



Metacognition and fluid intelligence in value-directed remembering

Dillon H. Murphy¹ · Karina Agadzhanian¹ · Mary C. Whatley¹ · Alan D. Castel¹

Received: 30 July 2020 / Accepted: 15 March 2021 / Published online: / Published online: 01 April 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The ability to selectively focus on and remember important information, referred to as value-directed remembering, may be crucial for effective memory functioning. In the present study, we investigated the relationships between metacognitive monitoring and control accuracy, selectivity for valuable information, and fluid intelligence. Mediation analyses demonstrated that participants' monitoring assessments and later recall were influenced by the value of the to-be-learned words and the accuracy of participants' judgments was moderated by fluid intelligence. Moreover, recall, selectivity, metacognitive awareness of selectivity, and metacognitive accuracy all generally increased with task experience, demonstrating participants' ability to improve their memory by utilizing cognitive resources more effectively. Together the results suggest that people may be aware of the need to be selective, and engaging in value-directed remembering may be related to higher-level cognitive skills associated with problem-solving and reasoning. Specifically, the strategic use of memory may be involved in focusing on important information, and the metacognitive processes that allow for this prioritization of memory may be related to more general problem-solving abilities that involve identifying important features of information to guide cognition in a broader context.

Keywords Metacognition · Judgments of learning · Fluid intelligence · Value · Selectivity

In a world with massive amounts of information, we are often limited in terms of how much we can process and later remember. To function effectively, people tend to selectively focus on and best recall valuable information, an ability referred to as value-directed remembering (VDR; Castel 2008; Castel et al. 2002; Castel et al. 2012; Elliott et al. 2020). This ability to selectively remember valuable information, often at the expense of low-value information, may be driven by one's awareness and knowledge of their memory abilities, a concept known as metamemory (Nelson and Narens 1990).

✉ Dillon H. Murphy
dmurphy8@ucla.edu

¹ Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA

Metamemory is generally considered in terms of both metacognitive monitoring and control (Dunlosky and Metcalfe 2009). Metacognitive monitoring refers to self-assessments of learning and these judgments often occur in the daily evaluation of memory such as judging whether you will remember someone's name or if you have studied enough for an exam. On the flip side, metacognitive control refers to the self-regulation of learning based on information gained from monitoring such as choosing which information to study in preparation for an exam (Dunlosky et al. 2016; Nelson and Narens 1990; Rivers et al. 2020; Son and Metcalfe 2000; Thiede and Dunlosky 1999). To examine metacognitive monitoring, researchers often ask participants to make predictions of future memory performance, commonly in the form of judgments of learning (JOLs; see Rhodes 2016 for a review).

According to Koriat's (1997) cue-utilization framework, three types of cues inform JOLs: *intrinsic cues* (characteristics of items that influence or are believed to influence memory such as perceptual fluency), *extrinsic cues* (the conditions of encoding or testing such as study time or free recall versus cued recall), and *mnemonic cues* (memorial experiences with items such as how easily an item comes to mind in response to a cue). Judgments are often accurate when judgments and performance are based on the same factors (Dunlosky and Matvey 2001; Tiede and Leboe 2009) and metacognitive accuracy is paramount given that monitoring assessments often inform the metacognitive control processes that play an integral role in the strategic allocation of attention towards important or valuable information (Ariel 2013; Ariel and Dunlosky 2013; Ariel et al. 2009; Dunlosky and Ariel 2011a, 2011b).

Previous research has indicated that the metacognitive monitoring of one's memory for to-be-learned material drives study allocation decisions (Metcalfe and Finn 2008; Rhodes and Castel 2008, 2009) and can also influence later recall of the learned material (i.e., reactivity; Double et al. 2018; Mitchum et al. 2016; Soderstrom et al. 2015; Spellman and Bjork 1992). Furthermore, in the agenda-based regulation (ABR) framework of study time allocation, learners develop and use goal-oriented agendas to prioritize items for study (given their goals and task constraints) based on monitoring assessments to selectively focus on important information that they need to remember (Ariel 2013; Ariel et al. 2009; Dunlosky and Ariel 2011a). For example, when studying for an upcoming exam, students often track their learning by engaging in metacognitive monitoring and make decisions on whether to continue their study (metacognitive control) based on those introspective judgments and an item's value or importance. Thus, accurate metacognitive monitoring is crucial for the optimal allocation of study time and study choices (Mazzoni et al. 1990), and effective allocation of study time can enhance recall for important information (Middlebrooks and Castel 2017).

Although metacognitive monitoring and control are closely interconnected and potentially important for value-directed remembering, fluid intelligence (Gf: the ability to reason, think abstractly, and solve problems without any prior training or expertise; Horn 1982) may be a moderating factor in selectivity for valuable information and the accuracy of metacognition. Given the limited amount of attention one can devote to learning or remembering certain information, an important aspect of intelligence may be determining how one's attentional resources should be allocated. Previous research has elucidated how better and poorer reasoners allocate their time between two levels of strategy planning (i.e., global and local; Sternberg 1981). Specifically, more intelligent participants allocate more time to global planning (i.e., planning at a general level such as an entire list of test items) than to local planning (i.e., planning at a specific level such as a single test item). In other words, a better ability to reason may be related to a more efficient allocation of time, potentially leading to greater task performance.

Some research on the relationship between fluid intelligence and metacognition has indicated that they may be independent of each other (Maqsd [1997](#)) but other research has indicated a positive relationship between fluid intelligence and metacognition (Saraç et al. [2014](#); Sternberg [1985](#); Rozencajwaj [2003](#); Van der Stel and Veenman [2014](#)). Additionally, previous work has indicated that people with higher intelligence are better at regulating their cognitive activities, leading to better task outcomes (Nisbett et al. [2012](#)) and that metacognition predicts academic performance when controlling for intelligence (Ohtani and Hisasaka [2018](#)). Thus, individual differences in intelligence may account for selectivity for important information in value-directed remembering as a result of more efficient metacognitive monitoring and control processes.

While fluid intelligence may be important for metacognitive accuracy, potentially leading to enhanced selectivity for important information, memory self-efficacy (MSE), a self-assessment of one's memory abilities, often positively affects memory performance (Bandura [1989, 1997](#); Berry [1999](#); see Beaudoin and Desrichard [2011](#) for a review). According to self-efficacy theory, four main factors influence MSE: mastery (based on past successes and failures), modeling (the observation and adoption of the behaviors of other people), verbal persuasion or dissuasion (the explicit feedback on memory abilities from others), and level of physiological or psychological excitation or inhibition. As a result of these factors, researchers have suggested that heightened MSE can relate to increased effort and motivation (Bandura [1977](#)), goal setting (Locke et al. [1984](#)), resilience and persistence (Cervone and Peake [1986](#)), lower anxiety (Hertzog et al. [1989](#)), as well as better task strategy (Locke et al. [1984](#)).

Whereas metamemory refers more specifically to the monitoring of learning and control processes, memory self-efficacy captures the general belief in one's memory abilities. MSE is generally positively related to task performance and can also positively relate to performance when manipulated in laboratory settings. For example, Cervone and Peake ([1986](#)) used anchoring, a cognitive bias formed by individuals when presented with an initial piece of information (the anchor), to increase or decrease MSE, and this manipulation affected participants' performance. Additionally, providing encouraging, positive feedback has led to increased MSE and better performance but negative feedback has decreased MSE and performance (Sanna and Pusecker [1994](#)). Thus, MSE may be a malleable and potentially useful tool to alter participants' memory performance.

Together, the accuracy of metacognitive monitoring, the efficiency of metacognitive control processes, fluid intelligence, and MSE all may play a role in maximizing strategic control and selective recall for valuable information. Specifically, participants with greater fluid intelligence may be better able to assess their learning (as measured by JOLs) and subsequently spend more time studying valuable information yet to be learned. Similarly, participants with increased MSE may better remember information and also be more selective for high-value items as a result of increased effort and motivation or better task strategy.

The current study

Although there is some debate as to the role of cognitive abilities other than metacognition in value-directed remembering (e.g., working memory; Griffin et al. [2019](#); Robison and Unsworth [2017](#)), in the current study, we investigated how fluid intelligence and MSE relate to selectivity for high-value information and metacognitive accuracy in various value-directed remembering (VDR) paradigms. In Experiment 1, we attempted to manipulate MSE by

varying the question order on a fluid intelligence task to determine subsequent differences in recall performance, selectivity, and metacognitive accuracy in a VDR task. In Experiment 2, we allowed participants to self-pace their study time to assess the relationship between metacognitive monitoring and control, as well as the relationship between metacognitive accuracy and fluid intelligence. In Experiment 3, we added consequences for misguided metacognition to further assess the impact of item value and task experience on another form of metacognitive accuracy.

