Individual Differences in the Reactivity Effect of Judgments of Learning: Cognitive Factors

Zhao, W., Liu, S., Tian, X., Li, B., Shanks, D. R., Yang, C., & Luo, L.

Author Note

The data contained in this project are publicly available at Open Science Framework (https://osf.io/9td3a/).

Conflict of Interest

The authors declare no conflict of interest.

Abstract

An emerging body of studies has demonstrated that asking participants to make concurrent judgments of learning (JOLs) during learning can reactively change (typically enhance) their memory performance, a phenomenon known as the *reactivity effect*. The current study conducted the first exploration of individual differences in the JOL reactivity effect by employing a large-scale (N = 284 participants) approach. The reactivity effect was measured in a related word pair learning task, and each of four higher-order cognitive constructs, including working memory capacity (WMC), attentional control (AC), episodic memory (EM), and general fluid intelligence (gF), was assessed by multiple tasks. The results showed that making JOLs enhanced cued recall of related word pairs, reflecting an overall positive reactivity effect. WMC independently and positively predicted JOL reactivity and this prediction effect survived when controlling for the prediction effects of other cognitive constructs. After controlling for the effects of WMC, EM, and gF, AC negatively predicted JOL reactivity. Neither EM nor gF predicted reactivity. These findings lend support to the learning engagement and dual-task costs theories to jointly account for the JOL reactivity effect. Practical implications for guiding learning practices and for mitigating JOL reactivity in future metacognition research are discussed.

Keywords: Judgments of learning; reactivity; individual differences; working memory capacity; attentional control

Over the last half-century, numerous studies have been conducted to measure people's metacognitive ability by instructing them to concurrently perform a cognitive task (e.g., a learning or decision task) and make item-by-item metacognitive judgments (e.g., prospective estimates of the likelihood of remembering a studied item in a later memory test, or retrospective confidence ratings about decision accuracy) (Bjork et al., 2013; Dunlosky & Tauber, 2016; Koriat, 1997; Li, Hu, et al., 2024; Yang, Yu, et al., 2021). Metacognitive ability is typically measured as the level of consistency between objective task performance and subjective judgments (i.e., relative or absolute accuracy of metacognitive judgments; Fan et al., 2021; Lipowski et al., 2013; Rhodes, 2016; Schraw, 2009).

Although researchers frequently assume that metacognitive judgments provide a passive measure of metacognition and do not affect the underlying cognitive processes *per se*, an emerging body of studies has provided consistent findings challenging this assumption. These studies established that soliciting metacognitive judgments can affect the very things being judged, a phenomenon termed the *reactivity effect* (Mitchum et al., 2016; Shi et al., 2023; Soderstrom et al., 2015; Zhao et al., 2022; Zhao, Li, et al., 2023). For instance, many studies demonstrated that the requirement to report confidence ratings following each decision can reactively enhance individuals' decision accuracy and slow down their decision speed (e.g., Bonder & Gopher, 2019; Double & Birney, 2018; Lei et al., 2020; Li, Hu, et al., 2024). It has also been shown that asking individuals to explain what they are doing or thinking (i.e., concurrent verbalization) can reactively alter their task performance (for a review, see Fox, Ericsson, & Best, 2011).

Besides the reactivity effects of confidence ratings and concurrent verbalization, dozens of recent studies observed that asking participants to make item-by-item judgments of learning (JOLs; metacognitive estimates of the likelihood of remembering a studied item in a later memory test) can reactively change their ultimate memory performance (e.g., Double & Birney, 2019b; Janes et al., 2018; Rivers et al., 2021; Zechmeister & Shaughnessy, 1980; Zhao et al., 2022; Zhao, Yin, et al., 2023). These reactivity findings present substantial challenges for using metacognitive judgments to measure people's metacognitive ability, especially for using JOLs to measure people's metamemory ability (Double & Birney, 2019a; Double et al., 2018; Li et al., 2022).

Below, we first review previous findings concerning the JOL reactivity effect, then highlight the practical and theoretical importance of exploring individual differences in JOL reactivity, and finally provide a brief overview of the aims of the current study.

JOL reactivity

Prior research on the JOL reactivity effect suggested that making JOLs while learning can modify (typically enhance) memory itself. As an illustration, Soderstrom et al. (2015) randomly assigned participants to a JOL or a no-JOL group and instructed both groups to study a mixed list of word pairs, with half being strongly related (e.g., *blunt-sharp*) and the other half being weakly related (e.g., *boxer-terrible*). The JOL group provided JOLs for each word pair during the study phase, reporting their subjective likelihood of later being able to recall the target associated with each cue word, while the no-JOL group did not. Importantly, the total exposure duration for each word pair was matched between the two groups. In a subsequent cued recall test, the JOL group demonstrated significantly better recall performance for strongly related word pairs and a numerical advantage for weakly related word pairs compared to the no-JOL group, indicating a positive reactivity effect on memory for related word pairs.[1]

Similar to Soderstrom et al. (2015), Witherby and Tauber (2017) observed that making JOLs can facilitate both short-term learning and long-term memory for related word pairs. In this research, participants were instructed to study a pure list of related word pairs, with half of the participants providing concurrent JOLs and the other half not. Subsequently, participants completed a cued recall test either after 3 minutes (in the immediate test condition) or 48 hours (in the delayed test condition). The results revealed that the JOL group outperformed the no-JOL group in recalling more word pairs in both the immediate and

[1] The current study especially focused on individual differences in the reactive influences of immediate JOLs (i.e., JOLs made during or immediately after studying each item) on memory. Different from immediate JOLs, delayed JOLs refer to metamemory judgments made after the study phase, and there is always a time interval between the study and JOL-making phases in the delayed JOL paradigm. In this paradigm, participants are typically prompted with a cue to make a prediction about the likelihood of remembering the target in a later test (Nelson & Dunlosky, 1991; Rhodes & Tauber, 2011). Some studies observed that making delayed JOLs can also enhance learning and memory performance. However, the enhancement effect of delayed JOLs is due to the fact that participants tend to covertly retrieve the targets with the aim of assessing the retrieval accessibility of the targets. The covert retrieval process induced by the requirement of making delayed JOLs in turn improves memory performance, a phenomenon known as the covert retrieval effect (Akdoğan et al., 2016; Tauber et al., 2015; Tekin & Roediger, 2021). Clearly, the mechanisms underlying the enhancement effect of delayed JOLs and those underlying the reactivity effect of immediate JOLs are distinct in nature. Hence, we do not further discuss the enhancement effect of delayed JOLs. Interested readers can consult the cited references.

delayed cued recall tests. The effect size for the positive reactivity effect after a 48-hour delay was substantial (Cohen's $d = 0.66$), with minimal difference to the effect size in the immediate test (Cohen's $d = 0.71$), suggesting that the positive reactivity effect persists in the long term.

The positive reactivity effect on memory for related word pairs has been repeatedly obtained in many other studies (Chang & Brainerd, 2023; Double et al., 2018; Janes et al., 2018; Li, Zhao, et al., 2023; Maxwell & Huff, 2022, 2023; Mitchum et al., 2016; Myers et al., 2020; Tauber & Witherby, 2019). Moreover, it generalizes to other types of study materials (e.g., word lists, visual images) and other populations (e.g., elementary school children). In particular, Zhao et al. (2022) employed word lists as stimuli and elementary school students as participants, and found that making JOLs enhanced recognition performance in children in Grades 1, 3, and 5. This positive reactivity effect has also been robustly documented by many other studies which employed young adults as participants (e.g., Halamish, 2018; Halamish & Undorf, 2023; Senkova & Otani, 2021; Yang et al., 2015; Zechmeister & Shaughnessy, 1980). Moreover, this effect survives in a level-of-processing (LOP) paradigm (Tekin & Roediger, 2020). That is, making JOLs can even enhance learning performance when learners employ deep encoding strategies to memorize study items. A meta-analysis conducted by Double et al. (2018) further confirmed the positive reactivity effects on memory for related word pairs and word lists. Furthermore, Shi et al. (2023) recently observed that making JOLs can substantially facilitate memory for visual images, extending the effect to memory for non-verbal materials (i.e., visual memory).

The aforementioned findings provide a basis for considering the extension of JOLs into education. Indeed, some researchers claim that instructing students to concurrently judge their learning status may act as an easy-to-implement intervention to boost learning (e.g., Double et al., 2018; Janes et al., 2018; Li, Zhao, et al., 2023; Soderstrom et al., 2015; Witherby & Tauber, 2017). However, before making such a recommendation, it is critical to explore individual differences in the JOL reactivity effect, a key question that has not been explored thus far.

Practical and theoretical importance

Soliciting JOLs, as claimed by many researchers, has the potential to be an easy-to-implement intervention to enhance learning and memory in daily life. However, before recommending it to practice, individual differences in JOL reactivity must be investigated because it is possible that generating JOLs may not produce an equivalent enhancement effect across all individuals. Indeed, memory in some individuals might even be harmed by making concurrent metamemory judgments (see below for detailed discussion). Hence, practically, it is important to determine for which sub-populations soliciting JOLs reactively enhances and for which it impairs memory. The practical implications of obtaining data on individual differences in JOL reactivity should be fruitful for guiding learning and memory practices.

Additionally, exploring individual differences in JOL reactivity may also provide practical guidance about how to revise research methods to mitigate JOL reactivity in future metacognition research (Double & Birney, 2019b; Janes et al., 2018; Zhao et al., 2022). As will be discussed below, exploring individual differences in JOL reactivity may provide new insights into its cognitive underpinnings. With better understanding of its underlying mechanisms, metacognition researchers may have tools to eliminate JOL reactivity or develop more elegant experimental procedures to mitigate JOL reactivity in future metacognition research (Double et al., 2018; Shi et al., 2023), thus allowing individuals' metacognitive ability to be measured in a less biased way (Li et al., 2022; Mitchum et al., 2016; Rivers et al., 2021).

Considering the aforementioned practical importance, the current study was specially designed to assess individual differences in the reactivity effect of JOLs on memory for related word pairs by probing the relations between JOL reactivity and four higher-order cognitive constructs, including working memory capacity (WMC; i.e., the capacity to hold information in immediate awareness, allowing it to be manipulated and transformed in the service of the current task goals; Unsworth & Spillers, 2010), attentional control (AC; i.e., the ability to focus, sustain, and shift attention as needed; Heitz & Engle, 2007), episodic memory (EM; i.e., the ability to form and recollect conscious memory of specific encountered events; Wheeler et al., 1997), and general fluid intelligence (gF; i.e., the ability to navigate complex and unfamiliar situations by applying logical reasoning and abstract thinking; Liu et al., 2024). Each of these cognitive constructs has been widely explored and found to relate to other learning and memory effects (e.g., the testing effect, the LOP effect, the spacing effect, and study strategy usage; Brewer & Unsworth, 2012; Ferretti & Butterfield, 1992; Minear et al., 2018; Robey, 2019; Toyota, 2015; Unsworth, 2019; Unsworth et al., 2013; Unsworth & Miller, 2024; Unsworth et al., 2020).

