CrossMark

Item strength affects working memory capacity

Zhangfan Shen[1,2] · Vencislav Popov[2,3] · Anita B. Delahay[2] · Lynne M. Reder[2,3]

Abstract Do the processing and online manipulation of stimuli that are less familiar require more working memory (WM) resources? Is it more difficult to solve demanding problems when the symbols involved are less rather than more familiar? We explored these questions with a dual-task paradigm in which subjects had to solve algebra problems of different complexities while simultaneously holding novel symbol–digit associations in WM. The symbols were previously unknown Chinese characters, whose familiarity was manipulated by differential training frequency with a visual search task for nine hour-long sessions over 3 weeks. Subsequently, subjects solved equations that required one or two transformations. Before each trial, two different integers were assigned to two different Chinese characters of the same training frequency. Half of the time, those characters were present as variables in the equation and had to be substituted for the corresponding digits. After attempting to solve the equation, subjects had to recognize which two characters were shown immediately before that trial and to recall the integer associated with each. Solution accuracy and response times were better when the problems required one transformation only; variable substitution was not required; or the Chinese characters were high frequency. The effects of stimulus familiarity increased as the WM demands of the equation increased. Character–digit associations were also recalled less well with low-frequency characters. These results provide strong support that WM capacity depends not only on the number of chunks of information one is attempting to process but also on their strength or familiarity.

Keywords Working memory · Familiarity · Knowledge · Math · Problem-solving

Working memory (WM) plays a central role in cognition by allowing people to temporarily maintain and process task-relevant information (Baddeley, 1992, 2012). The role of WM is especially evident in high-level cognitive tasks such as mathematical problem-solving, reading comprehension, and analogical reasoning. In tasks like these, people must continuously maintain relevant information, behavioral goals, and intermediate results, while at the same time manipulating this information through complex cognitive skills and procedures. As a result, WM capacity may be exceeded. These abilities, therefore, are limited by how much WM capacity is available. For example, individual differences in WM capacity and experimental manipulations of WM demands correlate strongly with laboratory measures of problem-solving (Beilock, Kulp, Holt, & Carr, 2004; Daily, Lovett & Reder, 2001; Hitch, 1978; Hitch & McAuley, 1991; Lovett, Reder & Lebiere, 1999; Passolunghi & Siegel, 2001; for a review, see Wiley & Jarosz, 2012) and intelligence (Au et al., 2015; Giofrè, Mammarella, & Cornoldi, 2013; Hicks, Harrison, & Engle, 2015; Unsworth, Fukuda, Awh, & Vogel, 2014), as well as with real-life outcomes such as academic and professional performance (Alloway, 2009; Alloway & Alloway, 2010; Bull, Espy, & Wiebe, 2008; Hambrick, Oswald, Darowski, Rench, & Brou, 2010). Given this tight link

---

Zhangfan Shen and Vencislav Popov contributed equally and should be considered co-first authors

✉ Lynne M. Reder
   reder@cmu.edu

[1] School of Mechanical Engineering, Southeast University, Nanjing 211189, China

[2] Department of Psychology, Carnegie Mellon University, 5000 Forbes Avenue, Baker Hall, 345F, Pittsburgh, PA 15213, USA

[3] Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

between WM capacity and performance in complex cognitive tasks, it is crucial to understand the factors that influence WM limitations.

One factor that affects WM capacity is the ability to chunk pieces of information together into a single unit, which allows people to greatly expand the amount of information they can temporarily hold and manipulate. This chunking ability has been demonstrated by numerous studies which have revealed that WM capacity is limited not by the absolute amount of information that needs to be processed but rather by the number of chunks into which this information can be compressed and organized (Miller, 1956; Simon, 1974; for recent reviews, see Cowan, 2001; Gobet et al., 2001). While there is some disagreement in the literature as to the specific number of units that constitute the limit of WM capacity (Cowan, 2001; Gobet & Clarkson, 2004; Luck & Vogel, 1997; Miller, 1956), and as to the nature of the mechanism behind this limitation (Cowan, 2001; Gobet et al., 2001; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016), most theories of WM share one thing in common—they treat chunks as all or nothing. They consider neither that chunks vary in strength nor that such variation in chunk strength would affect WM limitations. For example, the concept of chunk strength does not appear among the factors that limit WM discussed in a recent comprehensive review by Oberauer et al. (2016).

In this article, we test the hypothesis that the capacity of working memory depends not only on the number of chunks people are required to maintain and manipulate but also on the strength of these chunks in long-term memory (Reder, Paynter, Diana, Ngiam, & Dickison, 2007).

Background

Episodic memory theorists have long suggested that the representations of items in long-term memory differ in strength as a function of prior experience. One computational implementation of this idea, the Source of Activation Confusion (SAC) model of memory (Reder et al., 2000), operationalizes item strength as a continuous value stored alongside memory traces. This strength value can be thought of as a chunk's resting level of activation, which in SAC increases with repeated exposures and decays over time. It is important to note that, just as more exposures make a chunk stronger, they also make it appear more familiar. Although chunk strength and item familiarity are both affected by the frequency and recency of previous exposures, they are conceptually distinct concepts. The strength or resting level of activation is a property of representations in memory, and familiarity is a judgment (tacit or explicit) that is influenced by the strength of a concept.

There are good reasons to believe that items that differ in strength also differ in terms of how much WM resources are required to process and manipulate them (for a review, see Reder et al., 2007). As an extension of the SAC model, Reder et al. (2007) proposed that a number of puzzling findings in studies of episodic memory could be easily explained if we posit that the encoding of items and the creation of novel associations between them depletes WM resources as an inverse function of their strength. The addition of a WM resource to the theory was motivated by the desire to explain why long-term memory performance is not only influenced by the strength of the item being tested but also by the strength of other items that were studied alongside it. For example, recognition memory for a single stimulus (a picture or a word) is worse when it had been studied simultaneously with another low-frequency word compared to being studied alongside a high-frequency word (Diana & Reder, 2006). Relatedly, the presence of low-frequency items on a list hurts recall for high-frequency items on the same list (Dewhurst, Hitch, & Barry, 1998) and vice versa (high-frequency words help low-frequency words on the same list) in serial recall (Hulme, Stuart, Brown, & Morin, 2003). Similarly, recognition of low-frequency items improves as the proportion of high-frequency items on the list increases (Malmberg & Murnane, 2002). In addition, while single-item recognition is generally worse for high-frequency words compared to low-frequency words (Reder et al., 2000), the reverse is true for associative recognition of pairs of high-frequency or low-frequency words (Clark, 1992).