In each experiment, we expected participants' monitoring assessments and later recall to be influenced by the value of each word, consistent with prior work (e.g., Soderstrom and McCabe 2011; Yu et al. 2020). Specifically, we expected participants to best remember high-value words and for participants' judgments to map on to their selectivity. We also expected higher fluid intelligence to be related to increased selectivity for high-value words and metacognitive accuracy as the ability to reason, think abstractly, and solve problems may be related to the propensity to maximize memory utility by being selective for valuable information at the expense of less valuable information.

Experiment 1

In Experiment 1, we attempted to manipulate MSE and subsequent recall performance by having participants complete a fluid intelligence task in either ascending or descending order of difficulty before completing a VDR task. Prior research has explored the effect of different item-difficulty orders on test anxiety (Smouse and Munz 1968), performance (Klosner and Gellman 1973), and confidence (Weinstein and Roediger III 2010); however, research has yet to investigate the effect of different item-difficulty orders on MSE. Although most research reports no relationship between the order of questions and performance (see Hauck et al. 2017 for a review), compared to a descending order of difficulty, we expected participants completing the questions in ascending order to have elevated MSE as a result of better early task performance serving as an anchor for MSE judgments. Furthermore, we expected increased fluid intelligence to relate to elevated selectivity for high-value words, task scores, and metacognitive accuracy.

Method

Participants After exclusions, participants were 73 undergraduate students (age: $M = 20.19$, $SD = 1.93$) recruited from the University of California Los Angeles Human Subjects Pool who received course credit for their participation. The experiment was conducted online and lasted approximately 30 min. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in 1 exclusion from the ascending order group and 2 exclusions from the descending order group.

Materials and procedure. Raven's progressive matrices (RPM) Participants first completed the Raven's Progressive Matrices (RPM; Raven 1938), a non-verbal test of abstract reasoning and fluid intelligence (e.g., Jarosz et al. 2019; Staff et al. 2014). The task is composed of 12 problems (of varying difficulty), each of which presents participants with a pattern that has a piece missing. Participants were instructed to select the option (out of eight choices) that

correctly completes the pattern and then indicate their confidence in the accuracy of each response (from 0 to 100 with 0 being not confident and 100 being very confident). The task was self-paced, was not limited in time, and question difficulty was determined by Raven (1938) based on indices of participants' performance, such as mean response latency and accuracy rate. Participants were randomly assigned to complete the RPM with questions sequenced in either ascending or descending order of difficulty, and fluid intelligence scores were calculated as the proportion correct.

Memory self-efficacy questionnaire (MSEQ) Participants next completed the Memory Self-Efficacy Questionnaire for Items to assess participants' perception of their general memory abilities (MSEQ-I; Berry et al. 2013). The questionnaire consisted of 14 questions in which participants provided judgments of their ability to remember a list of items. They rated how confident they were that they could achieve a certain level of performance by selecting percentage responses ranging from 0 to 100% (in 10% increments). MSE scores were calculated by averaging participants' confidence ratings across all questions. See [Appendix](#) for the full MSEQ-I.

Value-directed remembering (VDR) After completing the RPM and the MSEQ-I, participants completed a value-directed remembering task. In this task, participants were presented with a series of to-be-remembered words with each word paired with an associated value between 1 and 12, indicating how much the word was "worth" (e.g., table: 5, toast: 12, plum: 7). Each point value was used only once within each list and the order of the point values within lists was randomized. The stimulus words were presented for 3 s each, were nouns that contained between four and seven letters, and had an everyday occurrence rate of at least 30 times per million (Thorndike and Lorge 1944). Participants were told that their score would be the sum of the associated values of the words they recalled (e.g., $5 + 12 + 7 = 24$) and that they should try to maximize their score.

After each word was presented, participants made a judgment of learning (JOL). Participants answered with a number between 0 and 100, with 0 meaning they definitely would not remember the word and 100 meaning they definitely would remember the word. Participants were given as much time as they needed to make their judgments. After the presentation of all 12 word-number pairs in each of the eight lists, participants were given a 20-s free recall test in which they had to recall as many words as they could from the list (they did not need to recall the point values). Immediately following the recall period, participants were informed of their score for that list but were not given feedback about specific items.

In addition to their point scores, participants were scored for efficiency via a selectivity index. For this metric, we calculated each participant's recall score relative to their chance and ideal score. The ideal score consisted of the sum of only the highest values for the particular number of words recalled. For example, if a participant remembered 3 words, ideally those words would be paired with the three highest values (e.g., 12, 11, 10). Chance scores reflected no attention to value and were calculated as the product of the average point value and the number of recalled words. At chance, the score in our example would be 6.5 multiplied by the number of recalled words. If a participant only recalled words paired with the highest values, the resulting selectivity score would be 1 while a participant who only recalled words paired with the lowest values would receive a selectivity score of -1 . Scores close to 0 indicate that a participant's recall was not sensitive to point values (see Castel et al. 2002 for more details).

Encoding strategies At the end of the VDR task, participants were asked to report which strategies (if any) they had used. Specifically, they were given a list from which they had to choose between reading each word as it appeared, repeating the words as much as possible, using sentences to link the words together, developing mental images of the words, grouping the words in a meaningful way, or utilizing some other strategy (participants could select some, none, or all of the strategies). To examine variation in encoding strategies, we computed an effective strategies variable which was the proportion of effective strategies reported as used by each participant. Prior research has indicated that effective encoding strategies lead to better recall performance and include imagery, sentence generation, and grouping, while less effective strategies involve passive reading and rote repetition (Hertzog et al. 1998; Richardson 1998; Unsworth 2016). In the present study, we coded self-reported encoding strategies based on their level of effectiveness and differentiated less effective strategies and strategies that support deeper levels of processing. Specifically, we computed an effective strategies variable which was the proportion of effective strategies reported as used by participants (i.e., using sentences to link the words together, developing mental images of the words, and grouping the words in a meaningful way).

Results

Although we initially hypothesized that manipulating the order of questions of the Raven's Progressive Matrices would result in differences in MSE and VDR, there were no differences of interest as a function of question order, consistent with prior work on item-difficulty order and performance (Hauck et al. 2017). Thus, we collapsed results across conditions for all subsequent analyses to investigate the relationships between selectivity for valuable information, metacognitive accuracy, strategy use, and fluid intelligence. Correlations between variables of interest can be seen in Table 1.

In our examination of VDR performance, we treated those data as hierarchical or clustered (i.e., multilevel), with items nested within individual participants. This approach helps to control for variation between individual participants and accounts for the non-independence of observations (i.e., 8 lists completed by the same individual are not independent observations). Also, multilevel approaches can account for an unequal number of observations across groups and participants and allow for both categorical and continuous predictor variables (Bolger and Laurenceau 2013; Gelman and Hill 2007; Jaeger 2008; Kenny et al. 1998; McElreath 2016). Thus, we used multilevel models (MLMs; sometimes also called mixed-effects or hierarchical models) in the present study. For all linear models, we used restricted maximum likelihood estimation (REML) to estimate coefficients, which is robust to small sample sizes at level-2 (i.e., the participant level in the current study; see McNeish 2017).

Because memory performance at the item-level was binary (i.e., correct or incorrect), we conducted a logistic MLM to assess performance. As a result, the regression coefficients are given as logit units, or the log odds of being correct. We report exponential betas (e^B), which give the coefficient as an odds ratio (e.g., the odds of being correct divided by the odds of being incorrect). Thus, e^B can be interpreted as the extent to which the odds of being correct changed with values greater than 1 representing an increased likelihood of being correct, values less than 1 representing a decreased likelihood of being correct, and a value of 1 indicating no change. Also, we report 95% confidence intervals for e^B because odds ratios are nonlinear and confidence intervals can be asymmetric. Lastly, centering of predictor variables

Table 1 Pearson (*r*) correlations between the primary variables of interest (collapsed across conditions) in Experiment 1

Measure	1	2	3	4	5	6	7	8	9	10	11
1) Raven's progressive matrices score	–										
2) Raven's progressive matrices response time	.408***	–									
3) MSEQ-I Score	.255*	.209	–								
4) Proportion of words recalled	.195	.122	.163	–							
5) Point scores	.286*	.124	.189	.930***	–						
6) Selectivity scores	.265*	.021	.073	–.050	.314**	–					
7) Reported effective strategy use	.074	.172	–.096	.168	.125	–.079	–				
8) Judgments of learning x value Pearson's <i>r</i>	.165	–.025	–.097	.069	.302**	.615***	–.084	–			
9) Output position x value Pearson's <i>r</i>	–.232*	–.006	–.232	–.079	–.209	–.367**	.013	–.361**	–		
10) Calibration	.081	.023	.161	–.210	–.177	.058	–.048	.039	–.036	–	
11) Resolution	.155	.012	–.152	.133	.268*	.388***	–.185	.328**	.023	–.271*	–

* = $p < .05$. ** = $p < .01$, *** = $p < .001$

in MLMs can be important for various reasons, and there are typically two ways to center item-level (level 1) predictors: around the grand mean (i.e., the average value across all observations), which we refer to as grand mean centering (GMC) or around the cluster mean (i.e., each level-2 unit's average value), which we refer to as cluster-based centering (CBC). Because level-2 variables only have one source of variance, they are always grand mean-centered. Cluster-based centering is important for isolating level-1 from level-2 effects, which becomes particularly important for interaction effects (Enders and Tofghi 2007; Ryu 2015). Thus, in all MLMs with an interaction term, we use CBC for level-1 predictors.