Aside from practical implications, exploring individual differences in JOL reactivity may also have important theoretical implications regarding the cognitive underpinnings of the effect. Several theoretical explanations have been proposed, but none of them has thus far received extensive empirical scrutiny. Below we describe two theories that are related to the current study: (1) *learning engagement* and (2) *dual-task costs*.[2]

Zhao et al. (2022) proposed a *learning engagement* theory to account for the positive reactivity effect of JOLs (for related discussion, see Davis & Chan, 2023; Shi et al., 2023). This theory asserts that learners' attention typically wanes and their tendency to suffer lapses of attention (e.g., mind wandering) steadily increases across a prolonged learning task (Brosowsky et al., 2020; Krimsky et al., 2017; Risko et al., 2012; Szpunar et al., 2013; Wammes et al., 2016). When participants are required to form and report item-by-item JOLs, they have to closely study and analyze each study item in order to find "diagnostic" cues to guide JOL formation. Put differently, the requirement of generating JOLs forces participants to sustain attention to the learning task, which in turn produces superior learning outcomes (i.e., a positive reactivity effect). To date, the learning engagement theory has only been empirically tested in a single study. Specifically, Shi et al. (2023) observed that the frequency of mind wandering during a learning task was lower in a JOL than in a no-JOL condition, and the difference in mind wandering rates statistically mediated the positive reactivity effect, suggesting that making JOLs enhances learning through improving engagement (i.e., through reducing mind wandering rates).

Of course, a single study does not permit a firm conclusion to be drawn about the validity of the learning engagement theory, and further tests are needed. The current study aims to test the learning engagement theory through probing the relation between AC and

[2] There are several other explanations of the JOL reactivity effect, such as the cue-strengthening theory (Soderstrom et al., 2015), the changed-goal theory (Mitchum et al., 2016), and the dual-mechanism theory (Janes et al., 2018), which are not related to the current study and hence not discussed here. Readers interested in other explanations can consult the cited references.

JOL reactivity. Numerous studies have established that individuals' AC ability strongly relates to the tendency to experience attention lapses, task-unrelated thoughts and mind wandering (Decker et al., 2023; Unsworth & Robison, 2020; Unsworth et al., 2022). For instance, Unsworth et al. (2021) found a correlation of $r = -.65$ between AC and rates of attention lapses. If making JOLs enhances learning performance through improving engagement (i.e., through sustaining attention), individuals with low AC ability should benefit more than those with high AC ability because they are generally poor at maintaining task attention, leaving more room for improvement. In brief, the learning engagement theory predicts a negative relation between AC and JOL reactivity.

Another potential explanation for JOL reactivity is the *dual-task costs* theory, developed by Mitchum et al. (2016) to explain why providing JOLs sometimes impairs learning. In Mitchum et al. (2016), participants were asked to study a mixed list of related and unrelated word pairs, with one group (i.e., a JOL group) making a JOL after studying each word pair and another group (i.e., a no-JOL group) not making JOLs. In a final cued recall test, although the JOL group recalled more related word pairs than the no-JOL group (reflecting a positive reactivity effect on memory for related pairs), recall performance of unrelated pairs was lower in the JOL than in the no-JOL group, reflecting a negative reactivity effect on memory for unrelated pairs (for related findings, see Janes et al., 2018; Li, Shanks, et al., 2024).

Mitchum et al. (2016) proposed that, during a learning task, the requirement of reporting JOLs may act as a secondary task that borrows limited cognitive resources (e.g., WM) from the primary learning task, in turn leading to an impairment effect on learning (i.e., a negative reactivity effect). This should be especially true when the primary learning task itself is quite challenging and resource-demanding (i.e., when the learning task itself is very difficult to perform, such as learning unrelated word pairs). Mitchum et al. (2016) provided one piece of evidence supporting this theory. In Mitchum et al.'s (2016) Experiments 1-3, two groups (JOL vs. no-JOL) of participants were asked to study a mixed list of related and unrelated word pairs, and they were allowed to spend as much time as they wanted to study each pair. After the study phase, participants completed a cued recall test on studied word pairs. The results showed that even though participants spent more time overall studying word pairs in the JOL than in the no-JOL group, cued recall performance in the final memory test did not statistically differ between the two groups. Put differently, Mitchum et al. observed that longer study time in the JOL group did not produce overall better memory performance, reflecting that the efficiency of study time was lower in the JOL than in the no-JOL group.

Clearly, it is difficult to make a firm conclusion about the validity of the dual-task costs theory based on the single piece of evidence provided by Mitchum et al. (2016). Hence, more tests on the dual-task costs theory are needed. The current study tests this theory by measuring the relation between WMC and JOL reactivity. WMC is crucial for managing and processing information simultaneously, which is essential in dual-task situations. Indeed, many studies found that higher WMC is associated with better performance in dual-task situations due to the greater ability to manage and coordinate tasks simultaneously (Bühner et al., 2006; Logan & Gordon, 2001; Shipstead et al., 2015). Hence, according to the dual-task costs theory and the essential role of WMC in multi-task performing, we expect a positive relation between WMC and JOL reactivity. Specifically, the primary learning task itself may be challenging for individuals with low WMC to perform, with few WM resources left for them to concurrently monitor their ongoing learning status. Hence, making JOLs should be less beneficial or even harmful for individuals with low WMC, because their WMC is more limited and sparing resources to make concurrent JOLs would induce greater costs to their learning performance.

Overall, the learning engagement theory predicts a negative relation between AC and JOL reactivity (that is, providing JOLs is expected to be more beneficial for individuals who are poor at controlling or maintaining attention), and the dual-task costs theory predicts a positive relation between WMC and JOL reactivity (that is, providing JOLs should induce greater dual-task costs and hence be less beneficial or even harmful to individuals with low WMC).

As mentioned above, besides WMC and AC, we also measured participants' EM and gF. A recent study by Zheng et al. (2024) showed that making JOLs reactively enhances not only familiarity-based but also recollection-based recognition of studied words, suggesting that making JOLs enhances memory at least partially through boosting EM (i.e., recollection) of studied information. Hence, it is possible that EM ability is related to the magnitude of the JOL reactivity effect at an individual differences level.

It has been found that individuals with higher gF exhibit greater metacognitive accuracy (for a review, see Ohtani & Hisasaka, 2018), and greater monitoring accuracy allows them to make more efficient adjustments (i.e., changing study strategies) during the study phase, therefore producing an enhancement effect (i.e., positive reactivity) on learning. Indeed, previous research found that gF is associated with the ability to employ effective learning strategies and adapt them as needed (Ferretti & Butterfield, 1992). Furthermore, a recent study by Shi et al. (2023) found that making JOLs enhances learning performance partially through driving learners to use more efficient study strategies (for related findings, see Sahakyan et al., 2004). Hence, it is reasonable to speculate that gF positively relates to JOL reactivity.

Overview of the current study

The current study employed a large-scale (*N* = 284 participants) individual-differences approach to explore cognitive individual differences in JOL reactivity. In total, four cognitive constructs were assessed: WMC, AC, EM, and gF. To enhance measurement reliability, each of these constructs was measured by multiple tasks. Specifically, WMC was assessed by operation span (Ospan), symmetry span (Symspan), and reading span (Rspan) tasks (see below for details). AC was measured by antisaccade, arrow flankers, and psychomotor vigilance tasks.[3] EM was evaluated by delayed recall of word lists, cued recall of word pairs, and picture source-recognition tasks. gF was appraised by Raven's progressive matrices and number series reasoning. The magnitude of JOL reactivity, which served as the dependent measure, was quantified as the difference in cued recall performance between the JOL and no-JOL conditions in a related word pair learning task.

[3] The antisaccade, arrow flankers, and psychomotor vigilance tasks have been widely used in previous studies to measure individuals' AC ability (Brewer & Unsworth, 2012; Draheim et al., 2022; Robison & Brewer, 2022; Robison & Unsworth, 2018; Unsworth & Robison, 2020). Furthermore, it has been established that performance in these tasks strongly relates to individuals' susceptibility to attention lapses, task-unrelated thoughts, and mind wandering in cognitive or non-cognitive tasks. For instance, Unsworth et al. (2022) observed that antisaccade performance closely relates to the occurrence of attention lapses and task-unrelated thoughts, Decker et al. (2023) showed that arrow flankers performance strongly relates to the frequency of attention lapses, and Unsworth and Robison (2020) found a strong association between psychomotor vigilance performance and attention lapses.

Employing multiple measures of each of the four cognitive constructs allowed us to compute composite scores with better psychometric properties (e.g., greater measurement reliability). After calculating composite scores for all constructs, we performed Pearson *r* correlation analyses to quantify the bivariate correlation between JOL reactivity and each of the four cognitive constructs. Next, the composite scores of all four constructs were simultaneously submitted into a multiple regression model to determine the unique variances in JOL reactivity that could be explained by each construct. By investigating individual differences, the current study also aimed to test two theoretical explanations of JOL reactivity: (1) learning engagement and (2) dual-task costs.

Method

Participants

Two hundred and eighty-four undergraduates (M_{age} = 19.40, *SD* = 0.79; 225 female) were recruited from XXX University (Institute information is masked for anonymous review). They participated as a course requirement and also received 100 RMB as compensation. Note that the sample size was not pre-determined according to a power analysis. Instead, the final sample size was determined by the number of students in the course. A *post hoc* sensitivity analysis showed that the sample size was sufficient to detect a weak-to-medium correlation (ρ = .20) with a statistical power of .93 (two-sided α = .05).

The study was approved by the Ethics Committee of XXX Faculty of Psychology (Institute information is masked for anonymous review). All participants provided informed consent and agreed to allow their data to be used for research purposes.

Materials and procedure

Each participant completed 12 tasks in total across four sessions in a computer room. The first three sessions were completed on the same day, with the fourth completed one week later. In Session 1, participants completed the Ospan, arrow flankers, picture source-recognition, and Raven’s progressive matrices tasks. After Session 1, they rested for 10 min. Then in Session 2 they completed the Symspan, antisaccade, delayed recall of word lists, and number series reasoning tasks. After resting for 10 min, they started Session 3, in which they completed the Rspan, cued recall of word pairs, and psychomotor vigilance tasks. These three sessions lasted about 2.5 h. One week later, they completed the related word pair learning task which measured JOL reactivity. The related word pair learning task took about 30 min. All stimuli in all tasks were presented via *PsychoPy* (2023.1.3; Peirce, 2007).

Related word pair learning task

The study stimuli in the related word pair learning task were 80 related word pairs (e.g., *computer-keyboard*), selected from the word pair database developed by Hu et al. (2016). Hu and colleagues asked participants to rate the strength of semantic relatedness between the cue and target word for each pair on a scale ranging from 1 (*not related at all*) to 4 (*strongly related*). The mean rating of relatedness was 3.431 (*SD* = 0.273). To prevent any item selection effects, for each participant, the program randomly divided the 80 pairs into four lists, with two lists randomly assigned to the JOL and the other two to the no-JOL condition. Furthermore, during the study phase, the presentation order of lists and the presentation sequence of word pairs in each list were also randomized.

The related word pair learning task was adapted from Li et al. (2022). Participants were instructed to study four lists of related word pairs in preparation for a later memory test. They were informed that they needed to make memory predictions for two lists of word pairs, but not for the other two lists. They were encouraged to remember all word pairs equally well regardless of whether they needed to make memory predictions or not because all pairs would be eventually tested.

Before studying each list, the computer informed participants whether or not they would make memory predictions for the following list of word pairs. Then, they studied the 20 word pairs in the list. For each of the two no-JOL lists, the 20 pairs were presented one-by-one in random order, with each pair presented for 6 sec, followed by a 0.5 sec fixation to highlight the interstimulus interval (ISI). This cycle repeated until the end of the list. For the two JOL lists, the experimental procedure was identical except that participants made a JOL for each word pair. Specifically, during the last 3 sec of the trial, participants reported their JOL by dragging and clicking a slider to predict their likelihood of recalling the target word when prompted by the cue word in a later cued recall test. JOLs were made on a scale ranging from 0 (*sure I will not remember it*) to 100 (*sure I will remember it*) shown below it.