Reder et al. (2007) demonstrated with a simple computational model that all of these results likely stem from the fact that low-frequency items require more WM resources to be encoded and to be bound to other items. According to the model, depleting more WM resources during the encoding of low-frequency words leaves fewer resources for processing the paired item (Diana & Reder, 2006), for binding the items to one another (Clark, 1992), or for encoding the remaining items on the list (Dewhurst et al., 1998; Malmberg & Murnane, 2002). In more general terms, the theoretical claims of the model are as follows:

1. Memory operations such as the encoding, updating, and binding of stimulus to context, to other stimuli, or to relational structures, deplete a limited pool of WM resources.
2. The WM resource pool recovers over time.
3. Memory operations deplete more WM resources for less familiar stimuli.
4. As a result of maintaining or manipulating less familiar chunks of information, there are less WM resources available for performing additional operations or for processing additional stimuli.

The first claim is shared by a number of other researchers (e.g., Blumenfeld, Parks, Yonelinas, & Ranganath, 2010; Blumenfeld & Ranganath, 2006, 2007; van Geldorp, Parra, & Kessels, 2015; Peterson & Naveh-Benjamin, 2017;

Wagner, 1999); in contrast, the claims about how stimulus familiarity affects WM resources are unique to our proposal.

While the Reder et al. (2007) model was able to provide a consistent, formal account for all the studies reviewed above, there are two limitations on the conclusions we can draw. These studies had quasi-experimental designs and depended on naturally occurring differences in word frequency. It is known that differences in normative word frequency closely track differences in many semantic and orthographic properties of words, and, as such, it is not clear whether frequency per se is the cause of the effects reviewed above. Another concern is that these studies on their own do not provide direct evidence for the WM component of our proposal, because they did not measure WM capacity but rather long-term episodic memory performance.

Recently, we provided direct experimental support for the hypothesis that weaker chunks would exhaust limited WM resources to a greater degree than stronger chunks (Reder, Liu, Keinath, & Popov, 2016). We measured associative memory as well as working memory capacity for Chinese characters that were previously unfamiliar to subjects. Before we tested memory performance, we differentially familiarized these characters using a visual search task in which characters were randomly assigned to be seen at either low or high frequency. This differential exposure involved hundreds of trials of visual search training over nine hour-long separate sessions in the course of a month. Characters randomly selected to be high frequency were seen 20 times as often as those selected to be low frequency. Thus, we manipulated item familiarity experimentally, rather than depending on preexisting familiarity.

At the end of each week of training, we tested the ability of subjects to associate high-frequency or low-frequency characters. Novel combinations of two high-frequency characters were learned better as a combined cue to recall an English word, and this effect increased over the course of training. Note that the pairs of characters were equally unfamiliar whether the characters were high frequency or low frequency because each week the pairs were novel combinations associated with new English words. Thus, high-frequency characters were easier to bind to one another and to the English word.

We also measured WM capacity for high-frequency and low-frequency characters at the end of training with an *N*-back task. The *N*-back task requires subjects to respond whether the currently presented stimulus is the same as the one presented *N* trials ago, where *N* can be 1, 2, or 3 in different trial blocks. In the *N*-back task subjects must actively maintain the last *n* items in short-term (or working) memory, bind each of them to a corresponding serial position, and rapidly update that binding on each new trial (Owen, McMillan, Laird, & Bullmore, 2005). As such, this task is perfectly suited to explore whether the familiarity of an item affects the amount of WM resources necessary for encoding, binding, and manipulating the item. As predicted, the *N*-back performance was better in blocks

involving the high-frequency characters, and this difference in performance grew with greater working memory demands. It is important to note that the results were not due to failure to chunk the low-frequency characters or due to differences in the ease of encoding the characters. Combined, these results demonstrated that both learning and WM performance are directly influenced by the strength of a chunk.

In summary, both the computational modeling fits to the quasi-experimental findings from recognition and recall memory (Reder et al., 2007), as well as the strong experimental cued-recall and *N*-back results (Reder et al., 2016), provide converging support for the idea that WM limitations depend not only on the number of chunks but also on their strength.

Current experiment

Despite the results discussed so far, the idea that WM capacity differs as a function of the strength of chunks is controversial. No other existing computational model of WM, aside from the SAC model, has a mechanism in place to account for direct effects of item strength on WM capacity. As a result, some researchers believe that the direct frequency effects shown by Reder et al. (2016) are not caused by item strength per se but rather by other effects that exposure frequency might have on long-term representations, such as making them more distinctive (K. Oberauer & E. Ahn, personal communication, July 19, 2017). Providing further evidence with a different working memory task than the task used in Reder et al. (2016) is thus crucial for constraining the possible explanations of the effects.

Based on our framework, we can make the following novel prediction: Familiarity of the stimuli should affect not only recognition/recall and direct measures of WM capacity but also performance in any cognitive task that relies heavily on WM processing. Specifically, if our theory is correct, then given that high-level cognition is crucially dependent on WM capacity, we should be able to observe better performance in problem-solving and reasoning tasks when the elements involved in the task are more rather than less familiar.

To explore this prediction, we experimentally manipulated the familiarity of symbols in an algebraic task and explored how differences in symbol familiarity, task complexity, working memory load, and their interaction affect solution performance. We adapted a mathematical problem-solving task that we have used previously (Anderson, Reder & Lebiere, 1996) in which increasing the task complexity increased the demands on WM. In Anderson et al., we asked subjects to hold a span of two, four, or six digits in memory while solving simple algebraic equations. The equations varied across trials on two dimensions—number of transformations required for solution and whether or not the equation required substitution of constants from the digit span. The equations required either

one (e.g., $3x = 6$) or two steps (e.g., $3x - 2 = 7$) to be solved, and on half of the trials, subjects had to substitute the first two digits from the digit span for variables (a and b) in the equation (e.g., $ax - 2 = b$). After solving the equation, subjects had to recall the current digit span. Anderson et al. (1996) found that performance in both the math task and the digit-span task was worse when the equations were more complicated, when elements of the memory span had to be used to solve the problems, and when the concurrent digit span was longer. Importantly, the impact of increasing concurrent working memory load on the solution performance was not additive with the other factors. Rather, detriments due to digit span length increased as the complexity of the equation increased and when substitution was required. The fact that digit span, number of steps, and substitution all interacted suggested that each variable adversely affected working memory resources, and we modeled those effects as such.

