Metacognition and Learning (2020) 15:367–390
https://doi.org/10.1007/s11409-020-09230-y

Prompting retrieval during monitoring and self-regulated learning in older and younger adults

Renée DeCaro[1] · Ayanna K. Thomas[1]

Received: 18 February 2020 / Accepted: 11 May 2020 / Published online: 15 June 2020

Abstract
We evaluated age differences in the relationship between judgments of learning (JOLs) and the choice to restudy a subset of items under two conditions: 1) when a retrieval attempt was explicitly prompted during monitoring; 2) when a retrieval attempt was not explicitly prompted. Young and older adults studied unrelated word pairs. Item-by-item cue-only judgments followed, where participants either attempted to recall the target before providing a JOL or only provided a JOL. After the monitoring phase, participants reported how many total items they wanted to restudy. However, during the selection phase, participants selected half of the presented items to restudy. After restudying selected items, participants received a final cued recall test. Requiring individuals to attempt retrieval increased monitoring reaction times (RT) and decreased JOL magnitude, but did not affect self-regulated learning. For both monitoring groups, individuals were more likely to select items they rated with lower JOLs and items that they spent more time monitoring (i.e., greater RTs). In addition, older adults demonstrated a weaker negative relationship between JOLs and restudy selections, but no difference in the relationship between RTs and restudy selections, compared to young adults. Older adults also indicated wanting to restudy more total items than younger adults. Explicitly prompting retrieval during monitoring did not impact these observed age effects, or interestingly, final test performance. Overall, the results of this study suggest that prompting explicit retrieval prior to monitoring may have little direct effect on self-regulated learning or final test performance.

Keywords Retrieval · Self-regulated learning · Aging

Consider a student who is trying to learn vocabulary words for an upcoming Spanish test. To prepare for the test, she may create cue cards where she attempts to retrieve the Spanish target word she has studied when presented its English counterpart (e.g., *cat–?*). Research would

✉ Renée DeCaro
renee.decaro@tufts.edu

[1] Department of Psychology, Tufts University, Medford, MA 02155, USA

suggest that her *assessment* of her own learning will be better when presented only the cue as compared to when both the cue and target are presented (e.g., *cat–gato*), (see Dunlosky and Nelson 1992). Her assessment of learning will likely inform how she regulates (or, makes decisions about) subsequent learning, such as whether to restudy or to stop studying the paired words. In this example, she *monitors* (i.e., assesses) learning and *controls* (i.e., makes decisions) based on this monitoring (Nelson and Narens 1990).

The purpose of the present work was to characterize the relationship between metacognitive monitoring and control by focusing on the search process engendered by the prompt to monitor learning in a paired-associates paradigm. In the present study, we examined metacognitive control, hereto after referred to as self-regulated learning (SRL), following a monitoring prompt where participants were explicitly required to try to produce the target before providing a judgment (i.e., prejudgment recall, Nelson et al. 2004b), and one where they were only asked to provide a judgment. We hypothesized that the explicit prompt to produce an answer would influence how long participants would spend searching for and/or attempting to retrieve a target, based on previous research (e.g., Son and Metcalfe 2005). We further hypothesized the time differences engendered by the different monitoring prompts would directly relate to SRL. Finally, we compared older and younger adults to assess how SRL could be impacted by the cognitive demands associated with the monitoring manipulation.

The mnemonic basis for monitoring and control

The present work is built on the premise that the cues learners base their monitoring on are important for SRL. In this sense, monitoring is an inferential process informed by access to various cues rather than direct access to the representation of studied information in memory. The cue-utilization framework (Koriat 1997) hypothesizes three classes of inferential cues. Intrinsic cues are those related to item features such as perceptual or semantic characteristics (e.g., font-size, concreteness). Extrinsic cues are those related to processing conditions such as the conditions of encoding and retrieval (e.g., presentation rate during encoding). Mnemonic cues are those related to the learner's individualized experience of a particular item, which can be informed by intrinsic and extrinsic cues. Mnemonic cues include how easily an answer comes to mind in response to a cue (Benjamin et al. 1998; Kelley and Lindsay 1993; Zakay and Tuvia 1998), memory for performance on a previous test (Finn and Metcalfe 2007), and/or the familiarity of the cue used to probe memory (Maki 1999; Metcalfe et al. 1993; Son and Metcalfe 2005; Vernon and Usher 2003). One important experimental distinction is that intrinsic and extrinsic cues can be directly manipulated by the experimenter (e.g., by setting more or less encoding time), while mnemonic cues are largely participant-driven (e.g., whether that particular participant does or does not have an experience of familiarity about a given cue).

In our study, we investigated the amount of time spent attempting to retrieve a target, a mnemonic cue. Focus on mnemonic cues is not without challenges. Direct measurement of mnemonic cues is difficult (Koriat 1997). Further, mnemonic cues are related to a wide range of encoding and retrieval processes (Benjamin et al. 1998; Finn and Metcalfe 2007; Kelley and Lindsay 1993; Maki 1999; Metcalfe et al. 1993; Son and Metcalfe 2005; Vernon and Usher 2003; Zakay and Tuvia 1998). Yet, participants' reliance on these cues can be useful, as the cues are often informative, or diagnostic, of the success of future retrieval. For example, how easily an answer comes to mind in response to a cue is often a useful indicator of future retrieval of the answer (e.g., Dunlosky and Nelson 1994). Alternatively, when there is a disconnect between accessible mnemonic cues and the nature of the later test, reliance on mnemonic cues may be inaccurate as an indicator of future test performance (e.g., Benjamin et al. 1998). Thus, while previous research has established some difficulties inherent in the characterization and measurement of mnemonic cues, research also clearly demonstrates the importance of these cues in predicting performance outcomes. The present research extends this investigation into their effects on SRL.

The present study examines mnemonic cues within the context of delayed, cue-only judgments of learning (JOLs). These judgments are made on an item-by-item basis with only the cue present following a delay from initial study and are prospective (participants predict how likely the target will be remembered in the future). We modeled our method after previous work by Son and Metcalfe (2005). In their study, differences in the mnemonic cue of search time were apparent when JOLs were solicited with an explicit instruction to make a retrieval attempt compared to when JOLs were made without such explicit instruction. More specifically, participants learned a list of paired associates and then made delayed, cue-only JOLs. Participants made judgments either with no special instructions or after being instructed to attempt retrieval prior to JOLs. For individuals instructed to attempt retrieval (covertly or overtly), reaction times (RT) were the longest for items given low JOLs, and RT decreased monotonically as JOLs increased. Put differently, the more time spent attempting retrieval of the target word, the lower the participant estimated the predicted probability of recalling it in the future. This pattern between RT and JOLs suggested that participants used the time searching for a target as a cue to assess future retrievability.

Importantly, without the explicit instruction to retrieve, the pattern of the time spent attempting retrieval of the target changed such that participants' RTs were shorter for items given low JOLs. Without retrieval instructions, RTs across JOLs demonstrated an inverted U-shaped function, with low and high JOLs associated with shorter RTs than JOLs in the mid-range (e.g., 40–60 on a 100-point scale). This pattern suggests that when explicit retrieval prior to JOLs was not instructed, participants may not have engaged in a search for the target for items quickly judged as not well learned (i.e., low JOLs). Rather, in this condition, low JOLs may have been based on pre-retrieval cues, such as the familiarity of the stimulus that was used to probe memory (e.g., *cat*). For the purposes of the present work, these monitoring conditions (JOL Only and Retrieval + JOL) may foster the reliance on different mnemonic cues of cue familiarity (JOL Only) and retrieval fluency (Retrieval + JOL). We hypothesized that reliance on mnemonic cues (i.e., search time) produced by our monitoring manipulation would have downstream consequences on SRL.

Consistent with this hypothesis, recent research has addressed how mnemonic cues may be relied upon for SRL differently based on the type of monitoring assessment. Robey et al. (2017) found that individuals asked to make retrospective confidence judgments (RCJs) were less likely to select correctly retrieved items to be restudied than were individuals who made JOLs. In this study, all participants attempted retrieval of previously studied information (i.e., prejudgment recall), followed by either a JOL, RCJ, or a baseline rating task where participants picked a random number. The authors hypothesized that participants who made JOLs would be less likely to use cues related to the target search and/or retrieval attempt as compared to individuals who made RCJs when deciding whether or not to select an item to be restudied. Indeed, when individuals made RCJs, they were more likely to base their restudy choices on the mnemonic cues related to the retrieval attempt (i.e., retrieval success and time spent searching for the target). The present study complements this research by examining SRL when the monitoring judgment is held constant, but a retrieval attempt is explicitly or not explicitly prompted.