After studying all four lists, participants completed a distractor task in which they solved as many mathematics problems (e.g., *56+38 = ?*) as they could in 5 min. Next, they completed the final test. The 80 cue words were presented one-by-one in random order. Participants were asked to recall the target word corresponding to each cue word and typed their answer into a blank box. There was no time pressure and no feedback in the final test. The magnitude of JOL reactivity was measured as the difference in recall performance between the JOL and no-JOL conditions, with positive scores representing positive reactivity and negative scores denoting negative reactivity.

In brief, the related word pair learning task involved a within-subjects design (study method: JOL vs. no-JOL). Although previous studies showed that there tends to be a list strength effect (that is, strongly encoded items harm memory for weakly encoded ones) in an intermixed list (e.g., Malmberg & Shiffrin, 2005; Osth et al., 2018; Ratcliff et al., 1990), Wilson and Criss (2017) found no list strength effect in paired associate cued recall. Furthermore, a meta-analysis conducted by Double et al. (2018) showed no moderation effect of experimental design (between- vs. within-subjects design) on JOL reactivity. These findings jointly mitigate potential worry about the list strength effect in the current study.

WMC tasks

Ospan

In the Ospan task, adapted from Unsworth et al. (2005), participants concurrently solved a set of mathematics problems (e.g., *5 + 7 > 10?*) while memorizing a set of English letters (e.g., *E, N, F*). After solving each mathematics problem, a letter was presented for 1 sec for the participant to remember. Then the letter disappeared and the next problem was presented. This cycle repeated until all problems and letters in the set were presented. The length of the letter series (i.e., set size) ranged from 3 to 7. There were three trials at each set size (i.e., 15 trials in total). After presenting each letter set, 12 letters including all studied ones in that set were simultaneously shown on the screen. Participants were asked to sequentially click the studied letters in the studied order. The dependent measure was the total number of letters recalled in the correct serial position.

Symspan

The Symspan task, adapted from Unsworth et al. (2009), required participants to perform a symmetry-judgment task while memorizing a serial sequence of red squares within a matrix. Specifically, participants were shown a series of grid locations one-by-one on a 4 × 4 grid.

Each grid with one cell filled red was presented for 0.65 sec. Participants remembered the locations of red cells and the order in which they appeared. The set size on each trial ranged from 3 to 5 and there were 4 trials at each set size.

After each grid set was shown, participants saw an 8 × 8 grid with a number of cells filled black to form a pattern. Half of the patterns were symmetrical along the vertical axis while the other half were not. Participants were asked to decide as quickly and accurately as they could whether the on-screen pattern was symmetrical or not. Next, they recalled the sequence of locations of the red cells in the preceding displays by sequentially clicking on the cells of an empty 4 × 4 matrix. The dependent measure was the total number of cells recalled in the correct location and serial order.

Rspan

The Rspan task was also adapted from Unsworth et al. (2009). Participants read sentences while remembering a set of unrelated letters. They first read a sentence and determined whether it was meaningful or not (half were meaningful and half not). Nonsense sentences were created by changing one word from an otherwise meaningful sentence. After judging each sentence, participants were shown a letter for 1 sec. Next, a new sentence was shown for them to judge whether it was meaningful. This cycle repeated until all letters in a given set had been shown. The set size on each trial ranged from 3 to 7 and there were 3 trials at each set size (i.e., 15 trials in total). During the test phase, participants were asked to recall the letters from the just-studied set by clicking on the appropriate letters shown on screen. The dependent measure was the total number of letters recalled in the correct serial position.

AC tasks

Antisaccade

The antisaccade task is a classic task to measure individuals' ability to control and regulate their attentional focus (Burgoyne & Engle, 2020; Unsworth et al., 2011). In this task (Hutchison, 2007), participants fixated on a "+" symbol shown at the center of the screen, which was presented for a random duration between 0.2 to 2.2 sec. Then a white cue ("=") appeared at either the left or the right of the fixation for 0.1 sec. Immediately afterward, a target ("O" or "Q") appeared at either the same position as the cue (prosaccade trials) or the symmetrical side (antisaccade trials) for 0.05 sec and was then masked by the letter "H".

Participants indicated whether the target was “O” or “Q” by pressing the corresponding key on the keyboard within a 2 sec deadline. They completed 85 trials in total, including 10 practice trials, 15 prosaccade trials, and 60 antisaccade trials. The dependent measure was the proportion of response errors on the antisaccade trials.

Arrow flankers

The arrow flankers task is a widely used test to measure individuals’ ability to constrain attention to task-relevant stimuli (McDermott et al., 2007; Unsworth & Robison, 2020). On each trial of this task, participants first fixated on a cross for 0.4 sec. Then a three-arrow sequence was presented above the fixation cross for 1.7 sec. Participants indicated the direction of the middle arrow by pressing the left or right arrow key on the keyboard. For congruent trials, the directions of all arrows were identical (e.g., ←←←) while for the incongruent trials, the middle arrow pointed in the opposite direction to the two surrounding arrows (e.g., →←→). There were 50 congruent and 50 incongruent trials in total. The dependent measure was the difference in response times (RTs) between the congruent and incongruent conditions.

Psychomotor vigilance

The psychomotor vigilance task is one of the most widely used tools to assess sustained attention (Sinclair et al., 2013; Thomann et al., 2014). In this task, participants first watched a millisecond clock displaying “0000” at the center of the screen. After a variable amount of time (ranging between 1 to 10 sec), the zeros began to count up in 1 ms intervals from 0 ms. Participants’ task was to press the spacebar as quickly as they could once the numbers began counting upwards. The entire task consisted of 80 trials in total. The dependent measure was the average RTs of the slowest 10% of trials.

EM tasks

Delayed word list recall

The delayed word list recall task was adapted from Brewer and Unsworth (2012). Participants studied four lists of two-character Chinese words, with each list containing 10 words in total. The 40 words were concrete nouns selected from the Chinese word database developed by Cai and Brysbaert (2010). In each list, the words were presented one-by-one in random order, for 1 sec each, with a 0.5 sec fixation presented between each pair of words.

After viewing each list, participants spent 30 sec solving mathematics problems. Next, they completed a free recall test, in which they recalled the studied words in any order they liked and typed their answers into a blank box. The total duration of the free recall test was 1 min. The procedure for the four word lists was identical, except that participants studied new words in each list. The dependent measure was the total number of words successfully recalled in the free recall tests.

Cued recall

In the cued recall task, participants learned and were tested on three lists of unrelated word pairs, with each list containing 10 word pairs (Brewer & Unsworth, 2012). To develop the stimuli, we first selected 60 two-character Chinese nouns from the word database developed by Cai and Brysbaert (2010) and randomly paired them to form 30 unrelated word pairs, which were then randomly divided into three lists. In the study phase of each list, the 10 word pairs were presented one-by-one in random order with each word pair presented for 2 sec for participants to encode. The cue word was shown at the top of the screen and the target word at the bottom. Participants next spent 30 sec solving mathematics problems before completing a cued recall test. Then the 10 cue words were presented one-by-one in random order and participants recalled the target word corresponding to each cue word. For each test trial, they had up to 5 sec to type in their answer into a blank box. The procedure for the three lists of word pairs was identical except that participants studied new word pairs in each list. The dependent measure was the total number of word pairs successfully recalled in the cued recall tests.

Picture source-recognition

The picture source-recognition task was also adapted from Brewer and Unsworth (2012). In this task, participants studied 30 pictures displayed in one of four quadrants on the screen. The 30 pictures were presented one-by-one in random order with each picture presented for 1 sec. The computer randomly decided which of the four quadrants a given picture was located in. Participants were asked to remember both the pictures as well as their locations. In the test phase, participants were presented with 30 old and 30 new pictures and indicated whether a given picture had been studied previously. If a picture was identified as “old”, they then indicated in which quadrant it was originally viewed. They had up to 5 sec to make each response. The dependent measure was the proportion of correct responses in the source recognition test.

gF tasks

Raven's progressive matrices

The Raven's progressive matrices task consisted of 18 puzzles selected from Zhang and Wang (1989). For each puzzle, participants saw a 3 × 3 matrix of geometric patterns with the bottom right pattern missing and selected one of eight alternative patterns according to an unstated rule to fill the missing part of the matrix. The total time to solve these 18 puzzles was 5 min. The dependent measure was the total number of correctly solved puzzles.

Number series reasoning

The number series reasoning task consisted of 15 puzzles selected from Thorndike et al. (1986). For each puzzle, participants saw a series of numbers following an unstated rule (e.g., an arithmetic sequence). They attempted to determine the unstated rule and then selected the most appropriate choice for the next number out of 5 options. The total time to solve these puzzles was 5 min. The dependent measure was the total number of correctly solved puzzles.

Data availability

All data and analysis scripts associated with the current study have been made publicly available via Open Science Framework (OSF; https://osf.io/9td3a/). The study was not pre-reregistered.

Results and discussion

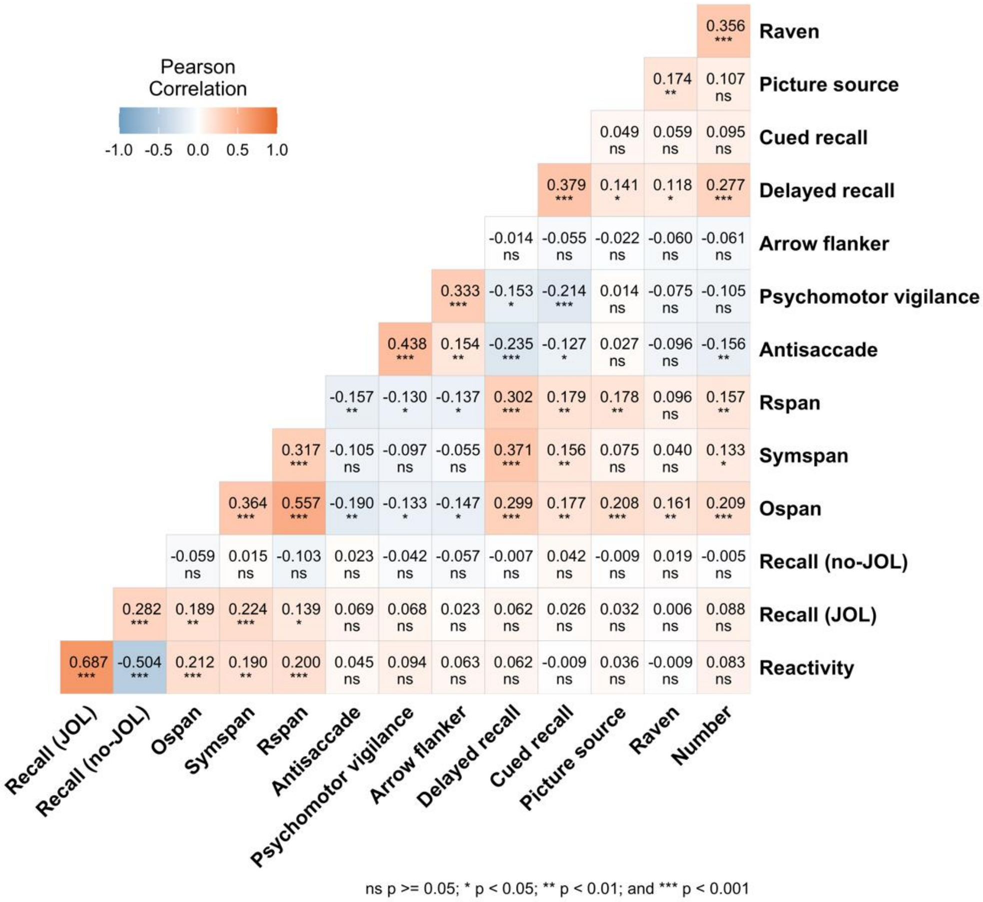
Descriptive statistics for all measures are reported in Table 1, and bivariate correlations among all measures are shown in Figure 1. Following previous individual differences studies (e.g., Robison & Brewer, 2020; Robison & Unsworth, 2018; Unsworth et al., 2021), we performed a confirmatory factor analysis (CFA) to determine whether the 11 cognitive tasks employed here could be represented by four latent factors: WMC, AC, EM, and gF. In the CFA model, we allowed the Ospan, Sysmspan, and Rsapn tasks to load on one factor, the antisaccade, psychomotor vigilance, and arrow flankers tasks to load on a second factor, the delayed recall of word lists, cued recall of word pairs, and picture source-recognition tasks to load on a third factor, and the Raven's progressive matrices and number series tasks to load on a fourth factor.