In the current study, we combined our math dual-task paradigm (Anderson et al., 1996) with our visual search training paradigm (Reder et al., 2016). Instead of giving subjects a set of digits to hold in memory while solving an equation, we presented subjects with two Chinese characters that were each paired with a single digit before the equation was presented. On half of the trials, the equations used those two Chinese characters as constants in the equation, and subjects had to substitute the corresponding digit to solve the problem. Whether substitution was involved or not, subjects had to identify the characters and recall the digit associated with those two characters, after they had attempted to solve the equation.

The key difference from Anderson et al. (1996) is in how we manipulated concurrent working memory load—rather than varying the number of digits that had to be remembered, we varied the strength of the two characters on a given trial. As in our previous study (Reder et al., 2016), we operationalized item strength as high or low (20:1 ratio) frequency of exposure during hundreds of trials of visual search training over several weeks. At the beginning of each week of visual search training, starting with Week 2, subjects performed the math task described above and completed a final one after Week 3 for a total of three math sessions.

The critical question was whether differential familiarity of the characters would affect performance in solving algebraic equations in the same manner that the size of the concurrent digit span affected performance in our previous study (Anderson et al., 1996). According to our theory, the processing of low-frequency characters consumes more WM resources compared to high-frequency characters, and thus we should observe impairments on solution performance and subsequent character recall when subjects maintain low-frequency characters in WM. Importantly, we also expected that, as the complexity of the equation increases (from one to two steps, and from the no substitution to the substitution condition), the effect of symbol familiarity on performance

should increase. If the demands on WM are low, such as in the no-substitution condition at Step 1, there should be sufficient WM capacity to process either low-frequency or high-frequency characters. However, as the equation solution demands on WM increase, there will be fewer resources available for processing the characters and maintaining intermediate results, and the impairment should be commensurate to how much of the resources are depleted.

Method

Subjects

Nineteen college students (ages ranging from 18 to 28) from Carnegie Mellon University participated in this study. No subject had prior knowledge of Chinese. In exchange for participation, they received a payment of $135 to $150, depending on performance. Performance was determined by the number of points earned as described below.

Materials, design, and procedure

Visual search task We used the same training procedure described in Reder et al. (2016), with a few modifications that will be noted below. There were 120 unique Chinese characters that were grouped into 30 sets of four. Characters in each set were more visually similar to each other than they were to characters from the remaining sets, and characters within a set had no unique individual features that would distinguish them from the other three characters in the set. The grouping in sets was done by a native Chinese speaker, and we subsequently confirmed the difference in within-set and between-set similarity using an independent orthographic analysis that was developed by Yang, McCandliss, Shu, and Zevin (2009). Each subject was exposed to a randomly drawn sample of 16 sets of characters (64 characters) from the total pool of items. Eight of the 16 sets for each subject were randomly selected to be in the high-frequency condition.

Subjects performed a visual search task for nine sessions[1] on different days across 3 weeks, and each session consisted of 672 trials. Trials began with a fixation presented in the center of the screen, and subjects had to press a button to continue (see Fig. 1, top panel). On each trial, subjects saw a randomly selected target character in the middle of the screen for 1 second, which was followed by an array of three to five characters. Subjects had to respond whether the target character was present or absent in the visual search array. There was

---

[1] Reder et al. (2016) used 12 training sessions, but the effects of frequency on learning were evident even after nine sessions, and by Session 9 there was sufficient evidence that the low-frequency characters were successfully chunked.
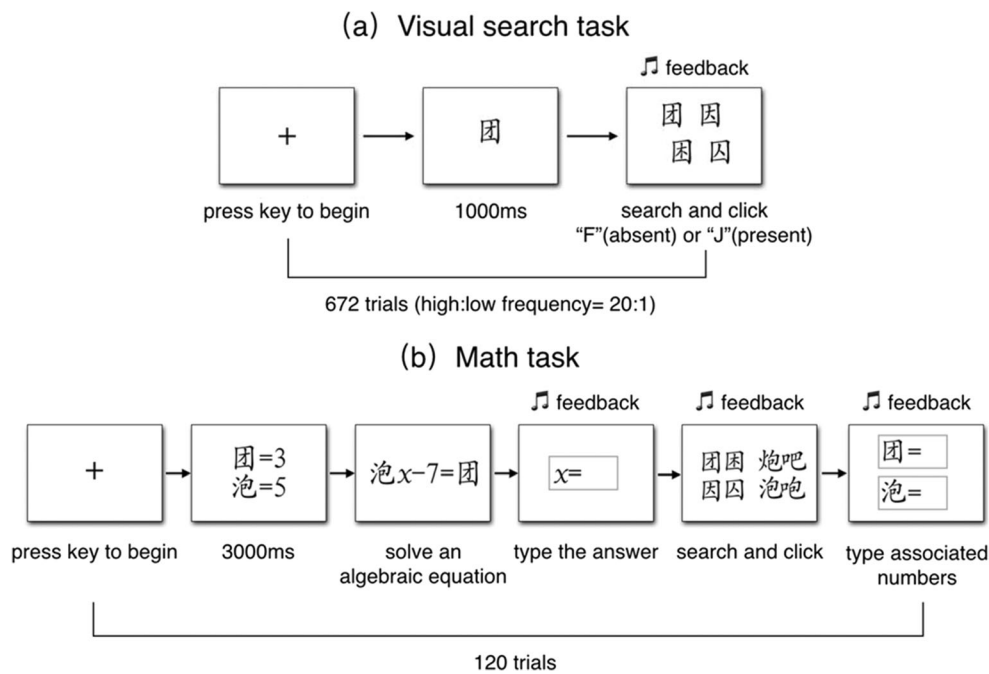
(a) Visual search task

(b) Math task

Fig. 1 Trial sequence for (**a**) the visual search task and (**b**) the math task

no time limit for responses, and after subjects pressed a button, they received auditory feedback that indicated whether their response was correct or not. The size of each character on the screen (1,280 × 800 pixels) was 130 × 130 pixels. The viewing distance was approximately 50 cm.