Explicitly prompting retrieval may serve as an important cue for SRL. The presence of an explicit retrieval prompt likely changes the nature of the task and orientation of participants to the task. As one example, a retrieval prompt likely encourages participants to search memory and orients participants to the idea that mastery of all items is unlikely. In support of this hypothesis, research has demonstrated that participants adjust metacognitive predictions after attempting retrieval. Specifically, in one study, participants made global predictions of future performance following study of 40 word pairs. Then, half the participants were given a practice retrieval trial (cued recall of one of the cue–target word pairs), while half the participants received no practice retrieval trial. Participants made a second performance prediction, and a final cued recall test was administered. Individuals in the practice retrieval condition who failed to successfully recall the target revised their global prediction of future performance, resulting in a reduction in the magnitude of confidence (Miller and Geraci 2014). In our study, when retrieval is explicitly prompted during the monitoring phase, participants may be better attuned to mnemonic cues related to the target search (e.g., retrieval success), consequently impacting JOL magnitude, and likely impacting SRL.

In summary, the present study investigated the impact of monitoring-dependent mnemonic cues on SRL. The present work complements previous work which evaluated how reliance on mnemonic cues affects SRL differently when the type of monitoring judgment (e.g., JOLs, RCJs) was manipulated. In our study, we held constant the monitoring judgment (i.e., delayed, cue-only JOLs), and instead, varied the reliance on mnemonic cues by directing participants to make a prejudgment recall attempt prior to monitoring or by directing them only to make the judgment without explicit instructions to attempt retrieval. With the former condition (Retrieval + JOL), the time spent searching for the target and perceived success in retrieval are cues that may influence JOLs. When there is no instruction to attempt retrieval (JOL Only), individuals will likely spend less time searching for the target, and instead may place more reliance on pre-retrieval cues when making JOLs. We predict that the different reliance on the mnemonic cues produced by these monitoring conditions will produce different strategies for SRL (e.g., restudy choices).

In SRL studies, the relationship between monitoring and restudy choices is typically calculated on an individual-by-individual basis and the direction of that relationship illustrates the adopted strategy. The direction may be negative, such that the items rated with the lowest JOLs are those most likely to be selected to be restudied. When individuals select or spend the longest time studying the lowest-rated items, the strategy is called *discrepancy reduction* (Dunlosky and Thiede 1998; Nelson and Leonesio 1988) as individuals are attempting to reduce the discrepancy between the least learned items and the desired level of learning and/or mastery. In the present study, individuals may select items rated with lower JOLs (i.e., demonstrate a negative relationship between JOLs and restudy choices). In addition, individuals could base selections on the mnemonic cue of search time. With regard to the search for the target, individuals in the present study may select items for which they spent the longest time attempting to retrieve (e.g., Robey et al. 2017). For individuals in the Retrieval + JOL group, they would be more likely to select their lowest-rated items (as these would demonstrate the longest RTs). For individuals in the JOL Only group, they would be more likely to select their mid-rated items as these would demonstrate the longest RTs (Son and Metcalfe 2005). In this study, we also evaluated these predictions with respect to individuals who may demonstrate differences in SRL, a healthy older adult population.

Age-related differences in monitoring and SRL

We compared how the two different monitoring conditions influenced restudy choices in younger adults and in older adults. Older adults, like their younger counterparts, must learn and remember new information, sometimes with consequences for health and wellbeing (e.g., take pill A with food, take pill B at night). Effective regulation of learning may act as or enable strategies that compensate for age-related encoding and/or retrieval deficits (e.g., Hertzog and Dunlosky 2011), and could therefore mitigate often-found age-related declines in learning and recollection (e.g., Salthouse 2010). However, what results in effective regulation for older and younger adults may differ as a function of the utility of specific cues for each age group and/or the cognitive demand instantiated by the monitoring process in and of itself.

Research has compared the accuracy of memory monitoring between young and older adults, and in some cases has found older adults to be less accurate (e.g., Souchay and Isingrini 2012), equally as accurate (e.g., Dunlosky and Hertzog 1997), and sometimes even more accurate (e.g., Hines et al. 2009) than younger adults (for review, see Castel et al. 2016). Spared monitoring may occur because older adults rely on many of the same cues accessible during monitoring as younger adults. For example, older adults use intrinsic cues such as stimuli concreteness (e.g., Tullis and Benjamin 2012) or normative difficulty (e.g., Price et al. 2010) to a similar extent as younger adults. Older adults have also shown sensitivity to cue–target relatedness (Hertzog et al. 2002; Thomas et al. 2013). Reliance on these intrinsic cues produced accurate monitoring (i.e., individuals were able to accurately predict what they would and would not later recall). Spared monitoring in older adults has also been demonstrated with mnemonic cues as well (DeCaro and Thomas 2019).

Our earlier work in this area suggests that despite age-equivalence in monitoring accuracy, age-differences in SRL were present (DeCaro and Thomas 2019). In that study, young and older adults attempted retrieval of the target, attempted to recall partial information about the target, and made cue-only JOLs during a monitoring phase. In terms of SRL (i.e., restudy choices), older and younger adults similarly relied on retrieval success in making SRL decisions; however, reliance on accessed partial information differed between these two age groups. For older adults, successfully accessing partial information was associated with an increased likelihood of restudying, whereas for young adults, it was associated with a decrease. In addition, older adults demonstrated a weaker negative relationship between JOLs and restudy choices compared to younger adults. While the results of this previous study underscored how young and older adults may rely differently on mnemonic cues for SRL, all participants made a prejudgment recall attempt. In the present study, we manipulated the presence of a prejudgment recall attempt as doing so may engender different reliance on mnemonic cues for SRL.

Older adults in the Retrieval + JOL group should have access to a greater number of mnemonic cues (i.e., cues produced both from the explicit retrieval attempt and the monitoring judgment) than older adults in the JOL Only group. In the JOL Only group, this same emphasis on retrieval is not present, thus fewer cues are present and/or can be relied upon to guide restudy choices. The additional support and cues produced by the prompt to retrieve may enhance older adults' SRL. This hypothesis is influenced by the understanding that older adults benefit under conditions that prompt explicit monitoring/SRL processes because such prompts serve as external environmental support (cf. Craik et al. 1987). For example, older adults' performance is aided when they're provided learning strategies at encoding (e.g., Froger et al. 2012), asked to explicitly “think aloud” on their mental processes (e.g., Fox and Charness 2010) or trained to explicitly self-test (e.g., Dunlosky et al. 2003). The additional and explicit support encourages thought processes that otherwise would not occur. Thus, the environmental support account would suggest that the increased cues available in the Retrieval + JOL group would enhance older adults’ SRL relative to the JOL Only group.

In contrast, an important factor in determining possible age differences may be the cognitive demands incurred by the explicit retrieval attempt itself. This perspective is informed by a resource-demand account. Exercising metacognitive control after demanding cognitive processes (e.g., encoding, monitoring) may result in too much of a cognitive burden on individuals who, on average, demonstrate a decline in available cognitive resources (Salthouse 2010). Therefore, manipulating the cognitive demands of monitoring (e.g., by prompting retrieval) could influence subsequent SRL in older adults. Requiring older adults to attempt retrieval could represent a task which exceeds available cognitive resources, thereby negatively impacting the relationship between monitoring and SRL for these individuals. If that is the case, then we would expect the relationship between monitoring and SRL to be stronger for older adults in the JOL Only group. For them, terminating search early (or, not engaging in a search at all) would spare valuable cognitive resources. In sum, we investigated how reliance on mnemonic cues (i.e., search time) produced by the addition of a retrieval prompt would have consequences on SRL. We tested whether the retrieval prompt would represent additional support and would therefore enhance older adults’ SRL, or would represent a cognitive burden and would therefore undermine older adults’ SRL.

Method

Design and participants

This experiment was a 2 (age group: young, old) X 2 (monitoring: Retrieval + JOL, JOL Only) between-subjects factorial design. Statistical power analysis using G*Power 3.1 (Faul et al. 2007) determined an estimated total sample size of 75 which was rounded up to 108 to accommodate counterbalancing. Input parameters included an estimate of effect size f ($f = .33$ or $\eta_p^2 = .10$), a standard α error probability (.05), and estimated power ($1-\beta$) of .80. Participants were randomly assigned to the JOL Only or Retrieval + JOL groups based on order of arrival to the lab.