Table 1. Descriptive statistics for all measures

Measure	*M*	*SD*	Skew	Kurtosis	Reliability
Related word pair learning task					
JOL reactivity	0.045	0.128	0.076	-0.338	0.586
Recall of JOL pairs	0.520	0.116	-0.137	-0.188	0.694
Recall of no-JOL pairs	0.475	0.097	-0.168	-0.184	0.621
WMC tasks					
Ospan	44.082	9.949	-0.119	-0.443	0.706
Symspan	27.611	4.982	-0.166	-0.253	0.702
Rspan	42.039	10.924	0.106	-0.314	0.539
AC tasks					
Antisaccade	0.379	0.170	-0.228	-0.492	0.897
Psychomotor vigilance	711.461	218.472	-0.589	-0.672	0.945
Arrow flankers	151.35	59.546	-0.045	-0.285	0.939
EM tasks					
Delayed recall of word lists	17.859	5.028	0.532	-0.161	0.741
Cued recall of word pairs	14.563	3.949	0.363	0.574	0.705
Picture source-recognition	18.782	3.542	-0.272	1.125	0.661
gF tasks					
Raven's progressive matrices	8.770	1.767	0.154	0.291	0.729
Number series	8.472	1.758	0.215	0.214	0.698

Note: WMC = working memory capacity; AC = attentional control; EM = episodic memory; gF = general fluid intelligence.

Figure 1. Correlation matrix for all measures



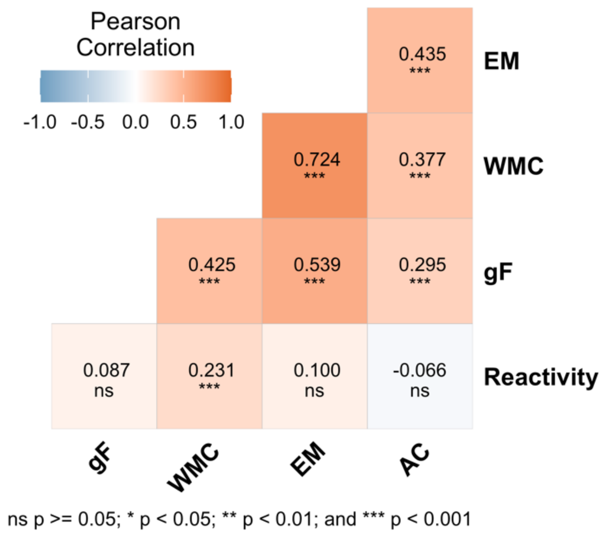
Note: Recall (JOL) = recall of JOL pairs; Recall (no-JOL) = recall of no-JOL pairs.

Several model fit statistics were employed to assess the level of model fit. Non-significant chi-square tests indicate good model fit (Kline, 2016). It is, however, worth noting that significance of chi-square tests is very sensitive to sample sizes. With large sample sizes (such as the in the current study), chi-square tests are nearly always significant. Comparative fit indices (CFI) $\geq$ .950 indicate excellent fit; Tucker-Lewis Index (TLI) $\geq$ .900 indicates good fit; root mean square errors of approximation (RMSEA) $\leq$ .050 indicate excellent fit; standardized root mean square residual (SRMR) values $\leq$.08 indicate good fit (Hu & Bentler, 1999; Kline, 2016). The CFA results showed that the overall fit of the four-constructs model was good to excellent, $\chi^2(38) = 56.271$, $p = .028$; CFI = .957; TLI = .937; RMSEA = .041, 90% CI [.014, .063]; SRMR = .044, supporting the validity of the proposed structure of the 11 cognitive tasks employed here. Standardized factor loadings for the CFA are shown in Appendix (see Table S1). Note that all loadings were statistically significant, *p*s $\leq$.004.

Next, composite scores for the four latent constructs (i.e., WMC, AC, EM, and gF) were calculated. The calculation of composite scores was performed using the factor loadings obtained from the CFA model. Note that, in the antisaccade, psychomotor vigilance, and arrow flankers tasks, higher scores represent inferior AC ability. To make it easy for readers to understand the results, we reversed the composite scores of AC (= 0 – the original composite scores of AC) to compute an AC measure. In a such way, a higher composite score represents superior AC ability.

Consistent with previous research, the composite scores of the four cognitive constructs were interrelated (see Figure 2). Specifically, WMC positively correlated with AC, $r = .377$, $p < .001$, $BF_{10} > 1{,}000$,[4] EM, $r = .724$, $p < .001$, $BF_{10} > 1{,}000$, and gF, $r = .425$, $p < .001$, $BF_{10} > 1{,}000$. AC positively correlated with EM, $r = .435$, $p < .001$, $BF_{10} > 1{,}000$, and gF, $r = .295$, $p < .001$, $BF_{10} > 1{,}000$. EM positively correlated with gF, $r = .539$, $p < .001$, $BF_{10} > 1{,}000$.

Figure 2. Correlation matrix among JOL reactivity, WMC, AC, EM, and gF



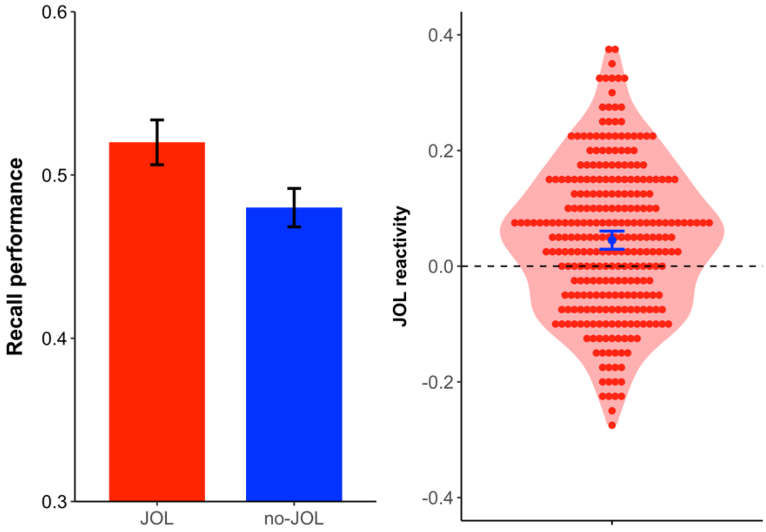
Note: WMC = working memory capacity; AC = attentional control; EM = episodic memory; gF = general fluid intelligence.

[4] The Bayes factor (BF_{10}) represents the relative strength of observed results favoring the alternative (i.e., the existence of an effect) over the null (i.e., the absence of an effect) hypothesis (Hoijtink et al., 2019; Keysers et al., 2020) given the observed data. For instance, $BF_{10} = 3$ indicates that the alternative hypothesis is 3 times as likely to be true as the null hypothesis, and $BF_{10} = 0.33$ indicates that the null hypothesis is about 3 (=1/0.33) times as likely to be true as the alternative hypothesis.

JOL reactivity

In the related word pair learning task, recall in the JOL condition (M = .520, SD = .116) was significantly greater than that in the no-JOL condition (M = .475, SD = .097), difference = .045, 95% CI = [.030, .060], $t(283) = 5.903$, $p < .001$, Cohen's $d = 0.350$, $BF_{10} > 1{,}000$ (see the left panel of Figure 3), successfully replicating the positive reactivity effect. Among the 284 participants, 60.9% showed a positive reactivity effect (i.e., superior recall in the JOL than in the no-JOL condition), 33.1% showed a negative reactivity effect, and the other 6.0% were ties (see the right panel of Figure 3). These findings suggest that even though generating JOLs produced an overall positive reactivity effect, about one-third of participants actually suffered from it, highlighting the necessity to explore individual differences in JOL reactivity. JOL accuracy (i.e., the extent to which JOLs correlated with actual recall) was not one of the main research interests and hence the corresponding results are reported in the Appendix.

Figure 3. Recall performance in the related word pair learning task



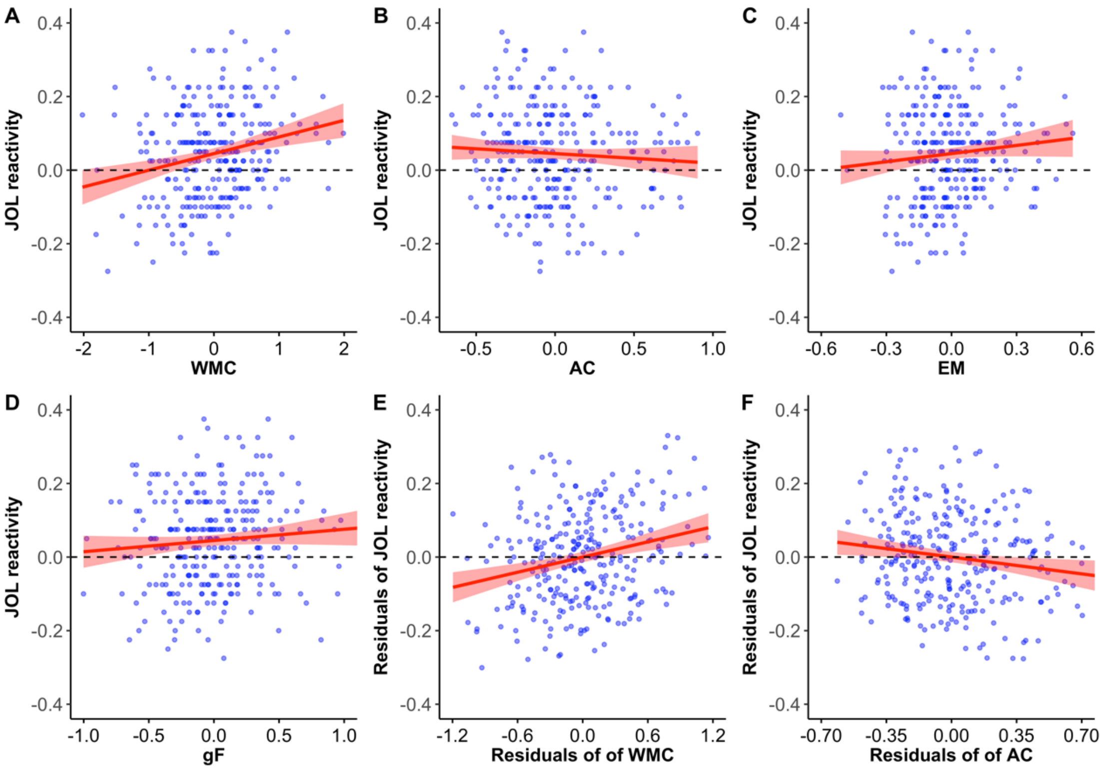
Note: In the violin plot (right panel) each red dot represents a given participant's difference score (i.e., JOL reactivity) based on recall in the JOL and no-JOL conditions. The blue point represents the group average of the difference scores. In both plots, error bars represent 95% CIs.