To first examine the proportion of words recalled ($M = .50$, $SD = .13$) as the task endured, a logistic MLM with accuracy (level 1) modeled as a function of list (level 1, GMC) revealed that list significantly predicted accuracy [$e^B = 1.08$, CI: 1.06–1.10, $z = 6.90$, $p < .001$] such that task experience resulted in greater recall. Similarly, a linear MLM with selectivity scores (level 2; $M = .17$, $SD = .24$) modeled as a function of list showed that list predicted selectivity [$b = .03$, CI: .02–.03, $t(6804) = 13.32$, $p < .001$] such that participants became more selective with increased task experience.

To determine if participants strategically organized their recall, we computed a Pearson correlation for each participant between each item's output position (with larger numbers meaning later output) and its value. A strong negative correlation would indicate that participants recalled high-value items before low-value items and a positive correlation would indicate the recall of low-value items before high-value items. While these correlations ($M =$

-.08, $SD = .23$) were different than 0 [one sample t -test: $t(72) = -2.87$, $p = .005$, $d = -.34$], a repeated measures ANOVA (8 levels) did not reveal a main effect of list [$F(7, 469) = 1.71$, $p = .104$, $\eta^2 = .03$]. Thus, participants generally recalled high-value items before low-value items and this did not change with task experience.

Most measures of monitoring, such as JOLs, are assessed as a probability, or percentage likelihood (same scale as the probability of recall), allowing for measures of *absolute* and *relative accuracy* (see Higham et al. 2016; Rhodes 2016). Absolute accuracy (i.e., *calibration*), is the overall relationship between judgment and performance and is calculated as the difference between mean judgments and the percentage of items recalled. For perfect calibration, participants' scores would be zero indicating a direct correspondence between prediction and recall. Results revealed that participants were well calibrated such that calibration ($M = .76$, $SD = 18.77$) was not different than 0 [one sample t -test: $t(72) = .35$, $p = .731$, $d = .04$]. Additionally, a repeated measures ANOVA (8 levels) on calibration revealed a main effect of list such that participants were initially overconfident but calibration improved with task experience [Mauchly's $W = .17$, $p < .001$, Huynh-Feldt corrected results: $F(4.75, 341.89) = 16.78$, $p < .001$, $\eta^2 = .19$].

Relative accuracy (i.e., *resolution*) is the degree to judgments discriminate between items that are or are not remembered and is often measured by Gamma correlations between each item's JOL and recall for each participant (see Masson and Rotello 2009 for alternative approaches). A perfect correlation between judgment and performance would exemplify the ability to distinguish between what will or will not be remembered; the individual remembers what they say they will remember. We computed Gamma correlations for each participant and these correlations ($M = .38$, $SD = .34$) were different than 0 [one sample t -test: $t(72) = 9.66$, $p < .001$, $d = 1.13$] indicating that participants' JOLs were relatively accurate. However, a repeated measures ANOVA (8 levels) on resolution did not reveal a main effect of list [$F(7, 378) = 2.48$, $p = .017$, $\eta^2 = .04$] indicating that relative accuracy did not change with task experience.

To further examine whether higher JOLs was related to better accuracy, we ran a logistic MLM with item-level accuracy measured as a function of JOLs (GMC), controlling for value and list, which showed that JOLs were a significant predictor of later accuracy [$e^B = 1.02$, CI: 1.016–1.020, $z = 17.89$, $p < .001$]. In other words, for each one-point increase in an item's JOL, the odds of remembering the word was expected to increase by 1.02. Thus, participants better remembered items after giving higher JOLs, suggesting they were generally metacognitively accurate.

In addition to metacognitive accuracy, we were also interested in whether participants were metacognitively aware of their selectivity. To assess the influence of value on participants' metacognitive judgments, we ran an MLM with item-level JOLs modeled as a function of the item's value (GMC), controlling for list (GMC). The analysis revealed that point value was a significant positive predictor of JOLs [$b = 1.85$, CI: 1.67–2.02, $t(6896) = 20.74$, $p < .001$] such that a one-point increase in an item's value is predicted to result in an increase of 1.85 in JOL, controlling for list position. Thus, participants were metacognitively aware of value effects on memory performance.

Next, to examine the relationship between fluid intelligence ($M = .58$, $SD = .27$) and VDR scores (sum of values of recalled words; $M = 41.94$, $SD = 10.62$), we used MLM to examine the predictive value of RPM score on average VDR score, taking into account both the number of words recalled and their point value. In this model, both RPM and average VDR scores were level-2 variables (e.g., collected at the participant level), and RPM performance was

centered around the grand mean (GMC). Results revealed that RPM score was a significant predictor of VDR scores [$b = 11.29$, CI: 2.39–20.20, $t(71) = 2.49$, $p = .015$], indicating that higher RPM scores predicted higher VDR scores.

Lastly, to examine differences in metacognitive accuracy as a function of fluid intelligence, we used a logistic MLM (Murayama et al. 2014). We modeled item-level accuracy as a function of item-level JOLs, participant-level (e.g., level-2) RPM score, and the interaction between JOLs and RPM score. To accurately represent the unique within-person effects, it is important to center item-level variables around the cluster means (e.g., each participant's JOLs are centered around their personal average JOL), rather than a grand mean (Enders and Tofghi 2007). Thus, JOLs were centered around participant means, while RPM scores were centered around the grand mean. Again, accuracy was modeled logistically (0 = incorrect, 1 = correct), so exponential betas (e^B) are reported.

Results revealed that JOLs were a significant predictor of accuracy for those who had average fluid intelligence scores [$e^B = 1.02$, CI: 1.018–1.022, $z = 19.40$, $p < .001$] but fluid intelligence scores were not a significant predictor of accuracy for items that were at participants' mean JOL [$e^B = 1.50$, CI: .90–2.48, $z = 1.56$, $p = .119$]. Of most interest, the interaction between JOL and fluid intelligence was significant [$e^B = 1.01$, CI: 1.01–1.02, $z = 3.50$, $p < .001$] indicating that an increase in fluid intelligence is expected to enhance the relationship between JOLs and recall. Specifically, this coefficient indicates that the relationship between JOLs and later recall (e.g., the slope of the regression line) is expected to increase by an odds ratio of 1.01 given a one-unit increase in RPM score (see Fig. 1).

We probed this interaction to determine the relationship between JOLs on later recall at three values of RPM scores: the mean and one standard deviation above and below the mean RPM score. Importantly, these values don't necessarily represent groups of participants, but rather the expected relationship between JOLs and recall performance for an example participant with a particular score on the RPM task. This revealed that JOLs were a significant predictor of accuracy for those who were one standard deviation below the mean [$e^B = 1.02$, CI: 1.01–1.02, $z = 11.71$, $p < .001$], at the mean [$e^B = 1.02$, CI: 1.018–1.022, $z = 19.40$, $p < .001$], and for those who were one standard deviation above the mean on the RPM task [$e^B = 1.02$, CI: 1.02–1.03, $z = 16.10$, $p < .001$], suggesting that most participants were metacognitively accurate, but those with higher RPM scores were more so. In other words, participants with higher fluid intelligence had a stronger relationship between JOLs and recall than those with lower fluid intelligence, suggesting they may be more metacognitively accurate.