We recruited young adults from Tufts University and older adults from the Greater Boston area. Fifty-four undergraduate students (23 males), ranging in age from 18 to 21 (M = 18.9, SD = 0.8), from Tufts University participated for partial fulfillment of class requirements. Fifty-four community-dwelling older adults (15 males), ranging in age from 58 to 79 years (M = 69.9, SD = 4.8), participated in exchange for nominal compensation. Common in cross-sectional studies (e.g., DeCaro and Thomas 2019), older adults had higher mean years of education, $M = 17.24$, $SD = 2.5$, compared to younger adults, $M = 12.57$, $SD = 0.7$, $t(62.2) = 13.1$, $p < .001$, as well as higher scores on the Shipley vocabulary scale (Zachary 1991), $M = 15.53$, $SD = 2.3$, relative to younger adults, $M = 13.78$, $SD = 1.7$, $t(106) = 3.48$, $p < .01$. Older adults were screened using the National Adult Reading Test (NART; Nelson and Willison 1991), validated for American older adults (Grober et al. 1991), and none were excluded on this basis.

Materials

Seventy-two words were selected from the MRC Psycholinguistics Database (Coltheart 1981) to be displayed as cues and targets. The cues had a neutral valence and the targets had either a positive or negative valence (Warriner et al. 2013). Cue words were randomly assigned targets to form unrelated cue-target pairs. We checked that no target word occurred in the set of possible associates for its cue, and vice versa, using the University of South Florida Free Association Norms (USF FAN; D. L. Nelson et al. 2004a).

Procedure

The procedure was approved by Tufts University Institutional Review Board, and participants provided informed consent prior to participating in the experiment. Some differences in experimental conditions for the age groups were undertaken. Specifically, younger adults were presented with 30 word pairs at a rate of 2.5 s per pair; whereas older adults were presented with 24 word pairs at a rate of 7.5 s per pair. Differences in numbers of words and/or study times are commonly used in cross-sectional experiments (e.g., DeCaro and Thomas 2019) to mitigate differences between age groups in initial learning. If initial learning is different, subsequent differences in self-regulated learning may be a consequence of those initial learning differences, and not metacognition per se (Schwartz and Metcalfe 1994). All other aspects of the procedure were the same between young and older adults. All participants were tested individually and underwent the following phases: initial study, monitoring, restudy goal, selection, restudy, and final test. Each phase is depicted in the diagram of the experimental procedure in Fig. 1. Only the initial study and restudy presentations were displayed at a fixed rate; all remaining phases of the experiment (monitoring, restudy goal, restudy selection, and final test) were self-paced. The experiment was conducted with E-Prime software (Psychology Software Tools, Pittsburgh, PA).

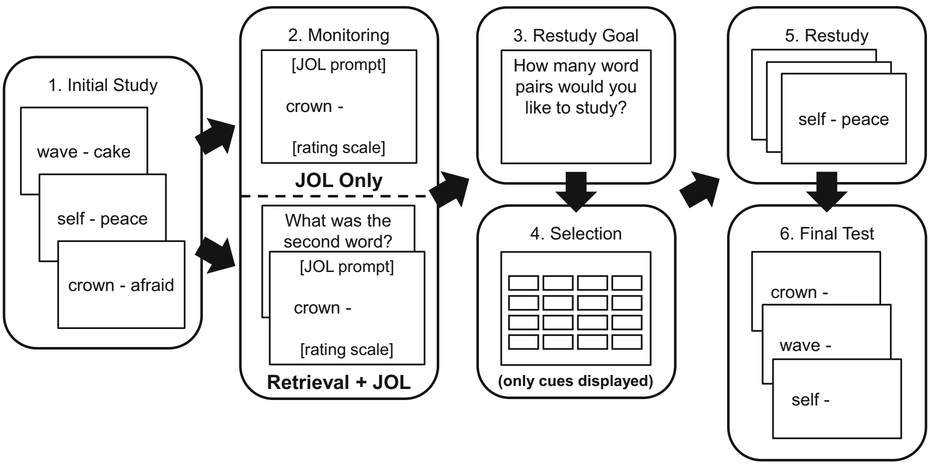

Fig. 1 Experimental flow

Initial study Word pairs were randomly presented to the participants (see Phase 1 in Fig. 1). Words appeared in black Garamond 32-point font on a white background in the center of the computer screen. A blank screen appeared between each word pair presentation for 0.5 s.

Monitoring After studying all word pairs, participants began the monitoring phase they had been randomly assigned to, either the Retrieval + JOL condition or the JOL Only condition (see Phase 2 in Fig. 1). Participants within the Retrieval + JOL condition viewed a screen prior to each JOL. The cue was presented on the screen with the following question: "*What was the second word?*" Participants reported their answer out loud (or said "next" if unable to recall it) while pressing the spacebar, which conveyed them to the next screen where the same cue remained presented with the JOL prompt (see below for description). Thus, participants within the Retrieval + JOL condition made two responses to the same cue, one after the other.

Participants in the JOL Only condition responded only to the JOL prompt. The cue was presented on the screen with the following question: "*How likely do you think it is you can remember the correct answer in the future memory test?*" A scale was displayed at the bottom of the screen from 0 to 100 (in increments of 10), where 0 was '*not likely*' and 100 '*very likely*'. Participants reported the number out loud while pressing the space bar, which conveyed them to the next screen where the next cue word was presented. This continued until participants had made a JOL for every studied cue. Thus, participants within the JOL Only condition made one response to each cue.

During this stage, a microphone recorded the audio and the experimenter transcribed the participant's responses after the study concluded. All participants had unlimited time to make their responses (Son and Metcalfe 2005). Reaction times were measured to the spacebar press (which participants made while speaking their answers aloud). Two reaction times were recorded for participants in the Retrieval + JOL group (the button press to retrieval screen and the JOL screen), and one reaction time was recorded for participants in the JOL Only group (the button press to the JOL screen). At the end of this stage, the experimenter took the keyboard and typed the participant's responses for the remainder of the study.

Restudy goal Following the monitoring phase, individuals indicated a goal for restudying (see Phase 3 in Fig. 1). A screen displayed the question: "*How many word pairs would you like to restudy?*" How many words they could choose from (e.g., "*out of 30"* for younger adults) was displayed in the center of the screen. Participants said aloud how many they wished to restudy and this constituted their global restudy goal. We note, however, that in the selection phase, the number of words participants could select was limited.

Restudy selection Individuals were next presented with an array of cue words in order to make their actual restudy selections (see Phase 4 in Fig. 1). At this stage, two counterbalancing measures were deployed: one which counterbalanced the words that were excluded from the selection set, and one which counterbalanced the presentation order of the cues included in the selection set. In the first case, one-third of the cues were excluded from the selection phase to measure monitoring prediction accuracy in the absence of the choice to restudy/inclusion in the selection phase. The premise for this counterbalancing was that the inclusion (i.e., re-presentation) of the cue words may impact monitoring accuracy. Excluded cues were counterbalanced across three subsets, so that each cue word set was excluded from the selection phase an equal amount of times.[1]

On the selection screen, the included cues were presented in an array comprised of 4 rows. For the second counterbalancing measure, three array orders were presented across each of an additional three counterbalancing orders, so that words appeared an equal number of times in three sections of the computer screen. The premise for this counterbalancing was to mitigate any order effects on restudy choices (e.g., Price 2017). Participants selected half of the items to restudy from the presented cues (8 out of a possible 16 for older adults and 10 out of a possible 20 for younger adults). Each cue was numbered, and participants reported aloud the corresponding number of the cue word they wished to select. The cues disappeared from the array when they were selected, and the number of items that remained to be selected was updated on a display on the screen. Once the number of items remaining to be selected reached '0', the next phase began.

Restudy After all selections were made, participants began the restudy phase. During the restudy phase, the selected cue-target word pairs were re-presented similarly to the initial study phase, but in a new random order (see Phase 5 in Fig. 1).

Final test A cued recall test followed, where cues were tested in a new random order (see Phase 6 in Fig. 1). Participants were asked to recall the appropriate target for the cue displayed in the center of the screen. The experimenter typed their response and hit ENTER, after which the next cue to be tested appeared.