We used a 2 × 2 within-subjects design, with independent variables of character frequency (high vs. low) and whether the target was present in the search array or not (present vs. absent). High-frequency characters were presented 20 times more often than low-frequency characters in each session. The target character was present in the search array on half of the trials. The search array always contained three distractors from the same set as the target character. An additional one or two distractors from different sets of the same frequency class appeared on some trials. The trial order, the set size, and whether the target was present or not in each trial were randomly determined for each subject and session. The dependent variables were accuracy and response times in reporting whether the target was present or not on each trial.

Math problem-solving task As described in the introduction, the math task was adapted from Anderson et al. (1996). The task involved solving simple linear equations with one unknown (i.e., solve for *x*) using addition, subtraction, multiplication, or division. We randomly generated 360 unique equations with the following constraints: (1) constants were single digits (1–9); (2) intermediate results were single digits (1–9); (3) the final answer was an integer ranging from −9 to 9.

Subjects performed three sessions of the math task after the third, sixth, and ninth visual search training session.

Each math session consisted of 120 different equations, with each condition equally represented in a random order among the trials. Trials began with a fixation presented in the center of the screen, and subjects had to press a button to continue (see Fig. 1, bottom panel). Following the fixation, two characters from the same frequency class were presented on the screen simultaneously for 3 seconds, and each character was associated with a different digit. After viewing the symbol–number associations, a subject was shown an equation to solve and given unlimited time to solve it in their heads (i.e., no pen and paper or calculator). Once the solution was found, the subject pressed a button to continue and then typed in the answer. After entering the answer, the subject had to identify which two characters were shown at the beginning of the trial, selecting each from two arrays of four characters. Each array was composed of characters from the same similarity set as the character that had to be identified. After identifying a character, the digit that had been associated with it also had to be entered. Auditory feedback was provided immediately after subjects entered the equation solution, after character recognition, and after digit recall.

The math task involved a 2 × 2 × 2 within-subjects design, with the following independent factors: the frequency of the Chinese characters presented on each trial (high vs. low), whether the equation contained Chinese characters as constants that had to be substituted for the corresponding digits (substitution vs. no substitution), and whether the equation required one step or two steps. Only the substitution variable and the number of steps variable affected what the subject did during the trial (see Fig. 2).
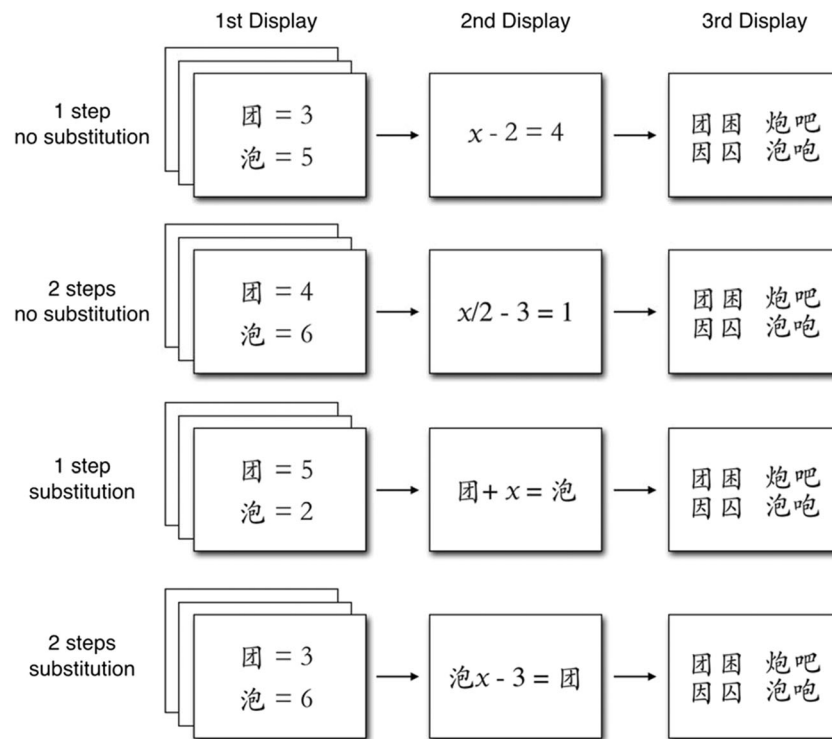
Fig. 2 Example trials for levels of the two variables (2 × 2 factorial) related to equation complexity: substitution versus no substitution and one versus two transformations. The third factor, character frequency (high or low), did not alter the procedure so is not shown here. The characters on each trial varied although they are repeated in this illustration

The Chinese characters used in each equation that required a substitution were the same as the ones used in the visual search task. On all trials, whether or not there was substitution, subjects first studied the assignment of a different digit to each of two characters, and on a given trial, both characters were either from the high-frequency or the low-frequency treatment condition. Half of the equations contained high-frequency characters, and the other half contained low-frequency characters. In each equation, the two characters were randomly selected from two different character sets from the same frequency condition, that is, either both were low-frequency or both were high-frequency. When the equation required a substitution, the position of the two characters during the assignment of numeric value was uncorrelated with their position order in the equation. Three different equation sets were used in the three separate testing sessions, and each set was presented in random order for each subject. We measured accuracy and speed in solving the equation, accuracy in recognizing the trial-specific characters, and accuracy in recalling the associated digit with each character. We instructed subjects to try to be as accurate as possible in solving the equations while trying to also remember the characters–digit bindings as best as they could. Subjects earned 10 points for correctly solving the equation, and 1.5 points for each correctly recalled character–digit binding.

Results

We analyzed the accuracy data via logistic mixed-effects regressions and reaction times via linear mixed-effects regressions (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). For the RT analyses, we considered only correct trials (6.9% error for the visual search task; 10.3% for the math task). Then we excluded from the analyses cases with RTs more than 3 median absolute deviations (Leys, Ley, Klein, Bernard, & Licata, 2013) above or below the median RT, calculated separately for each subject, session, and condition (2.7 % for the visual search task; 5.9% for the math task). RTs were log transformed because the residual plots revealed a lack of homoscedasticity.