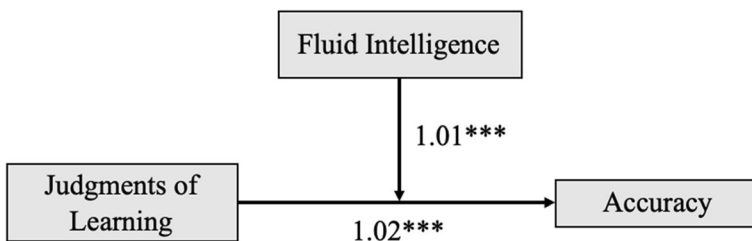


Fig. 1 The relationship between a person's judgments of learning and item-level recall accuracy and the extent to which this relationship is moderated by a person's fluid intelligence in Experiment 1. Coefficients shown are exponential betas (e^B), or odds ratios. * $p < .05$. ** $p < .01$. *** $p < .001$

Discussion

In Experiment 1, we manipulated the order of questions on the RPM to investigate potential differences in MSE and VDR as a consequence of completing easier or more difficult questions first. After completing the RPM task in ascending difficulty, we expected that these participants would report higher MSE as a result of better early task performance serving as an anchor for MSE judgments. However, results revealed no differences in MSE or any other variables of interest as a function of question order, so we collapsed the remaining analyses across conditions. Results revealed that the proportion of words recalled, selectivity for high-value words, and task scores (sum of values of recalled words) all increased with experience. In terms of the organization of participants' recall, participants tended to recall high-value items before low-value items but this did not change with task experience. Furthermore, participants were generally metacognitively accurate, both in terms of absolute and relative accuracy, and were metacognitively aware of their selectivity for valuable items and this awareness increased as a function of list.

The results observed in Experiment 1 are consistent with previous research on the influence of item value on selectivity, suggesting that people tend to focus more on high-value information and less on low-value information to maximize gains (Ariel et al. 2009; Castel 2008; Castel et al. 2009; Castel et al. 2002; Castel et al. 2007). Additionally, we demonstrated that increased fluid intelligence predicts greater VDR task scores, and participants with higher fluid intelligence had a stronger relationship between JOLs and recall than those with lower fluid intelligence, indicating that they were more metacognitively accurate (see Fig. 1). Thus, not only were participants with greater fluid intelligence more selective, but they were also more accurate in their metacognitive judgments.

Fluid intelligence positively correlated with MSE such that participants who performed better on the RPM task reported greater belief in their memory abilities. Since past successes and failures are known to influence MSE (Bandura 1977), heightened fluid intelligence may be related to greater feelings of mastery after success on previous and/or similar tasks, leading to higher self-efficacy judgments. However, MSE did not correlate with any other variables of interest, supporting accounts that MSE may not directly relate to performance (see Beaudoin and Desrichard 2011).

Fluid intelligence positively correlated with selectivity scores such that higher fluid intelligence was associated with better recall of higher valued items compared to less valuable items. However, fluid intelligence was not associated with the number of words recalled but was positively related to task scores (the sum of the values paired with recalled words). Moreover, fluid intelligence negatively correlated with the organization of retrieval such that participants with increased fluid intelligence demonstrated a greater tendency to recall valuable items before less valuable items. Thus, the positive relationships between fluid intelligence, selectivity, and task scores (but no relationship with total recall), together with the tendency to prioritize recall for high-value items, indicate that better reasoners engage in a more strategic utilization of cognitive resources to optimize performance in VDR. Finally, reported effective strategy use did not relate to task performance or fluid intelligence, contrary to previous findings (e.g., Ariel et al. 2015; Hennessee et al. 2019), suggesting that utilizing effective encoding strategies may not be necessary for good performance.

Experiment 2

In Experiment 1, participants were selective and metacognitively aware of their selectivity. Additionally, selectivity and task scores were positively related to fluid intelligence such that greater fluid intelligence was associated with better task scores, likely as a result of enhanced selectivity. In Experiment 2, rather than fixed study time, we investigated the role of fluid intelligence in recall performance, selectivity, and metacognitive accuracy when participants self-paced their study time to determine whether metacognitive awareness of selectivity maps is reflected in metacognitive control mechanisms. We expected participants to spend more time studying high-value words compared to low-value words and to again be metacognitively aware of their selectivity by giving high-value words high JOLs and low-value words low JOLs (see also Soderstrom and McCabe 2011). Similar to Experiment 1, we also expected increased fluid intelligence to be associated with selectivity (via selective allocation of study time), better recall, task scores, and metacognitive accuracy.

Method

Participants After exclusions, participants were 42 undergraduate students (age: $M = 19.90$, $SD = 1.36$) recruited from the University of California Los Angeles Human Subjects Pool and received course credit for their participation. The experiment was conducted online and lasted approximately 30 min. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in 3 exclusions.

Materials and procedure The materials in Experiment 2 were similar to those in Experiment 1. However, all participants completed the Raven's Progressive Matrices in the standard order (ascending difficulty; see Engle et al. 1999; Raven and Raven 2003), again followed by the MSEQ-I, and the VDR task. On the VDR task, instead of viewing each word for 3 s, participants self-paced their study time; participants were permitted to study each word for as long as they liked before advancing to the next word. No other specific instructions were given.

Results

Correlations between variables of interest can be seen in Table 2. To investigate study time per word (in seconds) throughout the VDR task ($M = 3.65$, $SD = 2.94$), we conducted a series of linear MLMs with items (level 1) clustered within individuals (level 2). First, examining the effect of list on study time, we found that list (level 1, GMC) predicted study time (in seconds), [$b = -.11$, CI: $-.18 - -.03$, $t(4895) = 2.86$, $p = .004$] such that each subsequent list was predicted to result in .11 fewer seconds of study time per word. A similar analysis of the proportion of words recalled (level 2) revealed that the average proportion of words recalled ($M = .49$, $SD = .12$) increased with task experience [$b = .01$, CI: $.01-.01$, $t(1983) = 11.59$, $p < .001$]. Further, analysis of point scores (level 2; $M = 42.78$, $SD = 11.57$) revealed that VDR scores increased with task experience [$b = 1.62$, CI: $1.48-1.76$, $t(2893) = 22.61$, $p < .001$] such that participants were expected to have a 1.62 point increase in score with each additional list of experience. Finally, participants also showed elevated selectivity scores (level

2; $M = .26$, $SD = .28$) with increased task experience [$b = .06$, $CI: .05-.06$, $t(3845) = 23.40$, $p < .001$].

To investigate whether participants recalled valuable items before less valuable items, a Pearson correlation was computed for each participant between each item's output position and its value. In contrast to Experiment 1, these correlations ($M = -.02$, $SD = .24$) were not different from 0 [one sample t -test: $t(41) = -.43$, $p = .668$, $d = -.07$] and a repeated measures ANOVA (8 levels) on this tendency did not reveal a main effect of list [$F(7, 273) = .71$, $p = .665$, $\eta^2 = .02$]. Thus, participants did not prioritize the order of their recall according to value and this did not change with task experience.

To determine if participants were metacognitively aware of their selectivity as in Experiment 1, we again ran an MLM with item-level JOLs modeled as a function of the item's value (GMC) and list (GMC). The analysis revealed that value was a significant positive predictor of JOL [$b = 2.88$, $CI: 2.67-3.09$, $t(3891) = 27.09$, $p < .001$], such that a one-point increase in an item's point value was expected to result in a 2.88 increase in JOL, controlling for list position.

To examine metacognitive accuracy, we first calculated participants' calibration. Results revealed that participants were well calibrated such that calibration ($M = 1.49$, $SD = 20.95$) was not different than 0 [one sample t -test: $t(41) = .46$, $p = .648$, $d = .07$]. Additionally, a repeated measures ANOVA (8 levels) revealed a main effect of list such that participants were initially overconfident but calibration improved with task experience [Mauchly's $W = .13$, $p < .001$, Huynh-Feldt corrected results: $F(4.83, 183.58) = 4.57$, $p < .001$, $\eta^2 = .11$]. In terms of resolution, a Gamma correlation between recall and each item's JOL was computed for each participant. These correlations ($M = .42$, $SD = .34$) were different than 0 [one sample t -test: $t(41) = 8.07$, $p < .001$, $d = 1.25$] and a similar repeated measures ANOVA (8 levels) revealed a main effect of list such that participants' relative accuracy increased with task experience [$F(7, 217) = 2.71$, $p = .010$, $\eta^2 = .08$]. To further assess whether JOLs were predictive of later accuracy, we also conducted an MLM with binary accuracy modeled as a function of JOL (GMC), controlling for value (GMC) and list (GMC). The analysis revealed a significant effect of JOL [$e^B = 1.025$, $CI: 1.022-1.028$, $z = 15.52$, $p < .001$], indicating that making higher JOLs for a given item was associated with a greater likelihood of correctly recalling that item, while controlling for list position and item value.