Results

First, we tested whether the expected differences in reaction times and JOL magnitude were present between our monitoring conditions and age groups. We tested monitoring accuracy for items excluded from the selection phase, and for comparison, we reported accuracy for items included in the selection phase. Of primary interest were the SRL outcomes. We evaluated how many word pairs individuals wished to restudy at the conclusion of the monitoring phase, prior to selecting items to be restudied. We tested whether the choices made during the selection phase differed as a consequence of monitoring and age groups, via intraindividual correlations (e.g., Gamma) as well as a mixed-effects model. Finally, we compared cued recall performance between participant groups across item types (e.g., excluded, selected, not selected).

[1] We compared average JOLs for the excluded sets to average JOLs for the included set of cues via paired-samples t-test and found no significant difference between the two, $M_{\text{Excluded}} = .38$ $SD = .22$, $M_{\text{Included}} = .38$ $SD = .22$, $t(107) = 0.24$, $p = .81$. The excluded sets were not rated as more or less difficult to remember (as indicated by average JOLs) than the sets of cues included in the selection phase.

Monitoring

Reaction times We used a generalized multilevel model to predict reaction times from JOLs. Reaction times larger than 2.5 standard deviations from an individual's mean RT (for each button press) were excluded, representing 133 cases from the data (4.6% of possible cases). As in Son and Metcalfe (2005), JOLs were assumed to add a fixed judgment reaction time for the Retrieval + JOL monitoring group, so we examined their cumulative RTs (i.e., reaction times to the retrieval and JOL button press were summed). We did not transform RT but instead modeled raw RT in milliseconds, using the Gamma distribution and the identity link function (Lo and Andrews 2015), so that regression coefficients could be interpreted as raw RT.

At the trial-level were two predictors, JOL and JOL squared. JOL was transformed into units of ten (e.g., a rating of 90 became 9), and was lower-limit centered (which required no transformation). Thus, the intercept represented a JOL of zero, the lowest possible rating (0%). The quadratic term (JOL squared) was the square of the transformed JOL. At the subject-level were two predictors, monitoring and age group, which were dummy-coded such that young adults and the JOL Only group were the referents. Random effects were included in the model for the participant and item intercepts and for the slope of JOL.

Regarding JOLs, we expected that JOLs would demonstrate a nonlinear relationship with reaction times as indicated by a significant effect for the JOL squared predictor. A negative coefficient for this predictor would indicate an inverted U-shaped relationship existed between JOLs and reaction times (Son and Metcalfe 2005). Asking individuals to make an explicit retrieval attempt would increase reaction times and would interact with the relationship between JOLs and reaction times. In other words, requiring retrieval would have a different effect on RT depending on values of JOL. Finally, older adults would demonstrate increased reaction times compared to young adults and would be differentially impacted by the explicit retrieval requirement, as indicated by a significant three-way interaction between JOLs, monitoring group, and age group. These hypotheses were confirmed and the relevant significant effects discussed below. All fixed and random effects in our model are reported in Table 1.

Table 1 Summary of generalized linear mixed model predicting reaction time

Fixed effects	β (SE)
Intercept	3256.54 (396.4)***
JOL_{trial}	371.35 (84.4)***
JOL_{trial}^2	- 39.93 (6.9)***
Monitoring (Retrieval + JOL)	3918.16 (568.7)***
Age (older)	2424.33 (596.0)***
JOL_{trial} Monitoring (Retrieval + JOL)	149.90 (159.7)
JOL^2_{trial} Monitoring (Retrieval + JOL)	- 45.12 (15.1)**
JOL_{trial} Age (older)	507.22 (174.8)**
JOL^2_{trial} Age (older)	- 67.87 (15.7)***
Monitoring (Retrieval + JOL) Age (older)	947.20 (854.1)***
JOL_{trial} Monitoring (Retrieval + JOL) Age (older)	- 855.98 (296.1)**
JOL^2_{trial} Monitoring (Retrieval + JOL) Age (older)	89.46 (28.43)**
Random effects	σ (*SD*)
Item Intercept	2.32 X 10^4 (152.2)
Subject Intercept	4.47 X 10^4 (211.5)
Subject JOL Slope	3.73 X 10^6 (1930.0)

~$p < .10$ *$p < .05$. **$p < .01$. ***$p < .001$

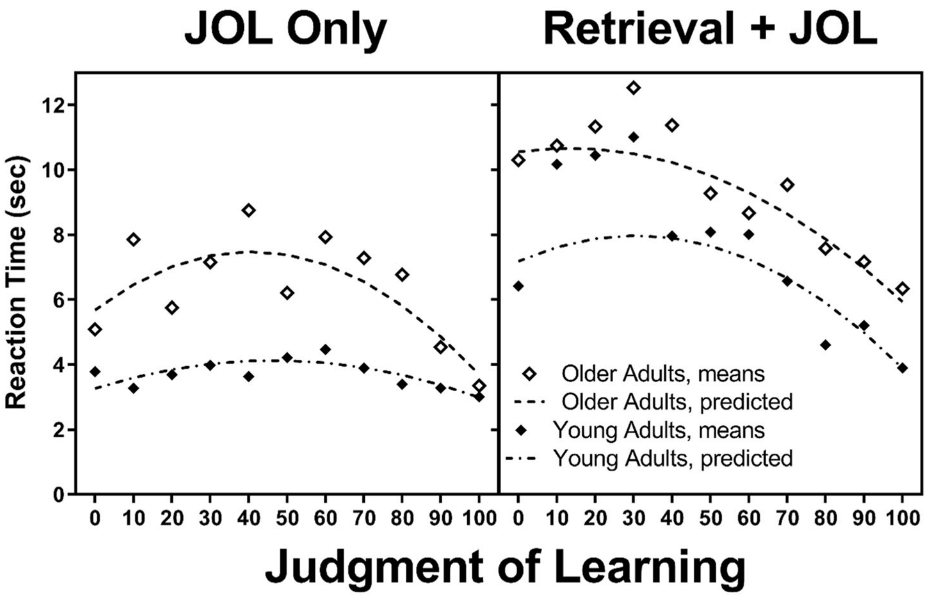

Fig. 2 Reaction times across judgments of learning (JOL) for the JOL Only and Retrieval + JOL monitoring groups plotted separately for young and older adults

To ease with the interpretation of the coefficients, predicted reactions times were plotted alongside mean reaction times in Fig. 2. Most importantly, there was a significant nonlinear relationship between JOLs and reaction time for young adults in the JOL Only group ($\beta = -39.93$). This JOL squared predictor significantly interacted with each of the subject-level variables, signifying that the nonlinear relationship between JOL and reaction times were different for young adults in the JOL Only monitoring group as compared to the young adults in the Retrieval + JOL monitoring group ($\beta = -45.12$), the older adults in the JOL Only monitoring group ($\beta = -67.87$), and the older adults in the Retrieval + JOL monitoring group ($\beta = 89.46$). As can be seen in Fig. 2., the predicted nonlinear relationship between JOLs and

Table 2 Summary statistics for items included and excluded from the selection phase by age group

	Selection phase					
	Included				Excluded	
	Selected		Not Selected			
Mean JOL Magnitude	*M* (*SD*)		*M* (*SD*)		*M* (*SD*)	
Young Adults	.25 (*0.21*)		.56 (*0.30*)		.40 (*0.22*)	
Older Adults	.27 (*0.23*)		.43 (*0.32*)		.36 (*0.22*)	
Mean Final Test Performance	*M* (*SD*)		*M* (*SD*)		*M* (*SD*)	
Young Adults	.72 (*0.22*)		.57 (*0.34*)		.34 (*0.24*)	
Older Adults	.69 (*0.23*)		.40 (*0.32*)		.27 (*0.22*)	
Monitoring Accuracy (γ)	*M* (*SD*)	N	*M* (*SD*)	N	*M* (*SD*)	N
Young Adults	0.25 (*.67*)	37	0.60 (*.67*)	40	0.85 (*.35*)	47
Older Adults	0.18 (*.72*)	42	0.84 (*.41*)	37	0.81 (*.30*)	41

JOL , judgment of learning

RTs was an inverted U, with the longest reaction times predicted for the mid-value JOLs and shorter reaction times predicted for the low- and high-value JOLs. Further, the degree (i.e., shape) of the nonlinear inverted U pattern was different for the participant groups. Thus, the present analysis confirmed that monitoring conditions influenced reaction times and did so differently for young and older adult participants.