Visual search task

We replicated the main results of Reder et al. (2016), namely that both accuracy and response times improved with training and with frequency of exposure (see Fig. 3). Specifically, over the 3 weeks of training, subjects became more accurate, ΔAIC = −942, LLR $\chi^2(1)$ = 944.177, $p < .001$, and faster, ΔAIC = −2236, LLR $\chi^2(1)$ = 2237.098, $p < .001$, in identifying whether the target character was present or absent from the search set. Importantly, subjects identified high frequency characters in the displays more quickly, ΔAIC = −51, LLR $\chi^2(1)$ = 53.632, $p < .001$, and more accurately, ΔAIC = −108, LLR $\chi^2(1)$ = 109.869, $p < .001$. There were no significant
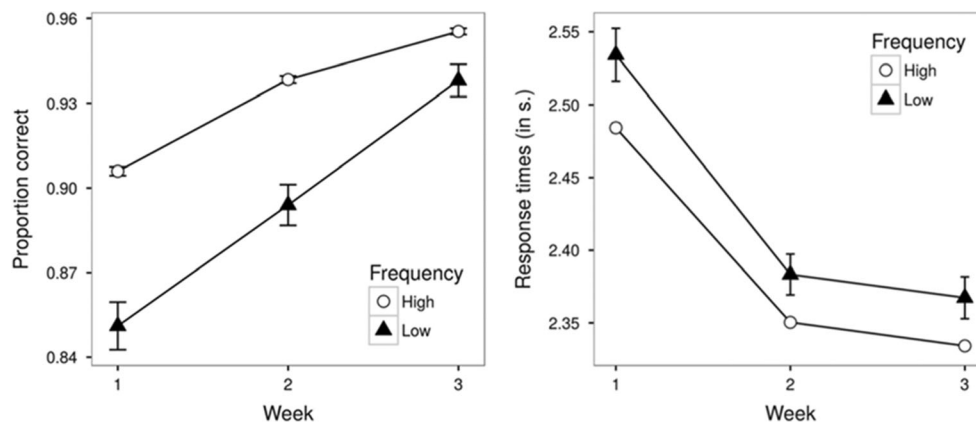
Fig. 3 Mean performance on visual search task trials for high and low frequency over 3 weeks of training. Left panel shows accuracy and right panel shows response times. Error bars indicate +/− 1 standard errors

interactions between training sessions and frequency on accuracy, ΔAIC = 2, LLR $\chi^2(1)$ = 0.113, p = .737, or response times, ΔAIC = 0, LLR $\chi^2(1)$ = 2.263, p =.133.[2] Thus, our experimental manipulation of frequency was successful.

Algebraic problem-solving task

Figure 4 shows the performance on the math task, averaged across weeks.[3] As predicted, performance decreased as the complexity of the equation increased, when it required a substitution, and when the Chinese characters were less familiar. Specifically, subjects performed significantly better in the no-substitution condition, being more accurate, ΔAIC = −120, LLR $\chi^2(1)$ = 122.134, $p <$ .001, and faster in solving the problems, ΔAIC = −1,574, LLR $\chi^2(1)$ = 1,576.28, $p <$ .001. Performance declined as the number of steps increased (from one step to two steps)—subjects became less accurate, ΔAIC =−43, LLR $\chi^2(1)$ = 44.985, $p <$ .001, and solved the equations more slowly, ΔAIC = −1150, LLR $\chi^2(1)$ = 1,152.67, $p <$ .001. Finally, when the Chinese characters they had to remember were more familiar, subjects solved the equations more accurately, ΔAIC = −34, LLR $\chi^2(1)$ = 35.971, $p <$ .001, and more quickly, ΔAIC = −14, LLR $\chi^2(1)$ = 15.68, $p <$ .001.

In addition to the main effects described above, there were a number of significant interactions. Most importantly, the detrimental effect of low familiarity of symbols increased as the equations became more demanding (see Fig. 4). Specifically, the effect of symbol frequency on accuracy was

larger in the two-step compared to the one-step condition, ΔAIC = −3, LLR $\chi^2(1)$ = 5.002, p = .025, and it was larger in the substitution than in the no substitution condition, ΔAIC = −2.9, LLR $\chi^2(1)$ = 4.913, p = .027. Post hoc z tests revealed that the difference in accuracy between high-frequency and low-frequency characters was not significant in the one-step no-substitution condition (z = −0.231, p = .817), but that it was significant in the other three conditions (z = −1.969, p = .025; z = −2.333, p = .01; z = −5.891, $p <$ .001, respectively for the two-step no-substitution, one-step substitution, and the two-step substitution conditions).

For response times, as can be seen from Fig. 4, frequency had a detrimental effect only in the most demanding two-step
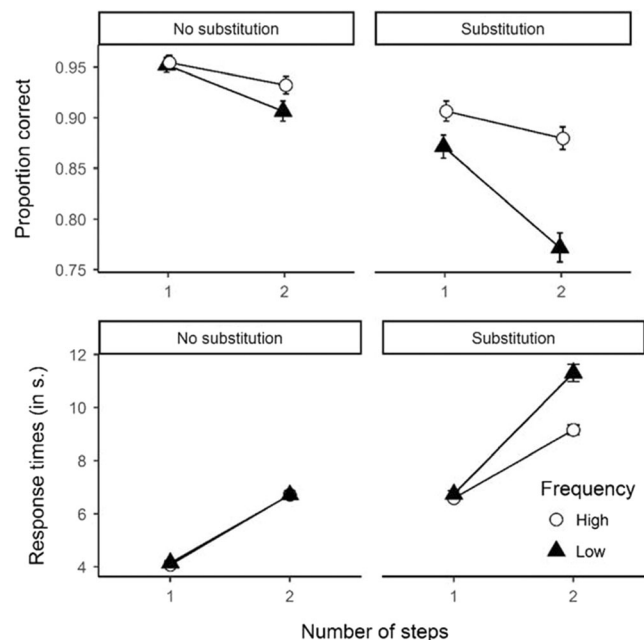
---

[2] In Reder et al. (2016), we did find an interaction between training sessions and frequency, but there we used 12 rather than nine sessions. The longer training time allowed performance to approach ceiling and could be responsible for the previously found interaction. Here, too, to the extent that the accuracy lines are not parallel for accuracy may also be the result of high-frequency characters approaching the ceiling.

[3] None of the effects of number of steps, substitution, or symbol familiarity changed over weeks (all ps > .10), so we report only the overall data.