We were also interested in the extent to which metacognition could explain the relationship between item value and later recall. To test this, we conducted a mediation analysis with study time as a mediator between item value and the later likelihood of correct recall. The conceptual model (and path coefficients) are depicted in Fig. 2. We ran a multilevel mediation model using a Bayesian framework to estimate the mediated effect (interpreted similarly to regression framework; see Vuorre and Bolger 2018). The Bayesian approach allowed us to better account for uncertainty in the estimation process than the traditional maximum likelihood method, and provides a full distribution of the estimate, allowing for some interpretation of its variability (Kruschke 2014). Furthermore, when testing for mediation, the model consists of various paths – first from the focal predictor to the mediator (a path) and also from the mediator to the outcome (b path). Also, there is a total effect, or the effect of the predictor on the outcome without considering the mediator (c path) and a direct effect (c' path), which is the effect of the predictor on the outcome while controlling for the mediator. The indirect effect (e.g., the influence of the predictor on the outcome *through* the mediator) is calculated as the product of the a and b coefficients (i.e., $a*b$). There are many ways to test for mediation (Fritz and MacKinnon 2007; Fritz et al. 2012; MacKinnon et al. 2004), but here we use Bayesian

Table 2 Pearson (*r*) correlations between the primary variables of interest in Experiment 2

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) Raven's progressive matrices proportion correct	—												
2) Raven's progressive matrices average response time	.397**	—											
3) MSEQ-I score	.492**	.130	—										
4) Proportion of words recalled	.495***	.208	.302	—									
5) Point scores	.435**	.089	.206	.916***	—								
6) Selectivity scores	.051	-.221	-.132	.259	.622***	—							
7) Reported effective strategy use	.170	.287	.336*	.323*	.173	.235	—						
8) Study time per word	.385*	.473**	.322*	.620***	.493***	-.046	.222	—					
9) Study time per word x value Pearson's <i>r</i>	.111	-.021	.227	.341*	.470**	.477**	.038	.200	—				
10) Judgments of learning x value Pearson's <i>r</i>	.203	-.023	.044	.300	.563***	.776***	-.065	.011	.425**	—			
11) Output position x value Pearson's <i>r</i>	-.112	-.178	-.108	-.144	-.192	-.201	-.136	.012	-.060	-.395***	—		
12) Calibration	-.195	-.024	.302	-.231	-.327*	-.342*	.168	-.123	-.162	-.156	.005	—	
13) Resolution	-.068	.039	-.225	.099	.342*	.656***	.009	-.154	.368*	.405**	-.235	-.115	—

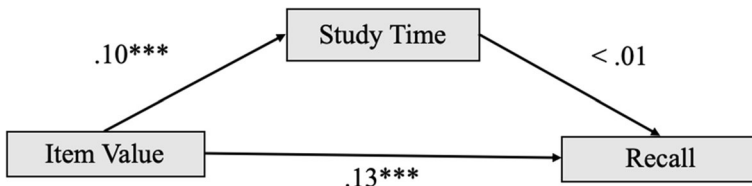


Fig. 2 Mediation analysis showing an item's study time (in seconds) as a mediator of its likelihood of recall in Experiment 2. Coefficients shown are unstandardized path coefficients, with coefficients for recall shown in logit units. * $p < .05$. ** $p < .01$. *** $p < .001$

estimation as described in Vuorre and Bolger (2018) using the *bmlm* package in R (Vuorre 2017).

We predicted that an item's value would be predictive of its probability of later recall, but that study time may mediate this relationship, such that items with higher values would be studied longer, and study time would relate to accuracy at the item-level. In other words, study time may explain some of the relationship between an item's value and later recall due to participants' metacognitive control of their selectivity and the effects of value on memory. To test for multilevel mediation, study time was modeled as a function of item-level point value (a path), and item-level accuracy was subsequently modeled as a function of item-level point value (c' path) and item-level study time (b path; see Fig. 2).

First looking at the a path, the analysis revealed that item-level value was a significant predictor of study time [$b = .10$, CI: $.05-.15$, $t(3892) = 3.97$, $p < .001$] indicating that a one-point increase in an item's point value was predicted to result in a $.10$ s increase in study time for that word. However, the b path was not significant [$b < .01$, $e^B = 1.00$, CI: $.99-1.02$, $z = .51$, $p = .614$] indicating that an increase in study time did not result in a greater likelihood of correct recall and the c' path remained significant while controlling for study time [$b = .13$, $e^B = 1.14$, CI: $1.12-1.17$, $z = 13.47$, $p < .001$], suggesting that value was a significant predictor of accuracy while controlling for study time. Next, looking at the test of mediation, we used 10,000 iterations to estimate the indirect effect, with 95% credible intervals (similar interpretation to confidence intervals). The estimated indirect effect was not significant [indirect effect = $.00$, CI: $.00-.01$] as the 95% credible interval included zero. Additionally, the proportion of the effect that was mediated was estimated to be $.02$ [CI: $-.02-.06$], again suggesting that study time was not a significant mediator of the relationship between point value and correct recall.

To assess the relationship between participants' fluid intelligence scores ($M = .65$, $SD = .22$) and their metacognitive accuracy, we conducted a multilevel analysis on item-level accuracy modeled as a function of item-level JOLs (CBC), RPM score (level 2, GMC), and their interaction. This analysis revealed that JOLs were a significant positive predictor of accuracy for those at the mean RPM score [$e^B = 1.03$, CI: $1.02-1.03$, $z = 18.00$, $p < .001$] and that RPM score positively predicted accuracy at each participant's mean JOL rating [$e^B = 2.57$, CI: $1.09-6.07$, $z = 2.15$, $p = .031$]. However, in contrast to what was found in Experiment 1, the interaction between RPM scores and JOLs was not significant [$e^B = 1.00$, CI: $.99-1.02$, $z = .71$, $p = .480$] suggesting that the relationship between JOLs and accuracy did not vary according to the level of fluid intelligence (see Fig. 3).

Because we also had a measure of metacognitive control (i.e., study time), we conducted a similar analysis on item-level accuracy with item-level study time (CBC), participant-level RPM score (GMC), and the interaction of these two variables as predictors (see Fig. 4). This analysis revealed that study time was a significant positive predictor of accuracy for those of

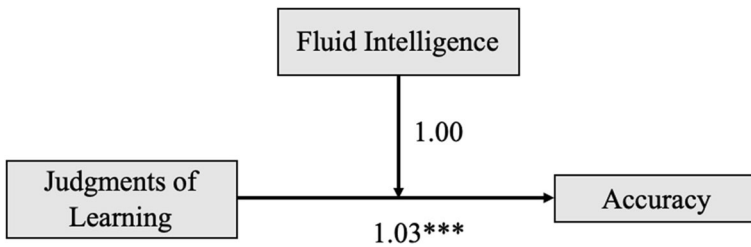


Fig. 3 The relationship between a person's judgments of learning and item-level recall accuracy and the extent to which this relationship is moderated by a person's fluid intelligence in Experiment 2. Coefficients shown are exponential betas (e^B), or odds ratios. * $p < .05$. ** $p < .01$. *** $p < .001$

average RPM score [$e^B = 1.05$, CI: 1.02–1.07, $z = 4.26$, $p < .001$] and that RPM score was a significant positive predictor of accuracy at each participant's mean study time [$e^B = 2.13$, CI: 1.16–3.91, $z = 2.42$, $p = .015$]. Of most interest, the interaction was significant [$e^B = .84$, CI: .76–.93, $z = 3.36$, $p < .001$] such that the relationship between study time and accuracy decreased as fluid intelligence increased. In probing the interaction, we found that the expected relationship between study time and accuracy for a participant who is one standard deviation below the mean on the RPM task was significant [$e^B = 1.09$, CI: 1.04–1.13, $z = 4.01$, $p < .001$] and also for a participant at the mean [$e^B = 1.04$, CI: 1.02–1.07, $z = 4.26$, $p < .001$]. However, there was not a significant relationship between study time and accuracy for a participant who scored one standard deviation above the mean on the RPM task [$e^B = 1.01$, CI: .99–1.02, $z = .72$, $p = .470$]. Thus, study time only related to recall accuracy for those who scored near or below the mean for fluid intelligence.