Monitoring magnitude Average JOL was subjected to a 2 (age group: young, old) X 2 (monitoring: Retrieval + JOL, JOL Only) between-subjects ANOVA. Monitoring group was significant, $F(1,104) = 20.38$, $p < .001$, $\eta_p^2 = .16$. Individuals in the Retrieval + JOL group rated words with lower average JOLs than individuals in the JOL Only group ($M_{\text{Retrieval + JOL}} =$ 29.42, $SD = 18.98$; $M_{\text{JOL Only}} = 46.58$, $SD = 20.45$). Neither the effect of age, $F(1,104) = 1.76$, $p = .19$, or the monitoring group by age interaction, $F(1,104) < 1$, $p = .98$, were significant (see Table 2 for marginal means). When individuals were asked to retrieve the answer prior to providing a JOL, they were less confident in whether they would be able to remember the item in the future. This was true for both young and older adults. To further illustrate, we plotted the distribution of JOLs in Fig. 3. As can be seen in Fig. 3., both young and older adults frequently used the lowest ratings on the scale (i.e., 0%) when first asked to attempt retrieval.

Monitoring accuracy We evaluated the relationship between JOLs and final test performance to compare group differences in monitoring accuracy. We computed Gamma (γ) correlations for individuals using JOLs and whether the item was recalled at final test (i.e., recalled, not recalled), for only those items that had been excluded from the selection phase. Note that γ could not be computed for 7 younger adults and 13 older adults, either because final test performance was a constant for these items or both final test performance and JOL were a constant for the items. A 2 (age group: young, old) X 2 (monitoring: Retrieval + JOL, JOL Only) between-subjects ANOVA was performed on mean gamma correlation. No significant effects were found, Fs < 1. Average gamma correlations were strong and positive, indicating that participants' JOLs accurately predicted final test performance (see Table 2 for descriptive statistics). For reference, we also computed γ for items included in the selection phase. These averages are also reported in Table 2.

In the previous analyses, we found that individuals in the Retrieval + JOL groups took longer to search for the targets than individuals in the JOL Only groups. Further, these individuals reported lower average JOLs. While the explicit prompt to attempt retrieval

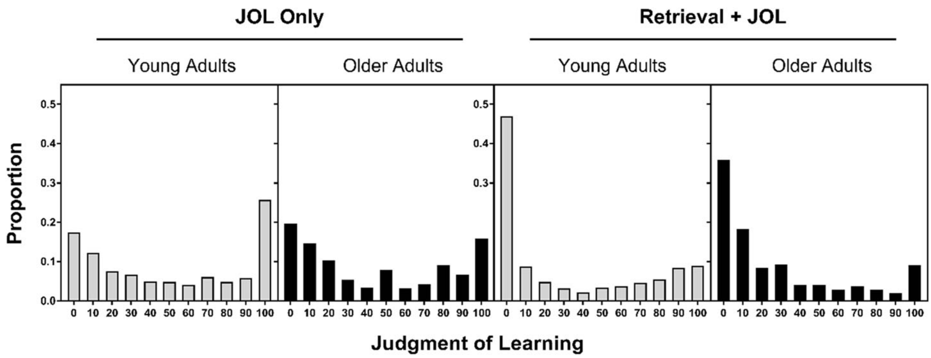

Fig. 3 Frequency distribution of judgments of learning, plotted separately by groups. Frequency is relative (proportional) as young and older adults received a different number of trials

decreased JOL magnitude to a similar extent for older and younger adults, older adults' RTs were differentially affected by the manipulation. Older adults' search times in the JOL Only group were longer than younger adults in the JOL Only group, and for older adults in the Retrieval + JOL group, search times were even longer. The presence of significant interactions between older adult groups and trial JOL demonstrated that the manipulation did not simply add a fixed amount of time at each value of JOL, but rather that increased search times for older adults were observed differently across JOL ratings. Finally, all groups were similarly accurate at predicting final test performance from JOLs.

Self-regulated learning

Next, we tested whether the differences in the monitoring phase (i.e., JOLs, RTs) produced different strategies during the restudy goal and selection phases (Phases 3 and 4 in Fig. 1). We measured learners' overall restudy goal (as indicated by the number they wished to restudy) as well as what learners ultimately selected. Learner's selections were evaluated via intra-individual correlations (i.e., the relationship between one predictor and whether or not a word was selected to be restudied) and via generalized multilevel models to test multiple predictors simultaneously.

Restudy goal We examined differences in the number of word pairs individuals indicated they wished to restudy. A 2 (age: young, older) X 2 (monitoring: Retrieval + JOL, JOL Only) ANOVA was performed on the proportion of items individuals indicated they wished to restudy. Only age was significant, $F(1,104) = 4.85$, $p < .05$, $\eta_p^2 = .05$. On average, young adults wanted to restudy fewer words than older adults ($M_{young} = 0.47$, $SD = 0.24$; $M_{older} = 0.58$, $SD = 0.31$). No other effects were significant, $ps > .19$. Marginal means are reported in Table 3.

Restudy selections as intra-individual correlations In this and the following analysis, items excluded from the selection phase were not included, as these could never have been selected by the participant. To assess whether monitoring judgments were associated with the decision to restudy items, Gamma correlations (γ) were computed between JOLs and restudy choice (i.e., selected, not selected). Note that γ could not be computed for 1 older adult as JOLs did not vary. A 2 (age: young, older) X 2 (monitoring: Retrieval + JOL, JOL-only) ANOVA was performed on average γ. Age was significant, $F(1,103) = 5.77$, $p < .05$, $\eta_p^2 = .05$. Neither the

Table 3 Restudy selection indices and within-person correlations by age and monitoring group

	Young Adults		Older Adults	
	JOL Only	Retrieval + JOL	JOL Only	Retrieval + JOL
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Global restudy goal	.44 (*0.21*)	.50 (*0.27*)	.62 (*0.31*)	.55 (*0.31*)
JOL and Restudy (γ)	- .53 (*0.51*)	- .67 (*0.40*)	- .33 (*0.67*)	- .35 (*0.59*)
RT and Restudy (γ)	.18 (*0.36*)	.23 (*0.46*)	.33 (*0.38*)	.23 (*0.40*)

JOL, judgment of learning; *RT*, reaction time effect of monitoring group, $F(1,103) = 0.53$, $p = .47$, nor the monitoring group by age interaction, $F(1,103) = 0.27$, $p = .60$, were significant.

Each group demonstrated an average γ that was negative (marginal means are reported in Table 3). In effect, individuals selected items to be restudied for which they had assigned lower relative values of JOLs. Young adults demonstrated a stronger negative relationship than older adults ($M_{\text{young}} = -0.60$, $SD = 0.47$; $M_{\text{older}} = -0.34$, $SD = 0.63$). Importantly, both older and younger adults demonstrated a relationship between JOLs and restudy choice that was greater than chance. Specifically, each age group average was subjected to a one-sample t-test. Average γ for young adults, $t(53) = 9.63$, $p < .001$, and older adults, $t(52) = 4.00$, $p < .001$, was significantly different than zero.

To assess whether reaction times were associated with the decision to restudy items, Gamma correlations (γ) were computed between RT and restudy choice (i.e., selected, not selected). Note that RTs greater than 2.5 standard deviations from an individual's mean were excluded from the computations. A 2 (age: young, older) X 2 (monitoring: Retrieval + JOL, JOL Only) ANOVA was performed on average γ. No effects were significant. There were no significant differences in the relationship between reaction time and restudy choice between the age groups, $F(1,104) = 0.86$, $p = .36$, monitoring groups, $F(1,104) = 0.12$, $p = .73$, as well as no significant interaction, $F(1,104) = 0.89$, $p = .35$. Each group demonstrated an average γ that was positive, indicating individuals selected items for which they had spent a greater amount of time searching for the target. Average γ for the participants in this study was significantly different than zero, $M_{\gamma} = 0.24$, $SD = 0.40$, $t(107) = 6.37$, $p < .001$. Marginal means are reported in Table 3.

Restudy selections and mixed-effects models We conducted generalized linear mixed models (GLMMs) to test multiple predictors of restudy choice. The purpose was twofold. First, GLMMs would enable us to test the relationships observed in our intra-individual correlation analysis (i.e., the presence of age group differences but not monitoring group differences) with JOL as an *absolute* measure. Second, GLMMs would enable us to test whether reaction times predicted the likelihood of restudying when added to a model containing our initial predictors (e.g., JOL). The second aim was accomplished by comparing an initial model to a second model with reaction time predictors added. Model 1 (our initial model) which included JOL, age group, monitoring group and their interactions was fit to the data.[2] Model 2 which included reaction time as a predictor as well as interactions with subject-level predictors (e.g., monitoring group) was fit to the data. Note that reaction time (in seconds) was centered within each subject such that each subject's mean reaction time was subtracted from their trial-level reaction time (e.g., in the model, a reaction time of '0' corresponded to the subject's mean reaction time). Regarding model comparison, there was a significant reduction in model deviance when the reaction time fixed effects were added, as indicated by a likelihood ratio test, $\chi^2(4) = 51.96$, $p < .001$. Fixed and random effects for both models are reported in Table 4. As reaction time significantly predicted restudy choices, only Model 2 is interpreted below.