Individual differences

To examine individual differences in JOL reactivity, we first computed Pearson *r* correlations between JOL reactivity and each of the four cognitive constructs (i.e., composite scores across multiple measures of each cognitive construct). Only WMC positively correlated with JOL reactivity, $r = .231$, $p < .001$, with strong Bayesian evidence supporting the existence of this relationship, $BF_{10} = 277.645$ (see Figure 4A). By contrast, there was no detectable relation between AC and JOL reactivity, $r = -.066$, $p = .271$, $BF_{10} = 0.251$ (see Figure 4B), between EM and JOL reactivity, $r = .100$, $p = .092$, $BF_{10} = 0.559$ (see Figure 4C), or between gF and JOL reactivity, $r = .087$, $p = .142$, $BF_{10} = 0.398$ (see Figure 4D).

The positive correlation between WMC and JOL reactivity is consistent with the dual-task costs theory: Individuals with high WMC benefit more from generating JOLs while those with low WMC benefit less or even suffer from generating JOLs. Indeed, as shown in Figure 4A, for participants with low WMC, making JOLs tended to impair their recall performance. On the face of it, the absence of a relationship between AC and JOL reactivity runs counter to the learning engagement theory, which predicts greater benefits of making JOLs for individuals with low AC ability.

Figure 4. Scatter plots depicting relations between JOL reactivity and each of the four cognitive constructs



Note: Panels A-D depict the bivariate relationships between JOL reactivity and each of the composite scores of working memory capacity (WMC), attentional control (AC), episodic memory (EM), and general fluid intelligence (gF), respectively. Panels E and F depict the relationships between the residuals of JOL reactivity and those of WMC and AC, respectively. In the scatter plots, red lines represent the regression trend between the two variables, with error bars representing 95% CIs.

Because the four cognitive constructs were interrelated (see Figure 2), we next performed a multiple regression analysis in which the composite scores of all four cognitive constructs were simultaneously entered as the independent variables, with JOL reactivity as the dependent variable. This analysis assessed the unique variances in JOL reactivity that could be explained by each cognitive construct.

The regression model was successful overall, $F(4, 279) = 6.517$, $p < .001$, $R^2 = .085$, $BF_{10} = 175.167$. As shown in Table 4, WMC positively predicted the magnitude of JOL reactivity, $b = 0.069$ [0.037, 0.101], $t = 4.215$, $p < .001$, $BF_{10} = 762.153$, confirming the finding of the bivariate analysis. Unlike the null effect in the bivariate analysis, the regression analysis showed that AC negatively predicted the magnitude of JOL reactivity, $b = -0.065$ [-

0.115, -0.015], t = 2.570, p = .011, BF_{10} = 5.083.[5] Neither EM, b = -0.078 [-0.211, 0.054], t = -1.167, p = .244, BF_{10} = 0.442, nor gF, b = 0.015 [-0.032, 0.062], t = 0.642, p = .522, BF_{10} = 0.283, predicted the magnitude of JOL reactivity.

Table 4. Regression coefficients from the multiple regression analysis

Predictor	b	SE (b)	95% CI (b)	β	t	p	BF_{10}
WMC	0.069	0.016	[0.037, 0.101]	.352	4.215	< .001	762.153
AC	-0.065	0.025	[-0.115, -0.015]	-.165	-2.570	.011	5.083
EM	-0.078	0.067	[-0.211, 0.054]	-.107	-1.167	.244	0.442
gF	0.015	0.024	[-0.032, 0.062]	.044	0.642	.522	0.283

Note: WMC = working memory capacity; AC = attentional control; EM = episodic memory; gF = general fluid intelligence.

To visually depict the effects of WMC and AC on JOL reactivity in the multiple regression, we first extracted the residuals of JOL reactivity, WMC, and AC from the multiple regression model and then formed two scatter plots (see Figures 4E and 4F). For instance, we first regressed JOL reactivity on AC, EM, and gF, and obtained the residuals of JOL reactivity that could not be explained by these three constructs. Then, we regressed WMC on AC, EM, and gF, and obtained the residuals of WMC. Finally, we used the residuals of JOL reactivity and those of WMC to generate the scatter plot in Figure 4E. Figure 4F was plotted in a similar manner.

As shown in Figure 4E, after controlling for the effects of AC, EM, and gF, WMC reliably and positively predicted the magnitude of JOL reactivity. Similarly, as shown in Figure 4F, after controlling for the effects of WMC, EM, and gF, AC negatively predicted the

[5] Among the three AC tasks (i.e., antisaccade, psychomotor vigilance, and arrow flankers), the psychomotor vigilance task directly assessed sustained attention. To further test the learning engagement theory of JOL reactivity, we performed another multiple regression analysis, in which AC composite scores were replaced by psychomotor vigilance performance. In this multiple regression analysis, we reversed the average RTs in the psychomotor vigilance task, so that greater measurement scores in this task represent superior ability to sustain attention. The results showed the same pattern. After controlling the confounding effects of other variables, sustained attention (i.e., psychomotor vigilance performance) negatively predicted JOL reactivity, b = -0.017 [-0.033, -0.002], t = -2.270, p = .024, BF_{10} = 2.658, suggesting that making JOLs is more beneficial for those who are poor at sustaining attention. This finding supports the learning engagement theory to account for JOL reactivity.

magnitude of JOL reactivity. This finding is consistent with the learning engagement theory: Generating JOLs is more beneficial for individuals who are poor at controlling their attention.

General Discussion

Although an emerging body of studies has explored the reactive influences of generating JOLs on memory performance (e.g., Double & Birney, 2019a; Double et al., 2018; Li et al., 2022; Mitchum et al., 2016; Shi et al., 2023; Soderstrom et al., 2015; Zhao et al., 2022; Zhao, Li, et al., 2023; Zhao, Yin, et al., 2023), no research has examined this effect from the perspective of individual differences. The current study filled this important gap by assessing cognitive individual differences in the JOL reactivity effect on memory for related word pairs. The observed findings speak to the nature of JOL reactivity in several important respects.

First, the positive reactivity effect on memory for related word pairs was successfully replicated here, with extremely strong Bayesian evidence supporting its existence (BF_{10} > 1,000). Second, the magnitude of JOL reactivity varied substantially across participants. That is, although many participants' memory performance benefited from generating JOLs, there were about one-third (33.1%) who were harmed by it (6.0% showed neither a positive nor negative effect). Thus, despite the fact that providing JOLs facilitated memory overall, it is important to uncover individual differences in this effect and, more importantly, to determine for whom it is beneficial and for whom it is detrimental to memory. Third, there was a positive relation between WMC and JOL reactivity, and after controlling for individual differences in WMC, EM, and gF, a negative relation between AC and JOL reactivity emerged. Neither EM nor gF correlated with JOL reactivity.

It is intriguing that AC itself did not relate to JOL reactivity, as reflected by the absence of a bivariate correlation between these two variables. However, after removing the effects of WMC, EM, and gF, a negative relation between AC and JOL reactivity became evident (BF_{10} = 5.83). The non-significant (weak) bivariate correlation between AC and JOL reactivity and the significant (stronger) relationship between these two variables in the multiple regression analysis mirror the *statistical suppression* (SS) effect (Martinez Gutierrez & Cribbie, 2021; Tzelgov & Henik, 1985; Velicer, 1978). Specifically, the SS effect refers to the phenomenon that after controlling for the effects of other predictors (i.e., suppressors), the relationship between the suppressed predictor and the dependent variable becomes stronger. The SS effect occurs due to the fact that other predictors (i.e., the suppressors) explain irrelevant variance within the suppressed predictor and the dependent variable, therefore strengthening the relationship between the suppressed predictor (i.e., the residuals of the suppressed predictor) and the dependent variable (i.e., the residuals of the dependent variable).

As summarized by Martinez Gutierrez and Cribbie (2021), the SS effect is quite common in psychological research: about one-third of psychological research articles showed evidence of this effect. Consistent with previous studies (e.g., McCord et al., 2014), the SS effect was also observed here. As shown in Figure 2, the four cognitive constructs were interrelated. For instance, AC was positively related to WMC, which means that participants who had greater WMC also had superior AC ability. It is possible that the negative relation between AC and JOL reactivity was masked (or suppressed) by the positive relation between WMC and JOL reactivity, in turn leading to the "null" bivariate relationship between AC and JOL reactivity. Critically, after controlling for the effects of the other three constructs (especially after removing the suppression effect of WMC), the negative relation between AC and JOL reactivity emerged. More specifically, after removing the variance in AC and JOL reactivity that could be explained by the other three cognitive constructs, the remaining residuals of JOL reactivity significantly correlated with the residuals of AC.

The SS effect observed here highlights the complex interplay between AC and the other three cognitive constructs in predicting JOL reactivity. Although AC itself did not directly correlate with JOL reactivity, it explained unique variance in JOL reactivity that could not be explained by the other three cognitive constructs. This finding underscores the importance of considering the SS effect in multiple regression analyses to uncover more accurate relationships between predictors and outcomes.

Theoretical implications

As elaborated in the Introduction, the dual-task costs theory proposes that generating JOLs serves as a secondary task which borrows WM resources from the primary learning task, in turn leading to dual-task costs on the primary learning task and ultimately impairing learning performance (Ariel et al., 2009; Griffin et al., 2008; Hertzog et al., 2002; Janes et al., 2018; Mitchum et al., 2016). Accordingly, this theory predicts that making JOLs should induce greater dual-task costs and be less beneficial for individuals with limited WMC. In line with this prediction, the current study observed strong evidence for a positive relation between WMC and JOL reactivity, regardless of whether the effects of AC, EM, and gF were controlled for or not. The positive relation observed here and the finding of low efficiency of study time documented by Mitchum et al. (2016) jointly support the dual-task costs theory as an account for the JOL reactivity effect.

Even though the dual-task costs theory can readily explain the positive relation between WMC and JOL reactivity, this theory has difficulty in accounting for the overall positive reactivity effect observed here and elsewhere (Chang & Brainerd, 2023; Double et al., 2018; Janes et al., 2018; Li, Zhao, et al., 2023; Maxwell & Huff, 2022, 2023; Mitchum et al., 2016; Myers et al., 2020; Tauber & Witherby, 2019), because it mainly focuses on mechanisms by which generating JOLs can be harmful to memory (Mitchum et al., 2016). In contrast, the learning engagement theory can readily explain the overall positive reactivity effect (Shi et al., 2023; Tauber & Witherby, 2019; Zhao et al., 2022). This theory asserts that generating JOLs produces an enhancement effect by sustaining learners’ attention. Consistent with this proposal, we found that, after controlling the effects of WMC, EM, and gF, AC negatively predicted JOL reactivity, suggesting that JOLs are relatively more beneficial for individuals who are poor at controlling their attention. The negative relation between AC and JOL reactivity and the finding of reduced mind wandering observed by Shi et al. (2023) jointly support the learning engagement theory to account for JOL reactivity.