Discussion

Unlike Experiment 1, all participants in Experiment 2 completed the Raven's Progressive Matrices in ascending order of difficulty before completing the MSEQ-I and the VDR task. We were interested in whether allowing participants to self-pace their study time would result in a greater allocation of study time for high-value words and subsequent better recall of these words. We also further assessed the role of fluid intelligence in recall performance, selectivity, and metacognitive accuracy. Results revealed that study time decreased as the task endured but participants generally studied valuable items more than low-value items. Moreover, selectivity in recall was related to this strategic study pattern as well as the tendency to give high-valued words high JOLs and low-valued words low JOLs, indicating participants' awareness of selectivity both in terms of metacognitive monitoring and control. Similarly, better point

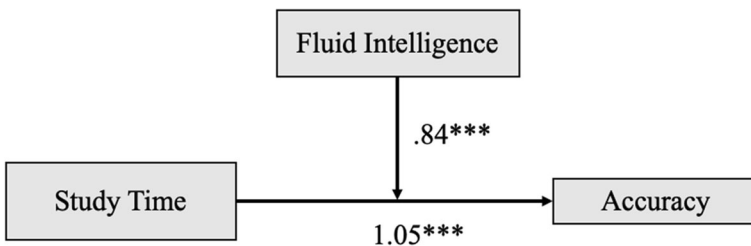


Fig. 4 The relationship between an item's study time and item-level recall accuracy and the extent to which this relationship is moderated by fluid intelligence in Experiment 2. Coefficients shown are exponential betas (e^B), or odds ratios. * $p < .05$. ** $p < .01$. *** $p < .001$

scores were related to the tendency to spend more time studying valuable items as well as the metacognitive awareness of selectivity. Results also revealed that reported effective strategy use was related to the proportion of words recalled, such that the participants with greater recall reported using more effective encoding strategies.

Additionally, the proportion of words recalled, point scores, and selectivity all increased with task experience, similar to Experiment 1, although participants did not prioritize the organization of their recall according to word value. Furthermore, participants were metacognitively aware of their selectivity, and this metacognitive awareness of selectivity, calibration, and resolution also improved with task experience. Thus, allowing participants to self-pace their study time likely better informed monitoring assessments, leading to increased relative accuracy as a function of task experience.

The tendency to spend more time studying high-value words and the tendency to give high-value words higher JOLs were related such that participants' judgments indicated that they believed that studying an item more would lead to better recall. This generally resulted in relatively accurate metacognitive monitoring, as measured by resolution, and resolution was positively related to scores and selectivity, indicating that accurate metacognition may play a role in optimizing memory efficiency. To further assess the relationship between value, monitoring, and recall, mediation analyses revealed that value was a significant predictor of JOLs, and JOLs were related to recall. Additionally, value influenced study time but an increase in study time did not result in a greater likelihood of correct recall. Thus, study time was not a significant mediator of the relationship between point value and correct recall.

Moderation models indicated that the relationship between JOLs and accuracy did not vary according to fluid intelligence scores, and study time was related to recall accuracy only for those who scored lower or average for fluid intelligence. Thus, the higher the fluid intelligence, the smaller the effect of study time on the accuracy of metacognitive monitoring. Furthermore, in Experiment 1, greater fluid intelligence was related to better selectivity for valuable information, but in Experiment 2, higher fluid intelligence was associated with a greater proportion of words recalled and better point scores, but not increased selectivity. Additionally, when able to self-pace study time, fluid intelligence positively related to study time, indicating that rather than selectivity remembering high-value words, participants spent more time studying each word to increase the number of words recalled to maximize task scores. As such, the proportion of words recalled strongly positively related to the amount of time spent studying the words. Thus, increased fluid intelligence was associated with better point scores by spending more time studying each word to recall more words rather than increasing point scores by being more selective.

Experiment 3

In Experiment 1, fluid intelligence was positively related to selectivity for valuable information but when self-pacing study time in Experiment 2, fluid intelligence was not related to selectivity. Rather, higher fluid intelligence was related to the tendency to maximize total recall to increase task scores. In Experiment 3, we further assessed the impact of item value and fluid intelligence on recall performance, selectivity, and metacognitive accuracy with fixed study time. However, rather than making JOLs for each word (as in Experiments 1 and 2), participants decided whether to "bet" on each word to instill consequences for misguided metacognition (see Hanczakowski et al. 2013; McGillivray and Castel 2011). Specifically,

while studying each word, if participants chose to bet on a word and later recalled it, they would receive the associated points, but if they failed to recall a word they bet on, they would lose the associated points. If participants chose not to bet on a word, they neither received the associated points if they recalled the word or were penalized if they failed to recall the word. We expected participants to bet more on high-value words and to improve task performance with increased experience. Additionally, we expected higher fluid intelligence to be associated with more strategic and accurate betting behavior, resulting in better task performance.

Method

Participants Participants were 37 undergraduate students (age: $M = 20.14$, $SD = 1.70$) recruited from the University of California Los Angeles Human Subjects Pool and received course credit for their participation. The experiment was conducted online and lasted approximately 30 min. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated) and this process resulted in no exclusions.

Materials and procedure The materials in Experiment 3 were similar to those in Experiment 2. All participants completed the Raven's Progressive Matrices in the same order (ascending difficulty) followed by the MSEQ-I, but a different version of the VDR task including a metamemory "betting" component (see McGillivray and Castel 2011). Rather than making JOLs, after each word was presented, participants had to decide if they wanted to "bet" on it (by clicking "yes" or "no"). Whether the participants chose to bet on a word or not, each word was displayed for 5 s. For the words participants bet on and later remembered, they received the associated points but if they failed to recall a word that they initially "bet" on, then participants lost those points. Conversely, if participants did not bet on a word, the points were not gained or lost regardless of whether the word was recalled. Point values (1–10, 15, and 20) were randomly paired with words within each list. The inclusion of the 15 and 20 point values was used to assess the impact of extreme incentive or loss potential (e.g., Loftus and Wickens 1970). Scores were calculated by summing the points associated with the words participants bet on and successfully recalled, and then subtracting the number of points associated with the words that were bet on but not recalled.

Results

Correlations between variables of interest can be seen in Table 3. To investigate performance throughout the VDR task, a logistic MLM with item-level accuracy modeled as a function of list (item-level, GMC) revealed a significant positive effect [$e^B = 1.04$, CI: 1.01–1.07, $z = 2.42$, $p = .016$] such that the likelihood of correct recall ($M = .43$, $SD = .13$) increased with task experience. Additionally, a linear MLM with list as a predictor of VDR score (level 2; $M = 26.28$, $SD = 27.50$) showed that scores were predicted to increase with task experience [$b = 3.12$, CI: 2.74–3.49, $t(3514) = 16.25$, $p < .001$]. However, another logistic MLM with list as a predictor of item-level betting (0 = no bet, 1 = bet) did not reveal an effect of list on the likelihood of betting ($M = .58$, $SD = .12$), [$e^B = 1.01$, CI: .98–1.04, $z = .30$, $p = .766$].

To determine if participants attempted to be selective, we conducted another logistic MLM on the likelihood of betting with item-level value (GMC) as a predictor. Results revealed that

Table 3 Pearson (r) correlations between the primary variables of interest in Experiment 3

Measure	1	2	3	4	5	6	7	8	9
1) Raven's progressive matrices proportion correct	–								
2) Raven's progressive matrices average response time	.579***	–							
3) MSEQ-I score	.039	–.094	–						
4) Proportion of words recalled	.033	.089	.156	–					
5) Point scores	.30	.257	.051	.849***	–				
6) Average number of bets	–.141	–.051	.237	.389*	–.034	–			
7) Reported effect strategy use	.058	.346*	.167	.428**	.437**	.226	–		
8) Output position x value Pearson's r	–.009	.047	.018	.158	.166	.063	.093	–	
9) Bets x value γ	.331*	.383*	.047	.482**	.731***	–.224	.406*	–.067	–

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

value had a significant positive effect on the likelihood of betting, such that participants were more likely to bet as word value increased [$e^B = 1.24$, CI: 1.21–1.26, $z = 21.10$, $p < .001$]. We also computed a Gamma correlation for each participant between whether or not they bet on each item and each item's value. These correlations ($M = .55$, $SD = .29$) were different than 0 [one sample t -test: $t(36) = 11.54$, $p < .001$, $d = 1.90$] and a repeated measures ANOVA (8 levels) revealed a main effect of list such that the tendency to bet more on high-value words and less on low-value words increased with task experience [Mauchly's $W = .14$, $p = .001$, Huynh-Feldt corrected results: $F(5.45, 168.97) = 3.67$, $p = .003$, $\eta^2 = .11$].

To investigate if participants tended to recall valuable items before less valuable items, a Pearson correlation for each participant between each item's output position and its value. These correlations ($M = -.06$, $SD = .21$) were not different than 0 [one sample t -test: $t(36) = -1.73$, $p = .092$, $d = -.29$] and a repeated measures ANOVA (8 levels) on this tendency did not reveal a main effect of list [$F(7, 175) = .69$, $p = .679$, $\eta^2 = .03$] such that participants did not prioritize the order of their recall according to value and this did not change with increased task experience.