Trial JOL significantly predicted restudy choice. For two items that differ in JOL ratings by 10% (e.g., one unit), the odds of selection for the higher JOL are 0.76 that of selecting an item rated with the lower JOL. Thus, as with our intra-individual correlations, we observed a

[2] Data coding was the same for all GLMMs. JOLs were transformed into units of 10 (e.g., 10% was transformed into '1') and were lower-limited centered. Young adults and the JOL Only group were the referents.

Table 4 Summary of generalized linear mixed models predicting restudy choice

	Model 1	Model 2
Fixed Effects	β (SE)	β (SE)
Intercept	- 1.32 (0.25)***	- 1.32 (0.24)***
Reaction time (RT)		- 0.24 (0.07)***
JOL_{trial}	- 0.28 (0.05)***	- 0.28 (0.05)***
Monitoring (Retrieval + JOL)	- 0.31 (0.33)	- 0.37 (0.32)
Age group	- 0.78 (0.35)*	- 0.76 (0.34)
JOL_{trial} Monitoring group	- 0.01 (0.07)	- 0.02 (0.07)
JOL_{trial} Age group	- 0.19 (0.07)**	- 0.20 (0.07)**
Monitoring group Age group	- 0.36 (0.47)	- 0.41 (0.45)
JOL_{trial} Monitoring group Age group	- 0.11 (0.10)	- 0.12 (0.10)
Age group RT		- 0.05 (0.09)
Monitoring group RT		- 0.14 (0.08)~
Age group Monitoring group RT		- 0.05 (0.10)
Random Effects	σ (*SD*)	σ (*SD*)
Item Intercept	0.08 (0.28)	0.07 (0.26)
Subject Intercept	0.79 (0.89)	0.67 (0.82)
Subject JOL Slope	0.04 (0.21)	0.04 (0.20)
–2LL	2276.4	2224.4

~$p < .10$ *$p < .05$. **$p < .01$. ***$p < .001$

negative relationship: as JOL increases, the likelihood of selecting an item to be restudied is reduced. We observed a significant interaction between age group and trial JOL. This significant interaction indicates that the relationship between trial JOL and restudy choice was different for older adults compared to younger adults (the referent). For older adults, as JOL increased by 10%, the odds of selecting the higher JOL were 0.92 that of selecting an item with the lower JOL. These odds are close to one, indicating a weak relationship for older adults between JOL and restudy choices (in line with findings from the intra-individual correlations). No other effects of trial JOL were significant.

Trial reaction time significantly predicted restudy choice. For two items that differed in reaction time by 1 s (e.g., one unit), the odds ratio was 1.27, holding constant trial JOL. Thus, for trial reaction time, there was a positive relationship: as reaction time increases, the likelihood of selecting an item to be restudied is increased. No other effects of trial reaction time were significant. Specifically, the increased odds of selection associated with increased reaction times did not differ for young adults in the Retrieval + JOL monitoring group ($\beta = -0.14$), the older adults in the JOL Only monitoring group ($\beta = -0.05$), and the older adults in the Retrieval + JOL monitoring group ($\beta = 0.05$). These results demonstrated that reaction time was related to restudy selection while accounting for the effects of trial JOL.

Final test

We examined final test performance conditional on whether items were studied (e.g., Price and Murray 2012). We examined differences in final test performance across the different item types: items restudied after being selected during the selection phase (i.e., selected words), items not restudied but presented during the selection phase (i.e., not selected words), and items not restudied/not presented during the selection phase (i.e., excluded words). A 2 (age: young, older) X 2 (monitoring: Retrieval + JOL, JOL Only) X 3 (item type: selected, not selected, excluded) mixed-effects ANOVA was performed on the proportion of items individuals recalled on the final test, with item type as the within-subjects factor. Item type was significant, $F(2, 208) = 123.47$, $p < .001$, $\eta_p^2 = .54$. Participants recalled a different proportion of items on the final test between selected items ($M_{Selected} = 0.70$ $SD = 0.23$), not selected items ($M_{Not\ Selected} = 0.48$ $SD = 0.35$), and excluded items ($M_{Excluded} = 0.31$ $SD = 0.23$). Age group was significant, $F(1, 104) = 4.58$, $p < .05$, $\eta_p^2 = .04$, as was the interaction between age group and item type, $F(2, 208) = 4.23$, $p < .05$, $\eta_p^2 = .04$. No other effect was significant, Fs < 1.35, ps > .26. Follow-up post hoc tests, which were Bonferroni corrected, are reported below.

Younger adults remembered a greater proportion of words they selected/restudied compared to words included in the selection phase but not selected/not restudied, $M_{Diff} = 0.15$ ($SE = 0.04$), $t(53) = 3.75$, $p < .001$, and compared to words that were excluded from the selection phase, $M_{Diff} = 0.37$ ($SE = 0.03$), $t(53) = 12.73$, $p < .001$. Younger adults also remembered a greater proportion of not selected words as compared to excluded words, $M_{Diff} = 0.22$ ($SE = 0.03$), $t(53) = 6.79$, $p < .001$. For older adults, this pattern was the same. Older adults remembered a greater proportion of words they restudied compared to those they did not select to be restudied, $M_{Diff} = 0.29$ ($SE = 0.04$), $t(53) = 7.88$, $p < .001$, and compared to those excluded from the selection phase, $M_{Diff} = 0.42$ ($SE = 0.04$), $t(53) = 11.51$, $p < .001$. They also remembered a greater proportion of not selected words compared to excluded words, $M_{Diff} = 0.13$ ($SE = 0.04$), $t(53) = 3.35$, $p < .01$. Unsurprisingly, restudying cue-target pairs conferred a performance benefit compared to items that were not restudied (i.e., both those not selected during and those excluded from the selection phase). Interestingly, for both age groups, performance was greater when items were included (but not selected) in the selection phase compared to when they were excluded from the selection phase. Mere inclusion in a cue-only form in the selection phase boosted performance.

We compared young and older adult performance across the three selection classes. There were no significant differences in performance between young and older adults for restudied items, $M_{Diff} = 0.03$ $SE = 0.04$, $t(106) = 0.69$, $p = .49$, or for items excluded from the selection phase, $M_{Diff} = 0.07$ $SE = 0.04$, $t(106) = 1.70$, $p = .092$. For items that were included in the selection phase but were not selected (and therefore not restudied), there was a significant difference in performance between young and older adults, $M_{Diff} = 0.17$ $SE = 0.07$, 95% CI_{Diff} [0.04, 0.30], $t(106) = 2.64$, $p < .01$. Both young and older adults demonstrated comparable final test performance for restudied items and items excluded from the selection phase. However, older adults performed worse on items not selected during the selection phase compared to younger adults. Marginal means are reported in Table 2.

General discussion

The amount of time spent attempting to retrieve a target could inform later decisions to restudy that target. In the present study, we either explicitly prompted retrieval of the target prior to JOLs or simply asked for JOLs without the explicit prompt to retrieve the target. The manipulation was built on the premise that the two methods of eliciting delayed, cue-only JOLs engender differences in the amount of time spent searching for the target (Son and Metcalfe 2005). The differences in reliance on mnemonic cues for monitoring judgments (Koriat 1997) may, in turn, produce differences in SRL. We found that when we explicitly prompted a retrieval attempt, participants took longer to make JOL assessments and average JOLs were lower than when participants were not explicitly prompted to retrieve. In terms of the consequences for SRL, our findings suggest that such explicit prompting does not, however, influence the choices one makes about restudying. Rather, in both conditions, individuals were more likely to select items for which the participant rated with the lowest values of JOLs and spent the greatest amount of time attempting to retrieve. Finally, when older adults were compared to younger adults, we found a weaker relationship between monitoring judgments and restudy choices, a similar relationship between reaction times and restudy choices, and age differences in restudy goals. With regard to the SRL age findings, none were affected by the monitoring manipulation.

