effect by assuming that free time is used for memory control processes assisting maintenance (Barrouillet et al., 2011; Tan & Ward, 2008). The present experiments provided three new insights into why WM benefits from longer free time between items:

First, many participants use free time for rehearsal. Consistent with previous strategy-report studies (Bailey, Dunlosky, & Hertzog, 2009; Bailey et al., 2008, 2011; Morrison et al., 2016), rehearsal was among the most commonly reported strategies (24-34% of trials). Yet, rehearsal was not an effective strategy. In trials in which people chose rehearsal their performance tended to be poorer than on other trials, in line with previous reports (Bailey et al., 2011; Dunlosky & Kane, 2007). When instructed, rehearsal yielded no benefit compared to most other strategies, also converging with previous experiments (Souza & Oberauer, 2018, 2020). This finding contradicts theories of WM for verbal materials in which rehearsal is assumed to be helpful for immediate serial recall or even necessary to counteract temporal decay (Baddeley et al., 1975; Camos et al., 2009; Cowan, 1999).

Second, people have some degree of insight into which strategies work relatively well. This could explain why in the present study, with its rather slow presentation rate of items, they reported using free time for elaborative strategies such as mental imagery and sentence generation on around 62% of the trials. This percentage was much higher than reported in previous studies, in which mental imagery was reported for about 6-14% of the trials, and sentence generation 2-13% (see Bailey, Dunlosky, & Hertzog, 2009; Bailey et al., 2008, 2011 for details). One reason for the increased prevalence of elaboration could lie in the longer free-time interval implemented here: Within 4.5 seconds of free time between items, participants have much more time to use time-demanding strategies than in most studies of immediate serial recall.

The frequent choice of elaboration strategies is remarkable as the long free time between items also provides the opportunity to engage in cumulative rehearsal, which is barely possible at the more common presentation rate of one item per second (Tan & Ward, 2008). Nevertheless, participants in our experiments reported choosing rehearsal less often (24-34% of trials) than those in earlier strategy-report studies. To conclude, one reason why free time between items is beneficial to WM performance is that participants use this time to engage more in the relatively effective strategies (i.e., elaboration), and less in ineffective ones (i.e., rehearsal). This is the case specifically in situations under which elaboration *is* indeed effective – meaning when remembering concrete word lists, and when words don’t repeat across trials.

A third insight is that part of the benefit of maintenance strategies for performance on WM tasks might be driven by LTM. In Experiment 1, self-reported mental imagery was found to boost immediate serial recall performance. Yet, this was no longer the case in Experiment 2 in which a closed pool of items was repeatedly used, thereby reducing the usefulness of item memory, and increasing demands on serial order memory. This suggests that mental imagery helps immediate recall through the formation of LTM traces of the items, and people draw on these traces for their responses. The same might be true for other elaborative strategies that are known to improve episodic LTM, such as sentence generation, although the present Experiment 2 was not suited to show that, because sentence generation helps to mitigate proactive interference. This conclusion remains tentative for two reasons. One is that we cannot unambiguously attribute the reduced choice of, and effectiveness of, mental imagery in Experiment 2 to the increased amount of proactive interference between trials. Another reason to be skeptical about this conclusion is that our recent work on proactive interference in serial recall (Oberauer & Bartsch, 2023) reveals that serial recall of word lists is largely immune to proactive interference between trials. That finding speaks against the idea that people draw on episodic memory to assist serial recall.

Conclusion

Across two experiments we have shown that instructing maintenance strategies that are correlated with poor performance in self-reported sessions leads to worse immediate serial recall than when participants freely choose their strategies. This includes the strategy most popular among participants as well as among WM researchers, rehearsal. By contrast, instructing strategies that were correlated with good performance (e.g., Exp 1: mental imagery [*M* = 68.96 % correct] and sentence generation [*M* = 73.41% correct]) did not lead to better performance than participant’s freely chosen strategy-mix (*M* = 74.20 % correct). Without strategy instructions, participants frequently switch between strategies, and apparently do so adaptively. Among the strategies leading to the best performance are variants of elaboration. The benefit of these strategies for performance on a WM might be mediated by their effect on episodic LTM. Our results show that we cannot push participants to do better than what they already manage to do on their own: instructing “effective strategies” can, at best, yield the same level of performance as when they are free to choose their own strategy.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, Issue C, pp. 89–195). Elsevier. https://doi.org/10.1016/S0079-7421(08)60422-3

AuBuchon, A. M., & Wagner, R. L. (2023). Self-generated strategies in the phonological similarity effect. *Memory & Cognition*, 1–19.

Baddeley, A. D. (1986). Working Memory. In *Current Biology* (Vol. 20, Issue 4). Oxford University Press. https://doi.org/10.1016/j.cub.2009.12.014

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/https://doi.org/10.1016/S0022-5371(75)80045-4

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging*, *24*(1), 82–92. https://doi.org/10.1037/a0014078

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition*, *36*(8), 1383–1390. https://doi.org/10.3758/MC.36.8.1383

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrouillet, P., & Camos, V. (2012). As Time Goes By: Temporal Constraints in Working Memory. *Current Directions in Psychological Science*, *21*(6), 413–419. https://doi.org/10.1177/0963721412459513

Barrouillet, P., Gorin, S., & Camos, V. (2020). Simple Spans Underestimate Verbal Working Memory Capacity. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000957

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175–192. https://doi.org/10.1037/a0022324

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage*, *199*, 585–597. https://doi.org/10.1016/j.neuroimage.2019.06.028

Bartsch, L. M., & Oberauer, K. (2021). The effects of elaboration on working memory and long-term memory across age. *Journal of Memory and Language*, *118*, 104215. https://doi.org/10.1016/j.jml.2020.104215

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance , and their contributions to long-term memory formation. *Memory & Cognition*, *46*(5). https://doi.org/10.3758/s13421-018-0805-9

Bartsch, L. M., Souza, A. S., & Oberauer, K. (2024, February 2). The benefits of memory control processes in working memory: comparing effects of self-reported and instructed strategy use. Retrieved from osf.io/m2hx3

Belletier, C., Doherty, J. M., Graham, A. J., Rhodes, S., Cowan, N., Naveh-Benjamin, M., Barrouillet, P., Camos, V., & Logie, R. H. (2023). Strategic adaptation to dual-task in verbal working memory: Potential routes for theory integration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(1), 51.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework.