To examine whether participants' fluid intelligence ($M = .56$, $SD = .24$) was related to their metacognitive accuracy, we conducted a logistic MLM with item-level recall modeled as a function of item-level betting (0 = no bet, 1 = bet), participant-level RPM score (GMC), and the interaction of these two variables (see Fig. 5). Results revealed that items that were bet on were more likely to be remembered than those that were not for participants who were average on the RPM task [$e^B = 23.75$, CI: 19.15–29.45, $z = 28.84$, $p < .001$]. Furthermore, RPM score did not show a significant relationship with accuracy for items that were not bet on [$e^B = .53$, CI: .15–1.87, $z = .99$, $p = .324$] but the interaction between RPM scores and betting behavior was significant [$e^B = 3.72$, CI: 1.53–9.03, $z = 2.90$, $p = .004$] such that the difference in accuracy for items that were bet on and those that were not was greater for participants who had higher RPM scores.

In probing the interaction at the mean and plus or minus one standard deviation, we found that the mean difference in accuracy for items that were bet on and those that were not bet on was significant for those who were below average on the RPM task [$e^B = 17.37$, CI: 13.03–23.16, $z = 19.46$, $p < .001$], those who were at the mean [$e^B = 23.75$, CI: 19.15–29.45, $z = 28.84$, $p < .001$], and for those who were above average [$e^B = 32.45$, CI: 23.70–44.45, $z = 2.69$, $p < .001$]. However, the predictive value of betting on later correct recall increased with increasing fluid intelligence. In other words, participants

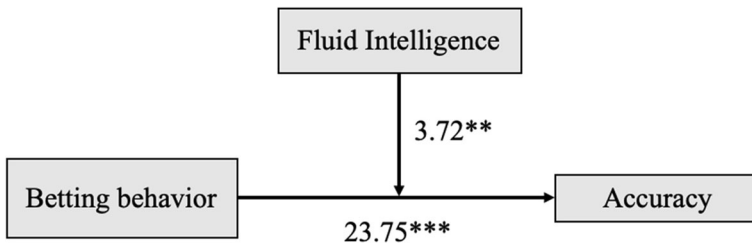


Fig. 5 The relationship between betting behavior and item-level recall accuracy and the extent to which this relationship is moderated by fluid intelligence in Experiment 3. Coefficients shown are exponential betas (e^B), or odds ratios. * $p < .05$. ** $p < .01$. *** $p < .001$

with higher fluid intelligence had a stronger relationship between betting on an item and later remembering that item than those with lower fluid intelligence.

Discussion

Analogous to Experiment 2, participants completed the RPM task in ascending order of difficulty, followed by the MSEQ-I, and a VDR task. However, in Experiment 3, we explored the role of fluid intelligence and consequences for misguided metacognition in a VDR task using a different measure of metacognitive monitoring by requiring participants to “bet” on each word, rather than providing JOLs. Results revealed that recall, point scores, and the tendency to bet more on high-value words and less on low-value words increased with task experience. However, the likelihood of betting did not change as a function of list but participants were more likely to bet as word value increased, indicating metacognitive awareness of selectivity. Further, selectively betting on high-value words and less on low-value words related to more words being recalled and better task performance. Together, these results suggest that more strategic betting and increased recall can enhance task performance.

In terms of fluid intelligence, unlike in Experiments 1 and 2, fluid intelligence was not directly related to task performance. However, fluid intelligence was positively related to the tendency to selectively bet more on high-value words and less on low-value words. Additionally, participants with higher fluid intelligence had a stronger relationship between betting on an item and later remembering that item than those who were lower on fluid intelligence. In other words, increased fluid intelligence related to more strategic and selective betting behavior, leading to better memory outcomes, further demonstrating participants’ metacognitive awareness of their selectivity with a binary form of metacognitive monitoring and consequences for forgetting.

Finally, the usage of more effective strategies was associated with a greater proportion of words recalled and higher task scores. Also, results revealed that the tendency to bet more on high-value words and less on low-value words was positively correlated with effective encoding strategy use such that participants using more effective encoding strategies tended to adopt better betting behavior. Collectively, the results further support participants’ metacognitive awareness of selectivity and the important role of fluid intelligence in selectivity.

General discussion

We are often exposed to much more information than we can process and remember. To function effectively, we tend to prioritize important information over peripheral information, and one's awareness or knowledge about their memory abilities (i.e., metamemory) may be crucial for such selectivity. There has been some disagreement about the role of fluid intelligence in metacognition (Maqsd [1997](#); Saraç et al. [2014](#); Sternberg [1981, 1985](#); Rozenčwajg [2003](#); Van der Stel and Veenman [2014](#)) and in the present study, we investigated the relationships between fluid intelligence, recall performance, selectivity, metacognitive accuracy, and metacognitive awareness of selectivity.

Previous work has revealed that accurate monitoring assessments should be sensitive to the cues that affect memory performance and impervious to those that have minimal effects (see Rhodes [2016](#)), consistent with the cue-utilization framework (Koriat [1997](#)). However, there are instances where the cues used to inform JOLs are unrelated or weakly related to actual memory performance, such as font size or word volume (see Rhodes and Castel [2008, 2009](#)), and relying on cues with poor predictive validity of later remembering can lead to a sub-optimal allocation of study time and poorer memory outcomes (e.g., Metcalfe and Finn [2008](#)).

When presented with words of different values, participants in the present study¹ generally rated valuable words as more likely to be remembered, rated low-value words as less likely to be remembered, and spent more time studying high-value words than low-value words. Participants were subsequently selective for high-value information in their recall, exemplifying accurate metacognition and demonstrating metacognitive awareness of their selectivity. Additionally, selectivity and task scores increased as the task endured, indicating that participants may become more aware of their limited memory capacity and more strategically selective with what they remember after gaining task experience (Ariel et al. [2009](#); Castel [2008](#); Castel et al. [2009](#); Castel et al. [2002](#); Castel et al. [2007](#); Elliott et al. [2020](#)).

Although item value was a significant predictor of JOLs and also influenced study time, study time did not mediate the relationship between point value and correct recall. Collectively, participants' metacognitive monitoring and control measures indicated that they believed that studying an item longer would lead to better recall, which generally resulted in accurate JOLs. This accuracy was positively related to scores and selectivity, indicating that accurate metacognition may play a role in optimizing memory efficiency. However, previous work has indicated that working memory is weakly related to selectivity (e.g., Castel et al. [2009](#); Griffin et al. [2019](#); Robison and Unsworth [2017](#)), thus, selectivity may rely on a strategic form of metacognition that is not purely memory-capacity based; abstract reasoning and problem-solving (in the form of fluid intelligence) may also play a crucial role in VDR.

While accurate metacognitive monitoring and control play a role in maximizing recall and strategic control for valuable information, the present study also revealed that fluid intelligence is an important component in optimizing task performance and the accuracy of metacognition. Fluid intelligence was generally related to better VDR scores (measured as the sum of point values associated with recalled words), however, the medium of achieving greater performance varied based on the pacing of the encoding phase. Specifically, when study time was fixed, higher fluid intelligence was related to enhanced selectivity for high-value words but

¹ Although our sample size is somewhat small for individual differences research, the use of multilevel regression models improves power compared to traditional ANOVAs. Additionally, we were able to find significant effects despite the smaller sample size; however, these findings should be replicated with larger samples in future work.

when able to self-pace study time, higher fluid intelligence was related to a more generous overall allocation of study time and subsequent greater recall. In other words, when study time was fixed, fluid intelligence correlated with selectivity for high-value words, but when able to self-pace the study time, participants did not limit their access to low-value words and spent more time studying each word, regardless of its value, to increase total recall.

Moreover, the relationship between JOLs and accuracy also varied according to fluid intelligence when study time was fixed but when self-pacing study time, only the relationship between study time and accuracy varied according to fluid intelligence scores. However, higher fluid intelligence scores were associated with progressively smaller effects of study time on metacognitive monitoring accuracy, thus, the current study suggests that fluid intelligence may play an important role in effective metacognition. Additionally, the positive relationships between fluid intelligence, selectivity, and task scores (but no relationship with total recall), together with the tendency to prioritize recall for high-value items, indicate that a better ability to use reasoning and logic to solve new problems involves a more strategic utilization of cognitive resources to optimize value-directed remembering performance.

In sum, the present study revealed that fluid intelligence plays an important role in participants' selectivity for high-value words, metacognitive awareness of selectivity, metacognitive monitoring and control, and task performance. Moreover, recall, selectivity, metacognitive awareness of selectivity, and metacognitive accuracy all increased with task experience, demonstrating participants' ability to improve their memory by utilizing cognitive resources more effectively. Overall, participants demonstrated awareness of the need to be selective to improve memory outcomes, and this tendency may be related to higher-level cognitive skills associated with problem-solving and reasoning. Thus, the strategic use of memory may be involved in focusing on important information, and the metacognitive processes that allow for this prioritization of memory may be related to more general problem-solving abilities that involve identifying important features of information to guide cognition in a broader context. While the present work shows correlational evidence using established tests of fluid intelligence and selective memory, additional research is needed to better understand the complex causal relationship between different forms of intelligence, metacognition, and the strategic use of memory.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11409-021-09265-9>.