The mnemonic basis for monitoring

In this experiment, we observed differences in reaction times when individuals were first asked to make an explicit retrieval attempt prior to monitoring (Retrieval + JOL) compared to when asked to make a monitoring judgment alone (JOL Only). Individuals in the Retrieval + JOL group took longer to make individual judgments than did individuals in the JOL Only group. Despite this longer retrieval/monitoring time, we note here that we did not entirely replicate findings from Son and Metcalfe (2005). Specifically, individuals in that study's Retrieval + JOL group demonstrated reaction times that were linearly-related to JOLs (i.e., a nonlinear relationship was *not* present between JOLs and RTs). In our Retrieval + JOL group, JOLs *did* show a nonlinear relationship with RTs. We suggest that this nonlinear relationship may be an indication of a termination of search for items judged as not known, consistent with previous research indicating individuals' ability to make fast "don't know" judgments (Kolers and Palef 1976). It is unclear why Son and Metcalfe did not find this pattern. We can only speculate that the nature of their stimuli was such that participants were able to acquire some trace information during the initial learning phase for all pairs presented.

Although the present pattern of RTs differs from those presented in one prior study, we argue that this difference does not impact our overarching conclusions. As the main purpose of our monitoring prompt manipulation was to create the conditions for different reliance on mnemonic cues and how those differences might manifest in changes in SRL, the pattern of RTs within a condition (e.g., within Retrieval + JOL) is less relevant than the pattern between conditions (e.g., between Retrieval + JOL and JOL only). That said, future research, with the explicit goal of teasing apart the mnemonic cues produced by these conditions, could help disambiguate the findings of our study and Son and Metcalfe (2005).

Important for the present study, the nonmonotonic relationships revealed a complex monitoring process wherein the time spent monitoring may have depended at times on pre-retrieval cues, such as the familiarity of the stimulus/cue used to probe memory, as compared to retrieval fluency alone (Son and Metcalfe 2005). This pattern is also consistent with the notion that people base monitoring on inferences from a variety of cues, in addition to the amount of effort expended in attempting to retrieve the target (e.g., Koriat et al. 2008). As measured by reaction time data, time spent attempting to retrieve the target differed in the present study depending on monitoring condition, and older adults were differentially affected by this manipulation.

The reliance on different monitoring cues also influenced the magnitude of JOLs. When individuals were first asked to retrieve the target, they demonstrated lower average JOLs than when they were not asked to retrieve the target. Differences in monitoring magnitude may have been driven by perceived retrieval success/failure because of its salience in the JOL + Retrieval condition (cf., Schwartz et al. 2016). This argument is consistent with those made by

Miller and Geraci (2014). Miller and Geraci found that a single, unsuccessful retrieval practice trial reduced overconfidence. Similarly, research has demonstrated that taking overt tests leads to reduced overconfidence across multi-trial learning (Koriat et al. 2002). The experience taking an overt test affects monitoring that takes place after the test has concluded, leading both young and older adults to adjust their predictions of future memory performance (Tauber and Rhodes 2012; Rast and Zimprich 2009). In our experiment, we extended these findings beyond global monitoring judgments and multi-trial learning to conditions where individuals instructed to explicitly attempt retrieval prior to each item-by-item monitoring judgment demonstrated lower JOLs than those not instructed to attempt retrieval. Lower monitoring magnitude was observed for both young and older adults, suggesting that both groups relied on the mnemonic cues produced by the retrieval prompt to a similar extent.

Despite differences in RT and JOL magnitude produced by the monitoring manipulation, individuals in all groups successfully predicted future test performance. In the present context, this equivalence in monitoring accuracy was important to establish so that any subsequent differences in SRL could be attributable to reliance on mnemonic cues, and not monitoring accuracy per se (see, for example, Kimball et al. 2012). To ensure that our monitoring conditions did not produce differences in monitoring accuracy, we tested predictive accuracy for items excluded from the selection phase. Consistent with previous research (Nelson et al. 2004b), individuals in the different monitoring conditions were equally accurate at predicting final test performance. Of note, individuals were more accurate at predicting final test performance for these items than for items which had been included in the selection phase but not selected, $t(68) = 2.09$, $p < .05$. This finding, while not central to our main hypotheses, is relevant for the SRL literature more broadly. In research examining item selection decisions, monitoring accuracy is typically calculated only for those items not selected for restudy (e.g., Tullis and Benjamin 2012). However, in the present study, we demonstrated an accuracy difference between items not selected for restudy and those excluded entirely from the selection phase. That predictive accuracy is affected by inclusion in the selection phase should be considered in the context of future SRL research. Finally, no age differences in monitoring accuracy were found, in line with previous research on delayed, cue-only JOLs (Hertzog and Dunlosky 2011).

Age differences in SRL

In the present study, regardless of explicitly prompting a retrieval attempt, age differences in SRL were found. Age-related SRL differences took several forms. First, the relationship between JOLs and restudy selections was weaker for older adults as compared to younger adults, consistent with previous research (DeCaro and Thomas 2019). This age difference in the monitoring-SRL relationship may indicate that aging may accompany a decoupling of monitoring and SRL; however, we also found age-invariance in the relationship between monitoring RT and SRL, suggesting that older adults and younger adults may use RT similarly when making restudy decisions. Second, older adults requested to restudy more word pairs as compared to young adults. The desire to restudy a different number of word pairs, unaffected by the monitoring manipulation, may indicate that older adults rely on a different type of cues, such as personal beliefs about their own memories, than younger adults. These age effects (in monitoring, RT, and goals) will be discussed in turn.

In our study, all participants demonstrated a negative relationship between JOLs and restudy choices, a pattern in line with the discrepancy reduction model of self-regulated learning (Dunlosky and Thiede 1998). Importantly, however, the relationship was less pronounced for older adults as compared to younger adults, in line with previous research suggesting that older adults are less likely to select items for restudy based solely on monitoring decisions (e.g., DeCaro and Thomas 2019). Importantly, the age-related differences found in the relationship between monitoring and control were unaffected by monitoring group. That is, requiring older adults to attempt retrieval of the target prior to providing JOLs did not influence the relationship between monitoring and control.

The examination between reaction times and restudy choices found a predicted pattern where the more time spent monitoring increased the likelihood of selecting an item to be restudied. Importantly, this pattern was statistically identical for older and younger adults suggesting that younger and older adults were similarly influenced by cues captured by search time (e.g., retrieval fluency). This similar reliance on the mnemonic cue of monitoring times stands in contrast to previous work (e.g., DeCaro and Thomas 2019). In DeCaro and Thomas (2019), several mnemonic cues were tested in the context of a paired-associates SRL paradigm. Importantly, in that study, successfully accessing partial information (i.e., the valence of the unrecalled target) predicted restudy for older but not younger adults. Retrieval success reduced the likelihood of selecting an item to be restudied similarly for young and older adults. The comparison between DeCaro and Thomas (2019) and the present study demonstrates the need to investigate multiple cues simultaneously (see also Hines et al. 2015).

Finally, prior to the selection phase we asked older and younger participants to indicate how many items they wanted to restudy. Older adults wanted to restudy more word pairs than younger adults. In fact, more older adults than younger adults indicated a desire to restudy the *entire* list. For older adults, nearly one third (27.8%) of the current sample indicated the desire to restudy the entire set. For younger adults, this number was just over 10% (11.1%). In the younger age group, the most frequently requested number of items for restudy was 10 (or, one third of the available items), preferred by approximately 40% of younger people in this study. Importantly, no differences were observed between monitoring groups conditions. That is, despite feeling less confident in the future recallability of targets (as indicated by lower magnitude JOLs), individuals in the Retrieval + JOL group did not, subsequently, wish to restudy more words. The presence of age group differences in the global restudy goal suggests reliance on information-based cues rather than mnemonic cues derived through the specific task experience. Information-based cues are those that reflect a person's beliefs about their own memories and other cognitive processes (Koriat et al. 2008). Here, older adults may have lower confidence in general in their memory ability, and subsequently, adjust their restudy goal to compensate accordingly (West et al. 2003). Such an interpretation would be in line with findings which have indicated older adults expect to learn less information, even before beginning the learning task (Price et al. 2010).