Overall, the individual differences findings observed here provide new evidence supporting the dual-task costs and learning engagement theories of JOL reactivity. It is worth noting that the mechanisms proposed by these two theories are not necessarily mutually exclusive and it is possible that the mechanisms proposed by them operate concurrently and independently (Zhao et al., 2022).

Besides the aforementioned theoretical implications, the positive and negative relationships may also shed light on the absence of JOL reactivity in older adults (Tauber & Witherby, 2019). Specifically, Tauber and Witherby (2019) found that, although providing JOLs enhanced cued recall of related word pairs in young adults, this pattern failed to generalize to older adults. It is well-established that WMC gradually declines as a function of aging across later adulthood (Craik & Bialystok, 2006; Dobbs & Rule, 1989; Greene et al., 2020; Park et al., 2002; Rhodes & Katz, 2017; Salthouse, 1994). Considering that older adults' WMC is generally limited, the requirement of making concurrent JOLs during the learning phase may induce stronger dual-task costs to their learning performance, which then offsets any benefits of generating JOLs and leads to little or no positive reactivity for older adults. Additionally, the negative relation between AC and JOL reactivity suggests that making JOLs enhances learning through sustaining task engagement. Numerous studies have found that, by comparison with young adults, older adults are generally more motivated to perform well and their minds typically wander less frequently during a prolonged learning task (Frank et al., 2015; Maillet et al., 2018; Maillet & Schacter, 2016). Hence, less room is left for JOLs to maintain older adults' attention, thus weakening positive reactivity.

The findings obtained here also provide new insights into the moderating effect of material type on JOL reactivity. Specifically, previous studies showed that generating JOLs can facilitate memory for simple materials such as word lists (e.g., Li, Zhao, et al., 2023; Li et al., 2022; Zhao et al., 2022; Zhao, Li, et al., 2023), related word pairs (e.g., Soderstrom et al., 2015; Tauber & Witherby, 2019; Witherby & Tauber, 2017), and visual images (Shi et al., 2023), but it does not affect memory for complex materials such as text passages (Ariel et al., 2021) and general knowledge facts (Schäfer & Undorf, 2023). By comparison with simple materials, encoding complex materials generally involves more complex mental processes (e.g., forming a mental situation model to represent the scenarios depicted in a passage) and consumes more WM resources (Alvarez & Cavanagh, 2004; Eng et al., 2005; Hardman & Cowan, 2015). Therefore, the requirement of making concurrent JOLs might induce stronger dual-task costs to the encoding process of complex materials (Hertzog et al., 2002; Mitchum et al., 2016), in turn leading to weakened reactivity.

Overall, the individual differences findings observed here suggest new explanations about the divergent reactivity effects on memory for simple and complex materials observed in prior studies. Additionally, they help explain why providing JOLs does not benefit older adults' memory. We acknowledge that the above discussions are mainly based on theoretical inferences. Direct tests of these inferences are called for.

We highlight that JOL reactivity is a multifaced phenomenon, and both dual-task costs and enhanced learning engagement (and other unknown cognitive or metacognitive mechanisms) may contribute to its occurrence. The current study was especially designed to explore individual differences in the JOL reactivity effect, rather than to specifically test the mechanisms underlying this effect. Furthermore, the current study did not tease dual-task costs and learning engagement apart to evaluate their unique roles in JOL reactivity. Hence, individual differences findings observed here do not permit us to draw firm conclusions about the validity of the dual-task costs and learning engagement theories. Further direct tests on these theoretical accounts are called for.

Practical implications

Besides the aforementioned theoretical implications, the present findings also have important practical implications. For instance, the absence of any relationship between EM and JOL reactivity or between gF and JOL reactivity implies that making JOLs can act as a memory enhancer across individuals with varying EM and gF abilities. The positive relation between WMC and JOL reactivity suggests that soliciting JOLs is most beneficial for individuals with high WMC but harmful for those with low WMC (especially for those whose WMC is lower than 1 *SD* below the mean). The corresponding practical implication is that JOLs can be treated as an easy-to-implement intervention to facilitate learning in individuals with high WMC, but should be avoided for individuals with low WMC. Additionally, other interventions should be considered to enhance learning in individuals with low WMC, such as retrieval practice (i.e., practice testing). It has been shown that testing, by comparison with restudying and other study strategies, helps to consolidate long-term memory and facilitate subsequent learning of new information, a phenomenon known as the testing effect or test-enhanced learning (e.g., Chan et al., 2018; Roediger & Karpicke, 2006; Yang, Luo, et al., 2021; Yang et al., 2018). More importantly, individual differences studies have demonstrated that testing produces either numerically or statistically stronger learning benefits for individuals with low WMC (Agarwal et al., 2017; Brewer & Unsworth, 2012; Nordstrand, 2018; Tse et al., 2019; Tse & Pu, 2012; Yang et al., 2020; but also see Zheng et al., 2023).

The negative relation between AC and JOL reactivity suggests that JOLs can be used to equalize learning across the AC ability range. The corresponding practical implication is that providing JOLs may serve as a method to sustain engagement and improve learning especially for individuals who are poor at maintaining attention. As shown in Figure 4F, when a participant's AC ability was better than the average, making JOLs tended to cause a negative reactivity effect. Hence, for individuals whose AC ability is better than the average, they should consider avoiding making concurrent JOLs, at least from the perspective of JOL reactivity.

Besides the practical implications for guiding learning practices, the observed findings also offer guidance regarding how to develop methods to minimize JOL reactivity in future metacognition research, which is important if researchers wish to measure individuals' metacognitive ability in a less biased way (Janes et al., 2018; Li, Zhao, et al., 2023; Li et al., 2022; Mitchum et al., 2016; Shi et al., 2023). For instance, the negative relationship between AC and JOL reactivity suggests that providing JOLs enhances learning performance through sustaining task attention. The corresponding recommendation for future JOL research is to shorten the overall duration of the learning task because the tendency for attention to wane increases as a task continues. During brief learning sessions, attention lapses occur less frequently (Krimsky et al., 2017; Mason et al., 2007) and hence there would be less opportunity for JOLs to sustain task attention and alter task performance. Besides shortening the task, another potential technique to mitigate reactivity is to divide a prolonged task into several brief sessions and include rest intervals between sessions, which have the potential to reduce attentional disengagement (Chen et al., 2021; Metcalfe & Xu, 2016; Walker & Trick, 2018).

The positive association between WMC and JOL reactivity suggests that generating JOLs tends to reactively impair learning performance through inducing dual-task costs (Mitchum et al., 2016). The corresponding implication for future JOL research is to decrease the level of difficulty of the primary learning task. In a such way, generating JOLs should induce weaker dual-task costs to the primary task. However, although decreasing the difficulty of the primary task may mitigate the negative reactivity effect (especially for individuals with low WMC), it may simultaneously increase the positive reactivity effect (especially for individuals with high WMC). Hence, a more suitable approach is to measure individuals' metacognitive ability in a more covert way (e.g., by recording eye-tracking behaviors or electrical brain activity) rather than asking them to make overt metacognitive judgments. Indeed, some studies have started to use electroencephalogram (EEG) to record alpha (α = 8–14 Hz) oscillations and prefrontal theta-band activity as indices of metacognitive monitoring in error detection tasks (Kononowicz & van Wassenhove, 2019; Wokke et al., 2017). Other studies have attempted to use eye-tracking techniques to measure metacognition in creative thinking tasks (Jiang et al., 2023).

Limitations and future research directions

Three limitations of the current study should be acknowledged. First, we investigated individual differences in the JOL reactivity effect on memory for related word pairs assessed in a cued recall test. Previous assessments suggest that the JOL reactivity effect tends to be moderated by material type and test format (Double et al., 2018; Mitchum et al., 2016; Myers et al., 2020). Hence, it remains unknown whether the conclusions about individual differences obtained here can generalize to the JOL reactivity effects on memory for other types of materials (e.g., word lists, visual images) assessed by different test formats (e.g., old/new recognition). Future research might profitably explore individual differences in JOL reactivity using different types of study materials and test formats.

Second, all the findings observed here are correlational, meaning that it is difficult to make firm causal inferences about the cognitive underpinnings of JOL reactivity. An important goal for future research is to test the dual-task costs and learning engagement theories through directly manipulating WM load (or task difficulty) and attentional engagement (or task motivation).

Finally, the current study measured JOL reactivity as the difference in test performance between the JOL and no-JOL conditions, as done in most (if not all) previous JOL reactivity studies (e.g., Janes et al., 2018; Rivers et al., 2021; Shi et al., 2023; Zhao et al., 2022; Zhao, Yin, et al., 2023). Participants in the no-JOL condition did not need to attend to any additional active control task. It is possible that the difference in test performance between the JOL and no-JOL conditions observed in the current study (as well as those observed in previous JOL reactivity studies) was just due to the fact that LOP in the JOL condition was deeper than that in the no-JOL (passive control) condition. Put differently, the observed difference in test performance between the JOL and no-JOL conditions might simply mirror a LOP effect (Craik & Lockhart, 1972; Nieznański, 2020; Tekin & Roediger, 2020), with nothing special about JOL reactivity. Although there is clear evidence that the effects are independent (see

Tekin & Roediger, 2020), future research should employ an active control condition (e.g., asking participants in the no-JOL control condition to make other forms of judgments unrelated to metamemory, such as semantic judgments) to examine JOL reactivity, which allows researchers to partial out the potential confounding effect of LOP when examining this effect.

Concluding remarks

The requirement to generate and report concurrent JOLs is relatively more beneficial for individuals with high WMC and low AC ability, while the magnitude of JOL reactivity tends not to depend on EM and gF abilities. The dual-task costs and learning engagement theories seem to be viable accounts for the JOL reactivity effect. Future research is encouraged to explore individual differences in the JOL reactivity effects on memory for other types of materials assessed by other test formats, and to further assess the validity of the dual-task costs and learning engagement theories in a more direct way.