Acknowledgments This research was supported in part by the National Institutes of Health (National Institute on Aging; Award Number R01 AG044335 to Alan D. Castel).

Open Practices Statement None of the experiments reported in this article were formally preregistered. Neither the data nor the materials have been made available on a permanent third-party archive; requests for the data or materials are available from the corresponding author upon reasonable request.

Declaration

Conflicts of Interest The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

References

- Ariel, R. (2013). Learning what to learn: The effects of task experience on strategy shifts in the allocation of study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1697–1711.
- Ariel, R., & Dunlosky, J. (2013). When do learners shift from habitual to agenda-based processes when selecting items for study? *Memory & Cognition*, *41*, 416–428.
- 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.
- Ariel, R., Price, J., & Hertzog, C. (2015). Age-related associative memory deficits in value-based remembering: The contribution of agenda-based regulation and strategy use. *Psychology and Aging*, *30*, 795–808.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*, 191–215.
- Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology*, *25*, 729–735.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*, 211–241.
- Berry, J. M. (1999). Memory self-efficacy in its social cognitive context. In T. M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 69–96). San Diego: Academic Press.
- Berry, J. M., Williams, H. L., Usualieva, A., & Kilb, A. (2013). Metacognitive awareness of the associative deficit for words and names. *Aging, Neuropsychology and Cognition*, *20*, 592–619.
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford.
- Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225–270). London: Academic Press.
- Castel, A. D., Balota, D. A., & McCabe, D. P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's disease. *Neuropsychology*, *23*, 297–306.
- Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085.
- Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, *35*, 689–700.
- Castel, A. D., McGillivray, S., & Friedman, M. C. (2012). Metamemory and memory efficiency in older adults: Learning about the benefits of priority processing and value-directed remembering. In M. Naveh-Benjamin & N. Ohta (Eds.), *Memory and aging: Current issues and future directions* (pp. 245–270). New York: Psychology Press.
- Cervone, D., & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology*, *50*, 492–501.
- 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.
- Dunlosky, J., & Ariel, R. (2011a). Self-regulated learning and the allocation of study time. *Psychology of Learning and Motivation*, *54*, 103–140.
- Dunlosky, J., & Ariel, R. (2011b). The influence of agenda-based and habitual processes on item selection during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 899–912.
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1180–1191.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks: Sage.
- Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (p. 23–37). Oxford University Press.
- Elliott, B. L., McClure, S. M., & Brewer, G. A. (2020). Individual differences in value-directed remembering. *Cognition*, *201*, 104275.
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*, 121–138.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239.
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47*, 61–87.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.
- Griffin, M. L., Benjamin, A. S., Sahakyan, L., & Stanley, S. E. (2019). A matter of priorities: High working memory enables (slightly) superior value-directed remembering. *Journal of Memory and Language, 108*, 104032.
- Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language, 69*, 429–444.
- Hauck, K. B., Mingo, M. A., & Williams, R. L. (2017). A review of relationships between item sequence and performance on multiple-choice exams. *Scholarship of Teaching and Learning in Psychology, 3*, 58–75.
- Hennessee, J. P., Patterson, T. K., Castel, A. D., & Knowlton, B. J. (2019). Forget me not: Encoding processes in value-directed remembering. *Journal of Memory and Language, 106*, 29–39.
- Hertzog, C., Hultsch, D. F., & Dixon, R. A. (1989). Evidence for the convergent validity of two self-report metamemory questionnaires. *Developmental Psychology, 25*, 687–700.
- Hertzog, C., McGuire, C. L., & Lineweaver, T. T. (1998). Aging, attributions, perceived control, and strategy use in a free recall task. *Aging, Neuropsychology and Cognition, 5*, 85–106.
- Higham, P. A., Zawadzka, K., & Hanczakowski, M. (2016). Internal mapping and its impact on measures of absolute and relative metacognitive accuracy. In J. Dunlosky & S. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 39–61). New York: Oxford University Press.
- Horn, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In E. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 237–278). New York: Plenum Press.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.
- Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence, 77*, 101387.
- Kenny, D. A., Kashy, D., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 233–265). New York: McGraw-Hill.
- Klosner, N. C., & Gellman, E. K. (1973). The effect of item arrangement on classroom test performance: Implications for content validity. *Educational and Psychological Measurement, 33*, 413–418.
- 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.
- Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial introduction with R* (2nd ed.). Burlington: Academic Press.
- Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology, 69*, 241–251.
- Loftus, G. R., & Wickens, T. D. (1970). Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology, 85*, 141–147.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99–128.
- Maqsud, M. (1997). Effects of metacognitive skills and nonverbal ability on academic achievement of high school pupils. *Educational Psychology, 17*, 387–397.
- Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 509–527.
- Mazzoni, G., Cornoldi, C., & Marchitelli, G. (1990). Do memorability ratings affect study-time allocation? *Memory & Cognition, 18*, 196–204.
- McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton: CRC Press.
- McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement in younger and older adults. *Psychology and Aging, 26*, 137–142.
- McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research, 52*, 661–670.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*, 174–179.

- Middlebrooks, C. D., & Castel, A. D. (2017). Self-regulated learning of important information under sequential and simultaneous encoding conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 779–792.
- 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.
- Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. (2014). Type-1 error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed effects model perspective. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *40*, 1287–1306.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), (p. 80) *The psychology of learning and motivation* (pp. 125–173). New York: Academic Press.
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159.
- 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.
- Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. Steve & R. S. McCallum (Eds.), *Handbook of nonverbal assessment* (pp. 223–237). New York: Kluwer.
- Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London: H. K. Lewis.
- Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65–80). New York: Oxford University Press.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*, 550–554.
- Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*, 597–614.
- Rivers, M. L., Dunlosky, J., & Persky, A. M. (2020). Measuring metacognitive knowledge, monitoring, and control in the pharmacy classroom and experiential settings. *American Journal of Pharmaceutical Education*, *84*, 7730.
- Robison, M. K., & Unsworth, N. (2017). Working memory capacity, strategic allocation of study time, and value-directed remembering. *Journal of Memory and Language*, *93*, 231–244.
- Rozencwajg, P. (2003). Metacognitive factors in scientific problem-solving strategies. *European Journal of Psychology of Education*, *18*, 281–294.
- Ryu, E. (2015). The role of centering for interaction of level 1 variables in multilevel structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 617–630.
- Sanna, L. J., & Pusecker, P. A. (1994). Self-efficacy, valence of self-evaluation, and performance. *Personality and Social Psychology Bulletin*, *20*, 82–92.
- Saraç, S., Önder, A., & Karakelle, S. (2014). The relations among general intelligence, metacognition and text learning performance. *Education and Science*, *39*, 40–53.
- Smouse, A. D., & Munz, D. C. (1968). The effects of anxiety and item difficulty sequence on achievement testing scores. *The Journal of Psychology*, *68*, 181–184.
- 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.
- Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1236–1242.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204–221.
- Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *5*, 315–316.
- Staff, R. T., Hogan, M. J., & Whalley, L. J. (2014). Aging trajectories of fluid intelligence in late life: The influence of age, practice and childhood IQ on Raven's progressive matrices. *Intelligence*, *47*, 194–201.
- Sternberg, R. J. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology*, *73*, 1–16.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-paced study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1024–1037.
- Thorndike, E. L., & Lorge, I. (1944). *The Teacher's work book of 30000 words*. New York: Bureau of Publications.

- Tiede, H. L., & Leboe, J. P. (2009). Metamemory judgments and the benefits of repeated study: Improving recall predictions through the activation of appropriate knowledge. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *35*, 822–828.
- Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influence of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 50–61.
- Van der Stel, M., & Veenman, M. V. J. (2014). Metacognitive skills and intellectual ability of young adolescents: A longitudinal study from a developmental perspective. *European Journal of Psychology of Education*, *29*, 117–137.
- Vuorre, M. (2017). Bmlm: Bayesian multilevel mediation. R package version 1.3.4. <https://cran.r-project.org/package=bmlm>.
- Vuorre, M., & Bolger, N. (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, *50*, 2125–2143.
- Weinstein, Y., & Roediger III, H. L. (2010). Retrospective bias in test performance: Providing easy items at the beginning of a test makes students believe they did better on it. *Memory & Cognition*, *38*, 366–376.
- Yu, Y., Jiang, Y., & Li, F. (2020). The effect of value on judgment of learning in tradeoff learning condition: The mediating role of study time. *Metacognition and Learning*, *15*, 435–454.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.