Fig. 4 Mean performance on math task trials for high and low frequency. Top panel shows accuracy and bottom panel shows response times. Error bars indicate +/− 1 standard errors

substitution condition. This effect was marked by significant two-way interactions between number of steps and frequency, $\Delta AIC = -1$, LLR $\chi^2(1) = 9.41$, $p = .002$, and substitution and frequency, $\Delta AIC = -10$, LLR $\chi^2(1) = 11.57$, $p < .001$, as well as a three-way interaction between all factors, $\Delta AIC = -8$, LLR $\chi^2(1) = 9.94$, $p = .002$. The three-way interaction can also be interpreted as that the slowdown due to the number of steps is strongest for low-frequency characters in the substitution condition (also supported by a two-way interaction of number of steps and substitution), $\Delta AIC = -22$, LLR $\chi^2(1) = 23.78$, $p < .001$. In summary, as we predicted, symbol familiarity interacted with WM demands and its effects on accuracy and response times were greater when the equation required more WM resources.

Character recognition and digit recall

After solving the equations, subjects had to recognize each of the two characters they had seen prior to solving the equation and then recall the number associated with each character. Figure 5 shows the proportion of correct trials, specifically, those for which both characters were identified correctly and both of their associated digits were also correctly recalled. All analyses are reported for the combined recognition and recall data because the pattern was the same when they are considered independently (for interested readers, we report the recognition and recall data separately in Tables 1 and 2 for all trials, and in Tables 3 and 4 for trials on which the equation was solved correctly). As expected, subjects recalled the digit associated with each character more accurately after they solved one-step equations, $\Delta AIC = -8$, LLR $\chi^2(1) = 10.669$, $p = .001$, and when the associated character was more familiar (more previous exposures), $\Delta AIC = -15$, LLR $\chi^2(1) = 17.171$, $p < .001$. In contrast, recall accuracy was significantly higher in the substitution condition, $\Delta AIC = -86$, LLR $\chi^2(1) = 88.798$, $p < .001$. This is likely because subjects had to use the associated digit to solve the equation, which strengthened the association between the character and the digit,
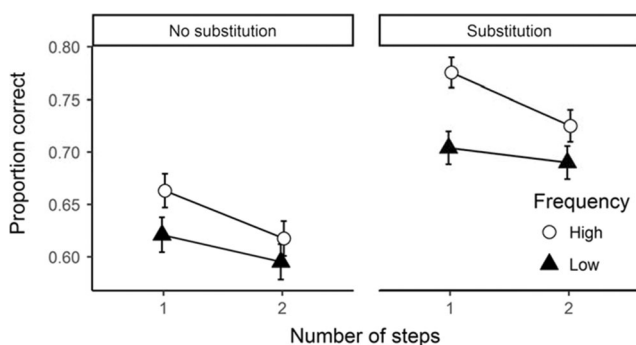
Table 1 Mean accuracy of character identification by condition

Frequency	No substitution		Substitution	
	1 step	2 steps	1 step	2 steps
High	.78	.77	.85	.83
Low	.73	.73	.79	.80

facilitating the recall of the character–digit binding. Alternatively, it could be due to reactivating the character representation, which facilitated the character recognition.

There were no significant interactions between the independent variables (all $ps > .10$).[4] Specifically, no significant interaction between frequency and steps, $\Delta AIC = 0$, LLR $\chi^2(1) = 2.06$, $p = .151$, no significant interaction between frequency and substitution condition, $\Delta AIC = 0$, LLR $\chi^2(1) = 1.7$, $p = .192$, and no significant interaction between steps and substitution condition, $\Delta AIC = 2$, LLR $\chi^2(1) = 0.021$, $p = .886$. Finally, the potential three-way interaction between steps, substitution condition, and frequency was also not reliable, $\Delta AIC = 2$, LLR $\chi^2(1) = 0.35$, $p = .554$.

Discussion

Do the processing and online manipulation of stimuli that are less familiar require more working memory resources? Is it more difficult to solve demanding problems when the symbols involved are less rather than more familiar? The current study suggests that the answer to both of these questions is yes. Here, we showed for the first time that processing more familiar symbols requires less WM resources during complex tasks such as mathematical problem-solving. Specifically, subjects solved algebraic equations faster and more accurately when the symbols they simultaneous held in WM were familiarized to a greater degree before the math task. The beneficial effect of symbol familiarity increased as the equations became more complex, either when the number of transformations required for solution was greater or when subjects had to substitute the symbols in the equation with the associated digits held in WM. In addition, it was easier to maintain a character–digit association in WM when the characters themselves were more familiar (rather than the character–digit association, which was novel on each trial) and when the concurrent problem-

Fig. 5 Mean performance on the Chinese character recognition and recall of corresponding digits for high and low frequency. Error bars indicate +/− 1 standard errors

---

[4] Indeed, in the original Anderson et al. (1996) experiment that varied the number of elements to hold in the digit span, performance on recalling the digits was only better in the substitution condition when there were exactly two digits. The explanation is the same as the one here, but in that study the other levels of digit span rule out alternative explanations. Also see Hitch (1978) for a similar result.

Table 2 Mean accuracy of digit recall to characters as a function of condition

Frequency	No substitution		Substitution	
	1 step	2 steps	1 step	2 steps
High	.66	.62	.78	.73
Low	.62	.60	.70	.69

Table 4 Mean digit recall accuracy to characters for trials that were correctly solved

Frequency	No substitution		Substitution	
	1 step	2 steps	1 step	2 steps
High	.67	.62	.82	.78
Low	.62	.61	.76	.78

solving task was less attentionally demanding. Because both tasks require the concurrent use of WM in order to maintain and manipulate the symbols, these results provide further support for the proposal that WM capacity depends not only on the number of chunks of information one is attempting to process but also on the strength or familiarity of those chunks in memory (Reder et al., 2016, 2007).