References

Agarwal, P. K., Finley, J. R., Rose, N. S., & Roediger, H. L. (2017). Benefits from retrieval practice are greater for students with lower working memory capacity. *Memory*, *25*, 764–771. https://doi.org/10.1080/09658211.2016.1220579

Akdoğan, E., Izaute, M., Danion, J. M., Vidailhet, P., & Bacon, E. (2016). Is retrieval the key? Metamemory judgment and testing as learning strategies. *Memory, 24*, 1390-1395. https://doi.org/10.1080/09658211.2015.1112812

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432–447. https://doi.org/10.1037/a0015928

Ariel, R., Karpicke, J. D., Witherby, A. E., & Tauber, S. (2021). Do judgments of learning directly enhance learning of educational materials? *Educational Psychology Review*, *33*, 693–712. https://doi.org/10.1007/s10648-020-09556-8

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Bonder, T., & Gopher, D. (2019). The effect of confidence rating on a primary visual task. *Frontiers in Psychology, 10*, 2674. https://doi.org/10.3389/fpsyg.2019.02674

Brewer, G. A., & Unsworth, N. (2012). Individual differences in the effects of retrieval from long-term memory. *Journal of Memory and Language*, *66*, 407–415. https://doi.org/10.1016/j.jml.2011.12.009

Brosowsky, N. P., DeGutis, J., Esterman, M., Smilek, D., & Seli, P. (2020). Mind wandering, motivation, and task performance over time: Evidence that motivation insulates people from the negative effects of mind wandering. *Psychology of Consciousness: Theory, Research, and Practice*, Advance online publication. https://doi.org/10.1037/cns0000263

Bühner, M., König, C. J., Pick, M., & Krumm, S. (2006). Working Memory Dimensions as

Differential Predictors of the Speed and Error Aspect of Multitasking Performance. *Human Performance, 19*, 253-275. https://doi.org/10.1207/s15327043hup1903_4

Burgoyne, A. P., & Engle, R. W. (2020). Attention Control: A Cornerstone of Higher-Order Cognition. *Current Directions in Psychological Science, 29*, 624-630. https://doi.org/10.1177/0963721420969371

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos One*, *5*, e10729. https://doi.org/10.1371/journal.pone.0010729

Chan, J. C. K., Meissner, C. A., & Davis, S. D. (2018). Retrieval potentiates new learning: A theoretical and meta-analytic review. *Psychological Bulletin*, *144*, 1111–1146. https://doi.org/10.1037/bul0000166

Chang, M., & Brainerd, C. J. (2023). Changed-goal or cue-strengthening? Examining the reactivity of judgments of learning with the dual-retrieval model. *Metacognition and Learning*, *18*, 183–217. https://doi.org/10.1007/s11409-022-09321-y

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review*, *33*, 1499–1522. https://doi.org/10.1007/s10648-021-09613-w

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*, 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Davis, S. D., & Chan, J. C. K. (2023). Effortful tests and repeated metacognitive judgments enhance future learning. *Educational Psychology Review, 35*, 86. https://doi.org/10.1007/s10648-023-09803-8

Decker, A., Dubois, M., Duncan, K., & Finn, A. S. (2023). Pay attention and you might miss it: Greater learning during attentional lapses. *Psychonomic Bulletin & Review, 30*, 1041-1052. https://doi.org/10.3758/s13423-022-02226-6

Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, *4*, 500–503. https://doi.org/10.1037/0882-7974.4.4.500

Double, K. S., & Birney, D. P. (2018). Reactivity to confidence ratings in older individuals performing the Latin square task. *Metacognition and Learning, 13*, 309–326. https://doi.org/10.1007/s11409-018-9186-5

Double, K. S., & Birney, D. P. (2019a). Do confidence ratings prime confidence? *Psychonomic Bulletin & Review, 26*, 1035–1042. https:// doi. org/ 10. 3758/s13423-018-1553-3

Double, K. S., & Birney, D. P. (2019b). Reactivity to measures of metacognition. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02755

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory*, *26*, 741–750. https://doi.org/10.1080/09658211.2017.1404111

Draheim, C., Pak, R., Draheim, A. A., & Engle, R. W. (2022). The role of attention control in complex real-world tasks. *Psychonomic Bulletin & Review, 29*, 1143-1197. https://doi.org/10.3758/s13423-021-02052-2

Dunlosky, J., & Tauber, S. K. (2016). *The Oxford Handbook of Metamemory*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.001.0001

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review*, *12*, 1127–1133. https://doi.org/10.3758/BF03206454

Fan, T., Zheng, J., Hu, X., Su, N., Yin, Y., Yang, C., & Luo, L. (2021). The contribution of metamemory beliefs to the font size effect on judgments of learning: Is word frequency a moderating factor? *Plos One*, *16*, e0257547. https://doi.org/10.1371/journal.pone.0257547

Ferretti, R. P., & Butterfield, E. C. (1992). Intelligence-related differences in the learning, maintenance, and transfer of problem-solving strategies. *Intelligence, 16*, 207-223. https://doi.org/10.1016/0160-2896(92)90005-C

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*, 316–344. https://doi.org/10.1037/ a0021663

Frank, D. J., Nara, B., Zavagnin, M., Touron, D. R., & Kane, M. J. (2015). Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. *Psychology and Aging*, *30*, 266–278. https://doi.org/10.1037/pag0000031

Greene, N. R., Naveh-Benjamin, M., & Cowan, N. (2020). Adult age differences in working memory capacity: Spared central storage but deficits in ability to maximize peripheral storage. *Psychology and Aging*, *35*, 866–880. https://doi.org/10.1037/pag0000476

Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, *36*, 93–103. https://doi.org/10.3758/MC.36.1.93

Halamish, V. (2018). Can very small font size enhance memory? *Memory & Cognition*, *46*, 979–993. https://doi.org/10.3758/s13421-018-0816-6

Halamish, V., & Undorf, M. (2023). Why do judgments of learning modify memory? Evidence from identical pairs and relatedness judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*, 547–556. https://doi.org/10.1037/xlm0001174

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: Do capacity limits restrict objects or features? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 325–347. https://doi.org/10.1037/xlm0000031

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General, 136*, 217–240. https://doi.org/10.1037/0096-3445.136.2.217

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging*, *17*, 209–225. https://doi.org/10.1037/0882-7974.17.2.209

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods, 24*, 539–556.

https://doi.org/10.1037/met0000201

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*, 1-55. https://doi.org/10.1080/10705519909540118

Hu, X., Liu, Z., Li, T., & Luo, L. (2016). Influence of cue word perceptual information on metamemory accuracy in judgement of learning. *Memory*, *24*, 383–398. https://doi.org/10.1080/09658211.2015.1009470

Hutchison, K. A. (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 645–662. https://doi.org/10.1037/0278-7393.33.4.645

Janes, J. L., Rivers, M. L., & Dunlosky, J. (2018). The influence of making judgments of learning on memory performance: Positive, negative, or both?. *Psychonomic Bulletin & Review*, *25*, 2356–2364. https://doi.org/10.3758/s13423-018-1463-4

Jiang, L., Yang, C., Pi, Z., Li, Y., Liu, S., & Yi, X. (2023). Individuals with high metacognitive ability are better at divergent and convergent thinking. *Journal of Intelligence*, *11*, 162. https://www.mdpi.com/2079-3200/11/8/162

Keysers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience, 23*, 788-799. https://doi.org/10.1038/s41593-020-0660-4

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Kononowicz, T. W., & van Wassenhove, V. (2019). Evaluation of self-generated behavior: Untangling metacognitive readout and error detection. *Journal of Cognitive Neuroscience*, *31*, 1641–1657. https://doi.org/10.1162/jocn_a_01442

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition*, *169*, 84–90.

https://doi.org/10.1016/j.cognition.2017.08.006

Lei, W., Chen, J., Yang, C., Guo, Y., Feng, P., Feng, T., & Li, H. (2020). Metacognition-related regions modulate the reactivity effect of confidence ratings on perceptual decision-making. *Neuropsychologia*, 144, 107502. https://doi.org/10.1016/j.neuropsychologia.2020.107502

Li, B., Hu, X., Shanks, D. R., Su, N., Zhao, W., Meng, L., Lei, W., Luo, L., & Yang, C. (2024). Confidence ratings increase response thresholds in decision making. *Psychonomic Bulletin & Review, 31*, 1093–1102. https://doi.org/10.3758/s13423-023-02380-5

Li, B., Shanks, D. R., Zhao, W., Hu, X., Luo, L., & Yang, C. (2024). Do changed learning goals explain why metamemory judgments reactively affect memory? *Journal of Memory and Language, 136*, 104506. https://doi.org/10.1016/j.jml.2024.104506

Li, B., Zhao, W., Shi, A., Zhong, Y., Hu, X., Liu, M., . . . Yang, C. (2023). Does the reactivity effect of judgments of learning transfer to learning of new information? *Memory*, 1–13. https://doi.org/10.1080/09658211.2023.2208792

Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., . . . Luo, L. (2022). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of learning: Empirical and meta-analytic tests. *Memory & Cognition, 50*, 1061–1077. https://doi.org/10.3758/s13421-021-01258-y

Lipowski, S. L., Merriman, W. E., & Dunlosky, J. (2013). Preschoolers can make highly accurate judgments of learning. *Developmental Psychology*, *49*, 1505–1516. https://doi.org/10.1037/a0030614

Liu, X., Han, X., Wang, T., & Ren, X. (2024). An account of the relationship between critical thinking and fluid intelligence in considering executive functions. *Thinking Skills and Creativity, 52*, 101538. https://doi.org/10.1016/j.tsc.2024.101538

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108*, 393-434. https://doi.org/10.1037/0033-295X.108.2.393

Maillet, D., Beaty, R. E., Jordano, M. L., Touron, D. R., Adnan, A., Silvia, P. J., . . . Kane, M. J. (2018). Age-related differences in mind-wandering in daily life. *Psychology and*

Aging, *33*, 643–653. https://doi.org/10.1037/pag0000260

Maillet, D., & Schacter, D. L. (2016). From mind wandering to involuntary retrieval: Age-related differences in spontaneous cognitive processes. *Neuropsychologia*, *80*, 142–156. https://doi.org/10.1016/j.neuropsychologia.2015.11.017

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 322-336. https://doi.org/10.1037/0278-7393.31.2.322

Martinez Gutierrez, N., & Cribbie, R. (2021). Incidence and interpretation of statistical suppression in psychological research. *Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 53*, 480-488. https://doi.org/10.1037/cbs0000267

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: The default network and stimulus-independent thought. *Science*, *315*, 393–395. https://doi.org/10.1126/science.1131295

Maxwell, N. P., & Huff, M. J. (2022). Reactivity from judgments of learning is not only due to memory forecasting: evidence from associative memory and frequency judgments. *Metacognition and Learning*, *17*, 589–625. https://doi.org/10.1007/s11409-022-09301-2

Maxwell, N. P., & Huff, M. J. (2023). Is discriminability a requirement for reactivity? Comparing the effects of mixed vs. pure list presentations on judgment of learning reactivity. *Memory & Cognition*, *51*, 1198–1213. https://doi.org/10.3758/s13421-022-01381-4

McCord, B., Rodebaugh, T. L., & Levinson, C. A. (2014). Facebook: Social uses and anxiety. *Computers in Human Behavior, 34*, 23-27. https://doi.org/10.1016/j.chb.2014.01.020

McDermott, J. M., Pérez-Edgar, K., & Fox, N. A. (2007). Variations of the flanker paradigm: Assessing selective attention in young children. *Behavior Research Methods, 39*, 62-70. https://doi.org/10.3758/bf03192844

Metcalfe, J., & Xu, J. (2016). People mind wander more during massed than spaced inductive learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 978–984. https://doi.org/10.1037/xlm0000216

Minear, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 1474-1486. https://doi.org/10.1037/xlm0000486

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, *145*, 200–219. https://doi.org/10.1037/a0039923

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition*, *48*, 745–758. https://doi.org/10.3758/s13421-020-01025-5

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science, 2*, 267-271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Nieznański, M. (2020). Levels-of-processing effects on context and target recollection for words and pictures. *Acta Psychologica, 209*, 103127. https://doi.org/10.1016/j.actpsy.2020.103127

Nordstrand, D. (2018). Test-enhanced learning, working memory, and difficulty of material (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-149176

Ohtani, K., & Hisasaka, T. (2018). Beyond intelligence: a meta-analytic review of the relationship among metacognition, intelligence, and academic performance. *Metacognition and Learning, 13*, 179-212. https://doi.org/10.1007/s11409-018-9183-8

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language, 103*, 91-113. https://doi.org/10.1016/j.jml.2018.08.002

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*, 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*, 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

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*, 163-178. https://doi.org/10.1037/0278-7393.16.2.163

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80).