Trends in Cognitive Sciences, *16*(2), 106–113. https://doi.org/https://doi.org/10.1016/j.tics.2011.12.010
Broadbent, D. E., & Broadbent, M. H. P. (1981). Articulatory suppression and the grouping of successive stimuli. *Psychological Research*, *43*(1), 57–67.
Camos, V., Johnson, M. R., Loaiza, V. M., Portrat, S., Souza, A. S., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences*, 1–14. https://doi.org/10.1111/nyas.13616
Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002
Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*(4), 1029–1034. https://doi.org/10.3758/s13423-014-0772-5
Cowan, N. (1999). An embedded-processes model of working memory. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, *20*, 506.
Craik, F. I. M., & Tulving, E. (1975). Depth of Processing and the Retention of Words in Eisodic Memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294.
Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075
Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 128.
Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403
Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135–178.
Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal Grouping Effects in Immediate Recall: A Working Memory Analysis. *THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY*, *49*(1), 116–139.
Johnson, M. K., Reeder, J. a, Raye, C. L., & Mitchell, K. J. (2002). Second thoughts versus second looks: an age-related deficit in reflectively refreshing just-activated information. *Psychological Science : A Journal of the American Psychological Society / APS*, *13*(1), 64–67. https://doi.org/10.1111/1467-9280.00411
Jonker, T. R., & Macleod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *41*(3), 831–840. https://doi.org/http://dx.doi.org/10.1037/xlm0000069
Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634–646.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572
Liu, Y. S., & Caplan, J. B. (2020). Temporal grouping and direction of serial recall. *Memory & Cognition*, *48*(7), 1295–1315.
Loaiza, V. M., & Lavilla, E. T. (2021). Elaborative strategies contribute to the long-term benefits of time in working memory. *Journal of Memory and Language*, *117*(December 2020), 104205. https://doi.org/10.1016/j.jml.2020.104205
Loaiza, V. M., & Mccabe, D. P. (2012). Temporal–contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition*, *40*(2), 191–203. https://doi.org/10.3758/s13421-011-0148-2
McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence

from delayed recall tests. *Journal of Memory and Language*, *58*(2), 480–494. https://doi.org/10.1016/j.jml.2007.04.004
McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition*, *29*(1), 10–17. https://doi.org/10.3758/BF03195736
Morey, R. D., Rouder, J. N., & Jamil, T. (2015). BayesFactor: Computation of Bayes factors for common designs. *R Package Version 0.9*, *9*.
Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9
Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 369.
Oberauer, K. (2019a). Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends in Cognitive Sciences*, *23*(9), 798–809. https://doi.org/10.1016/j.tics.2019.06.002
Oberauer, K. (2019b). Working Memory Capacity Limits Memory for Bindings. *Journal of Cognition*, *2*(1).
Oberauer, K. (2022a). The importance of random slopes in mixed models for Bayesian hypothesis testing. *Psychological Science*, *33*(4), 648–665.
Oberauer, K. (2022b). When does working memory get better with longer time? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(12), 1754–1774. https://doi.org/10.1037/xlm0001199
R Core Team. (2017). *R: A language and environment for statistical computing. Vienna, Austria; 2014*. https://www.r-project.org/.
Raye, C. L., Johnson, M. R. M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. M. K. (2007). Refreshing: A Minimal Executive Function. *Cortex*, *43*(1), 135–145. https://doi.org/10.1016/S0010-9452(08)70451-9
Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*(4), 597–614.
Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308.
Ryan, J. (1969a). Grouping and short-term memory: different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. https://doi.org/10.1080/14640746908400206
Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *The Quarterly Journal of Experimental Psychology*, *21*(2), 148–155.
Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-017-1357-y
Souza, A. S., & Oberauer, K. (2018). Does articulatory rehearsal help immediate serial recall? *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2018.09.002
Souza, A. S., & Oberauer, K. (2020). No Evidence That Articulatory Rehearsal Improves Complex Span Performance. *Journal of Cognition*, *3*(1), 1–25. https://doi.org/10.5334/joc.103
Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences*, *1339*(1), 20–31. https://doi.org/10.1111/nyas.12603
Souza, A. S., Vergauwe, E., & Oberauer, K. (2018). Where to attend next: guiding refreshing of visual, spatial, and verbal representations in working memory. *Annals of the New York Academy of Sciences*, 20–22. https://doi.org/10.1111/nyas.13621
Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin &*

Review, *15*(3), 535–542.
Thalmann, M., Souza, A. S., & Oberauer, K. (2019). Revisiting the attentional demands of rehearsal in working-memory tasks. *Journal of Memory and Language*, *105*, 1–18.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*(4), 446–468. https://doi.org/10.1016/S0749-596X(03)00095-0
Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 50.
Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1072.
Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory and Language*, *96*, 23–35.
Vo, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin affective word list reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538.
Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*(JUL), 1–22. https://doi.org/10.3389/fpsyg.2013.00433