The results presented here extend our understanding of how familiarity affects memory and cognition in several ways. It has been well established that item strength (e.g., word frequency, object familiarity) influences the availability of items in long-term memory as measured by recognition memory, free and cued recall performance, lexical decisions, and naming times (e.g., Appelman & Mayzner, 1981; Carroll & White, 1973; Clark, 1992; Grainger, 1990; MacLeod & Kampe, 1996; Ratcliff, Clark, & Shiffrin, 1990; Reder et al., 2016, 2007); however, to our knowledge, it has not been demonstrated previously that item familiarity also affects higher level cognition. This extension of item familiarity effects is theoretically significant because it suggests that the strength of items influences not only their availability or their ease of retrieval from long-term memory (LTM) but also the ease with which they are subsequently manipulated in WM. By showing a benefit of stimulus familiarity in a high-level cognition task that requires WM itself, rather than in a task designed to measure WM capacity directly (Reder et al., 2016), we can be more confident about the construct and ecological validity of our measures, as well as about the theoretical and real-life implications of these findings.

Finally, the few extant studies on the effect of familiarity on WM capacity have mostly involved quasi-experimental designs that relied on preexisting differences in familiarity of the stimuli (Blalock, 2015; Cowan, Ricker, Clark, Hinrichs,

& Glass, 2015; Jackson & Raymond, 2008; Siedenburg & McAdams, 2017; Xie & Zhang, 2017a, b; but see Reder et al., 2016). In these cases, it would be difficult to rule out potentially confounding inequalities in the stimuli. For example, in cases where performance for trained or already familiar stimuli was compared to performance for entirely novel stimuli (Blalock, 2015; Chen, Eng, & Jiang, 2006; Jackson & Raymond, 2008; Siedenburg & McAdams, 2017), it is possible that WM was worse for novel stimuli because they lacked stable unitized representations in the first place. In contrast, we differentially pretrained subjects with previously unknown visually complex symbols (i.e., Chinese characters) in a separate visual search task for nine sessions over 3 weeks, exposing half of the items 20 times more often than the other half. Thus, by the end of the training, none of the items were novel, yet their representations differed in terms of their familiarity. This manipulation allows us to be more confident that the results presented here are due to familiarity of the representation per se rather than to its unitized existence or other confounding factors.

Implications for theories of WM capacity

The finding that the strength of chunks affects WM capacity has important implications for current theories of WM. Current theories fall into three general cases: (a) decay-based, (b) interference-based, and (c) resource-based theories (for reviews, see Baddeley, 2012; Oberauer et al., 2016). We will briefly discuss whether, how, and to what degree these theories might be adjusted to account for the chunk-strength effect presented here and in Reder et al. (2016).

According to decay-based theories, currently active information in WM decays over time, and unless it is reactivated before it falls below a certain threshold, it becomes unavailable for further processing (e.g., Barrouillet, Bernardin, & Camos, 2004; Camos, Lagner, & Barrouillet, 2009). Thus, WM is seen as limited not by the number of items but by how quickly their activation decays and by how often they can be rehearsed/reactivated. If concurrent attentional demands are low, then attentional-based reactivation can occur more frequently, and it can prevent item activation from decaying below threshold (Camos et al., 2009). Some decay-based models might be able to partially account for our results

Table 3 Mean accuracy of character identification for trials that were correctly solved

Frequency	No substitution		Substitution	
	1 step	2 steps	1 step	2 steps
High	.78	.77	.87	.86
Low	.73	.74	.82	.85

if they posit that familiarity affects either the decay rate, the activation threshold, or the maximum level of activation for each stimulus. If less familiar stimuli start with a lower activation level, decay faster, or are more difficult to reactivate, then attentional/executive processes needed for their reactivation will likely be drawn away from the math task more often, hurting solution performance as a result. One of these mechanisms might also explain why low-frequency characters are less likely to be recognized and their corresponding digit recalled—their activation is less likely to exceed the reactivation threshold.

In contrast to decay theories, interference-based theories posit that representations do not decay with time, but that attempting to simultaneously hold many such representations active in WM creates interference between them due to competition, confusion, or feature overlap (e.g., Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). Within this framework, one possibility is that items that are more familiar have representations that are stronger or more distinct and are thus less susceptible to interference. This assumption will likely be able to account for the familiarity effects with the *N*-back task (Reder et al., 2016), for the impaired recall of the character–digit binding with low-frequency characters in the current experiment, and for the performance difference in the equation substitution conditions. Specifically, the representations of low-frequency characters are more likely to interfere with each other because of their weaker and less distinct representations, and this would make it more difficult to substitute their associated digits in the equation and to recall them afterwards. However, it is less clear why greater interference between less familiar characters held in WM would impair equation solution accuracy in the two-step no-substitution condition.[5] In this condition, the symbols that were being held in WM were not used to solve the equation, so the interference on the intermediate results being held in WM should be equivalent in both frequency conditions. Although it is true that subjects had to maintain character–digit bindings even in the no-substitution condition, it is not obvious how low-frequency characters would cause more interference in solving equations that did not involve using them.

Finally, resource-based approaches deserve special attention in this discussion because only a subset of these are compatible with our results. This class of theories attributes WM limitations to the use of a shared pool of limited resources for the active maintenance and manipulation of information. Importantly, a recent debate in the literature concerns whether this WM resource is discrete or continuous in nature (e.g., Donkin, Nosofsky, Gold, & Shiffrin, 2013; Van den Berg,

Awh, & Ma, 2014; Zhang & Luck, 2008). Slot-based theories posit that WM can actively maintain a limited number of distinct representations by allocating them to discrete units/slots (Donkin et al., 2013; Zhang & Luck, 2008). Although we will not review here the relevant evidence for this debate, one thing is worth pointing out: Slot-based theories are fundamentally incompatible with the finding that WM performance is better for stimuli that are more familiar. This is because, in slot-based theories, the only thing that is supposed to limit performance is the number of distinct representations that have to be maintained simultaneously. Yet, in both the current experiment and in Reder et al. (2016), we held the number of chunks constant, while we varied the amount of exposure they had in an independent task.

Continuous resource theories, in contrast, can easily accommodate these results (Reder et al., 2007). Our resource theory posits that variable amounts of WM resources can be allocated for the active maintenance of representations or for executing cognitive processes over them. In line with this view, we suggested in the introduction (1) that the encoding, updating, and binding of stimuli to context, to other stimuli, or to relational structures depends on a limited pool of WM resources; (2) that these operations deplete more WM resources for less familiar stimuli; and (3) that as a result of maintaining or manipulating less familiar chunks of information, there are less WM resources available for performing additional operations or for processing additional stimuli.