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin, 137*, 131–148. https://doi.org/10.1037/a0021705

Rhodes, R. E., & Katz, B. (2017). Working memory plasticity and aging. *Psychology and Aging*, *32*, 51–59. https://doi.org/10.1037/pag0000135

Risko, E. F., Anderson, N., Sarwal, A., Engelhardt, M., & Kingstone, A. (2012). Everyday attention: Variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, *26*, 234–242. https://doi.org/10.1002/acp.1814

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: Support for the cue-strengthening hypothesis. *Memory*, *29*, 1342–1353. https://doi.org/10.1080/09658211.2021.1985143

Robey, A. (2019). The benefits of testing: Individual differences based on student factors. *Journal of Memory and Language, 108*, 104029. https://doi.org/10.1016/j.jml.2019.104029

Robison, M. K., & Brewer, G. A. (2020). Individual differences in working memory capacity and the regulation of arousal. *Attention, Perception & Psychophysics, 82*, 3273-3290. https://doi.org/10.3758/s13414-020-02077-0

Robison, M. K., & Brewer, G. A. (2022). Individual differences in working memory capacity, attention control, fluid intelligence, and pupillary measures of arousal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*, 1296-1310. https://doi.org/10.1037/xlm0001125

Robison, M. K., & Unsworth, N. (2018). Cognitive and contextual correlates of spontaneous and deliberate mind-wandering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*, 85-98. https://doi.org/10.1037/xlm0000444

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review, 11*, 131-136. https://doi.org/10.3758/BF03206472

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*, 535–543. https://doi.org/10.1037/0894-4105.8.4.535

Schäfer, F., & Undorf, M. (2023). On the educational relevance of immediate judgment of learning reactivity: No effects of predicting one's memory for general knowledge facts. *Journal of Applied Research in Memory and Cognition*, Advance online publication. https://doi.org/10.1037/mac0000113

Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, *4*, 33–45. https://doi.org/10.1007/s11409-008-9031-3

Senkova, O., & Otani, H. (2021). Making judgments of learning enhances memory by inducing item-specific processing. *Memory & Cognition*. *49*, 955–967. https://doi.org/10.3758/s13421-020-01133-2

Shanks, D. R., Don, H. J., Boustani, S., & Yang, C. (2023). Test-Enhanced Learning. In: Oxford University Press.

Shi, A., Xu, C., Zhao, W., Shanks, D. R., Hu, X., Luo, L., & Yang, C. (2023). Judgments of learning reactively facilitate visual memory by enhancing learning engagement. *Psychonomic Bulletin & Review*, *30*, 676–687. https://doi.org/10.3758/s13423-022-02174-1

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics, 77*, 1863-1880. https://doi.org/10.3758/s13414-015-0899-0

Sinclair, K. L., Ponsford, J. L., Rajaratnam, S. M., & Anderson, C. (2013). Sustained attention following traumatic brain injury: use of the Psychomotor Vigilance Task. *Journal of*

Clinical and Experimental Neuropsychology, 35, 210-224. https://doi.org/10.1080/13803395.2012.762340

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 553–558. https://doi.org/10.1037/a0038388

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*, 6313–6317. https://doi.org/doi:10.1073/pnas.1221764110

Tauber, S. K., & Witherby, A. E. (2019). Do judgments of learning modify older adults' actual learning?. *Psychology and Aging*, *34*, 836–847. https://doi.org/10.1037/pag0000376

Tauber, S. K., Dunlosky, J., & Rawson, K. A. (2015). The influence of retrieval practice versus delayed judgments of learning on memory: Resolving a memory-metamemory paradox. *Experimental Psychology, 62*, 254–263. https://doi.org/10.1027/1618-3169/a000296

Tekin, E., & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie, 228*, 278-290. https://doi.org/10.1027/2151-2604/a000425

Tekin, E., & Roediger, H. L. (2021). The effect of delayed judgments of learning on retention. *Metacognition and Learning, 16*, 407-429. https://doi.org/10.1007/s11409-021-09260-0

Thomann, J., Baumann, C. R., Landolt, H. P., & Werth, E. (2014). Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness. *Journal of Clinical Sleep Medicine, 10*, 1019-1024. https://doi.org/10.5664/jcsm.4042

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet intelligence scale: Fourth edition: Guide for administering and scoring the fourth edition*. Chicago: Riverside.

Toyota, H. (2015). The spacing effects in memory and individual differences in emotional intelligence. *奈良教育大学紀要. 人文・社会科学, 64*, 25-32.

Tse, C. S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied, 18*, 253-264. https://doi.org/10.1037/a0029190

Tse, C. S., Chan, M. H. M., Tse, W. S., & Wong, S. W. H. (2019). Can the testing effect for general knowledge facts be influenced by distraction due to divided attention or experimentally induced anxious mood? *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00969

Tzelgov, J., & Henik, A. (1985). A definition of suppression situations for the general linear model: A regression weights approach. *Educational and Psychological Measurement, 45*, 281-284.

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin, 145*, 79-139. https://doi.org/10.1037/bul0000176

Unsworth, N., & Miller, A. L. (2024). An examination of individual differences in levels of processing. *Memory*, 1-12. https://doi.org/10.1080/09658211.2024.2355309

Unsworth, N., & Robison, M. K. (2020). Working memory capacity and sustained attention: A cognitive-energetic perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*, 77-103. https://doi.org/10.1037/xlm0000712

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392-406. https://doi.org/10.1016/j.jml.2010.02.001

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. https://doi.org/10.3758/BF03192720

Unsworth, N., McMillan, B. D., Brewer, G. A., & Spillers, G. J. (2013). Individual differences in everyday retrospective memory failures. *Journal of Applied Research in Memory and Cognition, 2*, 7-13. https://doi.org/10.1016/j.jarmac.2012.11.003

Unsworth, N., Miller, A. L., & Robison, M. K. (2020). Are individual differences in attention control related to working memory capacity? A latent variable mega-analysis. *Journal of Experimental Psychology: General, 150*, 1332–1357. https://doi.org/10.1037/xge0001000

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654. https://doi.org/10.1080/09658210902998047

Unsworth, N., Robison, M. K., & Miller, A. L. (2021). Individual differences in lapses of attention: A latent variable analysis. *Journal of Experimental Psychology: General, 150*, 1303-1331. https://doi.org/10.1037/xge0000998

Unsworth, N., Robison, M. K., & Miller, A. L. (2022). On the relation between working memory capacity and the antisaccade task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*, 1420-1447. https://doi.org/10.1037/xlm0001060

Unsworth, N., Spillers, G. J., Brewer, G. A., & McMillan, B. (2011). Attention control and the antisaccade task: A response time distribution analysis. *Acta Psychologica, 137*, 90-100. https://doi.org/10.1016/j.actpsy.2011.03.004

Velicer, W. F. (1978). Suppressor variables and the semipartial correlation coefficient. *Educational and Psychological Measurement, 38*, 953-958. https://doi.org/10.1177/001316447803800415

Walker, H. E. K., & Trick, L. M. (2018). Mind-wandering while driving: The impact of fatigue, task length, and sustained attention abilities. *Transportation Research Part F: Traffic Psychology and Behaviour*, *59*(Part A), 81–97. https://doi.org/10.1016/j.trf.2018.08.009

Wammes, J. D., Boucher, P. O., Seli, P., Cheyne, J. A., & Smilek, D. (2016). Mind wandering during lectures I: Changes in rates across an entire semester. *Scholarship of Teaching and Learning in Psychology*, *2*, 13–32. https://doi.org/10.1037/stl0000053

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1997). Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. *Psychological Bulletin, 121*, 331–354. https://doi.org/10.1037/0033-2909.121.3.331

Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, *6*, 496–503. https://doi.org/10.1016/j.jarmac.2017.08.004

Wokke, M. E., Cleeremans, A., & Ridderinkhof, K. R. (2017). Sure I'm sure: Prefrontal oscillations support metacognitive monitoring of decision making. *The Journal of Neuroscience*, *37*, 781–789. https://doi.org/10.1523/JNEUROSCI.1612-16.2016

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin*, *147*, 399–435. https://doi.org/10.1037/bul0000309

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ science of learning*, *3*, 1–9.

Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*, *26*, 724–738. https://doi.org/10.1037/xap0000278

Yang, C., Yu, R., Hu, X., Luo, L., Huang, T. S. T., & Shanks, D. R. (2021). How to assess the contributions of processing fluency and beliefs to the formation of judgments of learning: Methods and pitfalls. *Metacognition and Learning*, *16*, 319–343. https://doi.org/10.1007/s11409-020-09254-4

Yang, H., Cai, Y., Liu, Q., Zhao, X., Wang, Q., Chen, C., & Xue, G. (2015). Differential neural correlates underlie judgment of learning and subsequent memory performance. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01699

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, *15*, 41–44. https://doi.org/10.3758/bf03329756

Zhang, H., & Wang, X. (1989). Standardization research on Raven's Standard Progressive Matrices in China. *Acta Psychologica Sinica*, *21*, 113–121. https://www.proquest.com/scholarly-journals/standardization-research-on-ravens-standard/docview/618016179/se-2

Zhao, W., Li, B., Shanks, D. R., Zhao, W., Zheng, J., Hu, X., . . . Yang, C. (2022). When judging what you know changes what you really know: Soliciting metamemory judgments reactively enhances children’s learning. *Child Development*, *93*, 405–417. https://doi.org/10.1111/cdev.13689

Zhao, W., Li, J., Shanks, D. R., Li, B., Hu, X., Yang, C., & Luo, L. (2023). Metamemory judgments have dissociable reactivity effects on item and interitem relational memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*, 557–574. https://doi.org/10.1037/xlm0001160

Zhao, W., Yin, Y., Hu, X., Shanks, D. R., Yang, C., & Luo, L. (2023). Memory for inter-item relations is reactively disrupted by metamemory judgments. *Metacognition and Learning, 18,* 549–566. https://doi.org/10.1007/s11409-023-09340-3

Zheng, J., Li, B., Zhao, W., Su, N., Fan, T., Yin, Y., . . . Luo, L. (2024). Soliciting judgments of learning reactively facilitates both recollection- and familiarity-based recognition memory. *Metacognition and Learning, 19*, 609-633. https://doi.org/10.1007/s11409-024-09382-1

Zheng, Y., Sun, P., & Liu, X. L. (2023). Retrieval practice is costly and is beneficial only when working memory capacity is abundant. *NPJ Science of Learning, 8*, 8. https://doi.org/10.1038/s41539-023-00159-w

Appendix

Participants provided item-by-item JOLs to 99.72% (*SD* = 1.12%) of word pairs in the JOL lists. The average JOL was 49.897 (*SD* = 4.755). A gamma (*G*) correlation was calculated to measure the relative accuracy of JOLs for each participant. Specifically, cued recall performance was dummy coded (correct = 1; incorrect = 0), and then we calculated *G* between JOLs and cued recall performance across word pairs for each participant. Average *G* across participants was .127 (*SD* = .311, 95% CI [.091, .163]), significantly greater than chance (i.e., 0), $t(283) = 6.878$, $p < .001$, Cohen's $d = 0.408$. Thus, participants had some ability to metacognitively discriminate which items they would be able to remember in a future test.

Table S1. Standardized factor loadings from the CFA

Tasks	WMC	AC	EM	gF
Ospan	.762***			
Rspan	.705***			
Symspan	.489***			
Arrow flankers		.39***		
Psychomotor vigilance		.771***		
Antisaccade		.565***		
Picture source-recognition			.209**	
Delayed recall of word lists			.775***	
Cued recall of word pairs			.472***	
Raven's progressive matrices				.444***
Number series				.802***

Note. WMC = working memory capacity; AC = attentional control; EM = episodic memory; gF = general fluid intelligence; ** $p < .01$; *** $p < .001$.