In summary, we believe that the chunk-strength findings presented here and in Reder et al. (2016) might prove to be among the key benchmark results that any theory of WM has to be able to explain (for a recent review of such benchmark findings, see Oberauer et al., 2016). Our results are incompatible with slot-based resource theories and mostly compatible with decay and interference theories, depending on their implementation. However, these findings were only predicted by the resource theory as presented here, and in a computational model presented in Reder et al. (2007) and a more complete model that is forthcoming.

Practical implications

Aside from informing theoretical accounts of WM, the chunk-strength effect can be potentially useful for improving instruction and educational practices. Prior research has suggested that an effective way to optimize learning is to design instructional materials and procedures in such a way as to reduce WM/cognitive load during knowledge and skill acquisition, as postulated by Sweller and others (e.g., Clark, Nguyen, & Sweller, 2011; Gerjets, Scheiter, & Catrambone, 2004; Gobet, 2005; Mayer, 2014; Pashler et al., 2007; Sweller, 1994). Several helpful strategies have been proposed to aid in knowledge acquisition, such as grouping single items into larger units and semantically related clusters (Dehn, 2011), or

---

[5] The effect might be small, but it is reliable, and its size is commensurate with the effect in the two-step no-substitution condition in Anderson et al. (1996), who manipulated the memory span (two vs. six digits concurrently in WM), rather than the frequency of the items in WM.

encouraging the formation of stable schemas of relationally organized concepts in LTM (Gerjets et al., 2004; Gobet, 2005; Pashler et al., 2007; Sweller, 1994). While schema induction likely enhances student learning by reducing cognitive load, our findings suggest that in order to achieve optimal learning and problem-solving performance, it is not enough to simply create such chunks and schemas—they have to be sufficiently strengthened before students can move on to acquire additional knowledge. Because highly familiar items require less WM resources for processing, we suggest that students would likely benefit from strengthening individual chunks before being required to use them in solving complicated problems or before having to combine them in more complex structures (Reder et al., 2016).

References

Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, 25(2), 92–98.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, 106(1), 20–29.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, 30(3), 221–256.

Appelman, I. B., & Mayzner, M. S. (1981). The letter-frequency effect and the generality of familiarity effects on perception. *Perception & Psychophysics*, 30(5), 436–446.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22(2), 366–377.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.

Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556–559.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63(1), 1–29.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83.

Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, 133(4), 584–600.

Blalock, L. D. (2015). Stimulus familiarity improves consolidation of visual working memory representations. *Attention, Perception, & Psychophysics*, 77(4), 1143–1158.

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. (2010). Putting the pieces together: The role of dorsolateral prefrontal cortex in relational memory encoding. *Journal of Cognitive Neuroscience*, 23(1), 257–265.

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *Journal of Neuroscience*, 26(3), 916–925.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: An integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, 13(3), 280–291.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33(3), 205–228.

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.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *The Quarterly Journal of Experimental Psychology*, 25(1), 85–95.

Chen, D., Eng, H. Y., & Jiang, Y. (2006). Visual working memory for trained and novel polygons. *Visual Cognition*, 14(1), 37–54.

Clark, R. C., Nguyen, F., & Sweller, J. (2011). *Efficiency in learning: Evidence-based guidelines to manage cognitive load*. New York: John Wiley & Sons.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, 20(3), 231–243.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, 24(1), 87–114; discussion 114–185.

Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, 18(1), 132–145.

Daily, L.Z., Lovett, M.C., & Reder, L.M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*, 315–353.

Dehn, M. J. (2011). *Working memory and academic learning: Assessment and intervention*. New York: John Wiley & Sons.

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 284.

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 805.

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, 120(4), 873–902.

Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science*, 32(1/2), 33–58.

Giofrè, D., Mammarella, I. C., & Cornoldi, C. (2013). The structure of working memory and how it relates to intelligence in children. *Intelligence*, 41(5), 396–406.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, 19(2), 183–204.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four . . . or is it two? *Memory (Hove, England)*, 12(6), 732–747.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5(6), 236–243.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language*, 29(2), 228–244.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167.

Hicks, K. L., Harrison, T. L., & Engle, R. W. (2015). Wonderlic, working memory capacity, and fluid intelligence. *Intelligence*, *50*, 186–195.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*(3), 302–323.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, *82*(3), 375–386.

Hulme, C., Stuart, G., Brown, G. D., & Morin, C. (2003). High-and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518.

Jackson, M. C., & Raymond, J. E. (2008). Familiarity enhances visual working memory for faces. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(3), 556–568.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Lovett, M.C., Reder, L.M., & Lebiere, C. (1999). Modeling Working Memory in a Unified Architecture: An ACT-R Perspective. In Miyake, A. & Shah, P. (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control* (pp. 135–182). Cambridge: Cambridge University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 132.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 616–630.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York: Cambridge University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779–819.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing instruction and study to improve student learning (IES Practice Guide, NCER 2007–2004). Washington, DC: National Center for Education Research.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology*, *80*(1), 44–57.

Peterson, D. J., & Naveh-Benjamin, M. (2017). The role of attention in item-item binding in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. https://doi.org/10.1037/xlm0000386

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review*, *23*(1), 271–277.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a double-edged sword: A computational model of the encoding/retrieval trade-off with familiarity. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 48, pp. 271–312). New York: Elsevier.

Siedenburg, K., & McAdams, S. (2017). The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory*, *25*(4), 550–564.

Simon, H. A. (1974). How big is a chunk? *Science*, *183*(4124), 482–488.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1–26.

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124.

van Geldorp, B, Parra, M. A., & Kessels, R. P. C. (2015). Cognitive and neuropsychological underpinnings of relational and conjunctive working memory binding across age. *Memory*, *23*(8), 1112–1122.

Wagner, A. D. (1999). Working memory contributions to human learning and remembering. *Neuron*, *22*(1), 19–22.

Wiley, J., & Jarosz, A. F. (2012). Working memory capacity, attentional focus, and problem solving. *Current Directions in Psychological Science*, *21*(4), 258–262.

Xie, W., & Zhang, W. (2017a). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, *45*(4), 677–689.

Xie, W., & Zhang, W. (2017b). Familiarity speeds up visual short-term memory consolidation. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication. https://doi.org/10.1037/xhp0000355

Yang, J., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language*, *61*(2), 238–257.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.