Implications for retrieval and SRL

Requiring individuals to attempt retrieval did not influence the number of words they indicated wishing to restudy, the relationship between JOLs and restudy choices, or the relationship between search time and restudy choices. These results suggest that the explicit retrieval prompt may result only in changes to monitoring magnitude (e.g., under-confidence). These results may also suggest that in the context of a paired-associates list learning paradigm, restudy choices may be based on cues other than those available during the monitoring phase. Specifically, participants may engage in a later covert retrieval attempt during the selection phase that may be more directly related to their restudy choices than cues generated during monitoring. Evidence for engagement in a later covert retrieval attempt during the selection phase is, in part, supported by our findings regarding final test performance. Namely, performance on the final test was superior for items included during the selection phase but ultimately not selected as compared to those that were excluded from the selection phase. It is possible that when presented the subset of cues during the selection phase, participants engaged in attempted retrieval of the target words and used this as a basis to inform their restudy selections. At that time, participants likely chose not to restudy those items they were able to successfully retrieve. Successful retrieval likely boosted final test performance relative to those items that had been excluded (and, therefore, retrieval not attempted), a finding consistent with the retrieval practice literature (e.g., Roediger and Butler 2011). Importantly, our results are also consistent with findings suggesting older adults benefit more from restudying paired-associates compared to retrieval practice, unless provided corrective feedback (Tse et al. 2010). In our study, the later, covert retrieval attempt during the selection phase was at least more temporally salient than cues present during the monitoring phase, with consequences for both restudy choices and final test performance.

Although the primary purpose of this research was to better understand the cues that may drive restudy choice and how they may differ between older and younger adults, these results have implications for the retrieval-based learning literature more generally. Recent research has found that retrieval attempts lead to increased memory performance regardless of the time individuals spend attempting to retrieve the target. Specifically, Vaughn et al. (2017) instructed individuals to attempt to retrieve answers to trivia questions for either 0, 5, 10, or 30 s. Following each prompt, the answer was revealed, and a retest given after a 48-h delay. Performance on the retest was worst following the 0-s retrieval period condition. Surprisingly, in contrast to the authors' predictions, there was no learning benefit to the increased time, or effort, spent attempting to retrieve the answer (i.e., performance was equivalent following the 5, 10, and 30 s delays). Instead, differences emerged only when the researchers manipulated the time spent for processing the corrective feedback. The main conclusion was that additional time spent in retrieval mode conferred no benefit for learning. Here, our study complements this research from the retrieval-based learning domain in that extra time spent attempting retrieval confers no benefit on self-regulated learning.

Results from our study also suggest that there is not an inherent detriment to the act of explicitly requiring retrieval for self-regulated learning for older adults. If prompting older individuals to attempt retrieval represented too much of a demand on cognitive resources, then we would have observed poorer SRL for those individuals in the Retrieval + JOL group. Similarly, individuals in the JOL Only condition would therefore demonstrate better SRL (i.e., regulation in line with that of younger adults') because ending or not engaging in a retrieval search would spare cognitive resources. In contrast to this resource account, explicit retrieval could act as a means to encourage monitoring processes that otherwise would not occur, akin to providing individuals environmental support (e.g., Froger et al. 2012). As no differences in SRL were observed between older adults in the two monitoring conditions, neither account of the effect of an explicit retrieval prompt was supported.

Finally, that the two monitoring conditions employed in the present study yielded similar effects for the regulation of learning warrants consideration with respect to one other prevailing line of research–research into the reactivity of monitoring assessments themselves. Previously, we considered our study in relation to others which have held retrieval constant but differed monitoring (Robey et al. 2017) or have manipulated the presence or absence of retrieval (Miller and Geraci 2014). We have not considered our study with respect to those which manipulated the presence or absence of monitoring. The purpose of these reactivity studies is to determine whether the effect of making a monitoring judgment (e.g., JOLs) in and of itself causes memory enhancement effects (e.g., Soderstrom et al. 2015; Mitchum et al. 2016) or effects on strategy use/SRL (e.g., Double and Birney 2018). One of the main hypotheses driving the SRL effects is that making a monitoring judgment shifts attention and/or goal emphasis to information or strategies that would otherwise not be emphasized (*changed goal hypothesis*; Mitchum et al. 2016). For example, the inclusion of the probability scale (e.g., 0–100%) during monitoring could suggest to the participant that all information will not, in fact, be remembered, and so the participant may adjust their attention and/or study behavior accordingly.

The results of the present study complement reactivity studies by holding constant the emphasis on probability/monitoring but increasing the emphasis on retrieval and the products of retrieval in one condition (Retrieval + JOL). In the present case, no memory enhancement or SRL effects were observed, suggesting that the presence of a retrieval prompt may not shift goals relative to those emphasized as a consequence of soliciting delayed, cue-only JOLs. Future work could examine whether the presence of a retrieval prompt could shift goals to a greater extent when it precedes JOLs that do not place the same emphasis on retrieval from long term memory (e.g., immediate JOLs or cue-target JOLs). The present work, like work investigating monitoring reactivity, underscores the necessity of carefully considering how the requirements placed on participants could carry intended, and sometimes unintended, consequences for learning and self-regulation.

Conclusion

In the present study context, the retrieval prompt affected monitoring search time and consequently the confidence in the future retrievability of the target. The retrieval prompt did not, however, affect the SRL strategies one adopted. Differences in SRL were present only between young and older adults. In terms of older adults' SRL, we observed deficits compared to young adults (e.g., a weaker negative relationship between JOLs and SRL), equivalence compared to young adults (e.g., a similar positive relationship between search time and SRL), and differences in terms of restudy learning goals (e.g., wanting to restudy more words). The source of these differences and similarities appear to depend on the qualitative nature of the cues (i.e., experience- and information-based), reflecting the need for more research in this area.

Acknowledgements The experiment reported was included in the first author's Master's Thesis. Portions of this work were presented at the 58th Annual Meeting of the Psychonomic Society in Vancouver, BC, Canada.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval This study was approved by the Tufts University Institutional Review Board (IRB #1303045).

Informed consent Informed consent was obtained for all participants in line with ethical standards for the collection of data with human subjects.

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55–68. https://doi.org/10.1037/0096-3445.127.1.55.

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring memory in old age: Impaired, spared, and aware. In J. Dunlosky & S. (Uma) K. Tauber (Eds.), *The Oxford handbook of metamemory* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.3.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497–505.

Craik, F. I. M., Byrd, M., & Swanson, J. M. (1987). Patterns of memory loss in three elderly samples. *Psychology and Aging, 2*(1), 79–86. http://dx.doi.org.ezproxy.library.tufts.edu/10.1037/0882-7974.2.1.79

DeCaro, R., & Thomas, A. K. (2019). How attributes and cues made accessible through monitoring affect self-regulated learning in older and younger adults. *Journal of Memory and Language, 107*, 69–79.

Double, K. S., & Birney, D. P. (2018). Reactivity to confidence ratings in older individuals performing the latin square task. *Metacognition and Learning, 13*(3), 309–326. https://doi.org/10.1007/s11409-018-9186-5.

Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 52(4), P178–P186.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20*(4), 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33*(4), 545–565. https://doi.org/10.1006/jmla.1994.1026.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*(1), 37–56. https://doi.org/10.1016/S0001-6918(97)00051-6.

Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging, 18*(2), 340–345. https://doi.org/10.1037/0882-7974.18.2.340.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 238–244. https://doi.org/10.1037/0278-7393.33.1.238.

Fox, M. C., & Charness, N. (2010). How to gain eleven IQ points in ten minutes: Thinking aloud improves Raven's matrices performance in older adults. *Aging, Neuropsychology, and Cognition, 17*(2), 191–204.

Froger, C., Bouazzaoui, B., Isingrini, M., & Taconnat, L. (2012). Study time allocation deficit of older adults: The role of environmental support at encoding? *Psychology and Aging, 27*(3), 577–588. https://doi.org/10.1037/a0026358.

Grober, E., Sliwinsk, M., & Korey, S. R. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology, 13*(6), 933–949. https://doi.org/10.1080/01688639108405109.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science, 20*(3), 167–173. https://doi.org/10.1177/0963721411409026.

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*(2), 209–225.

Hines, J. C., Touron, D. R., & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging, 24*(2), 462–475. https://doi.org/10.1037/a0014417.

Hines, J. C., Hertzog, C., & Touron, D. R. (2015). Younger and older adults weigh multiple cues in a similar manner to generate judgments of learning. *Aging, Neuropsychology, and Cognition, 22*(6), 693–711. https://doi.org/10.1080/13825585.2015.1028884.

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*(1), 1–24.

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 923–954. https://doi.org/10.1037/a0026936.

Kolers, P. A., & Palef, S. R. (1976). Knowing not. *Memory & Cognition, 4*(5), 553–558.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increasedunderconfidence with practice. *Journal of Experimental Psychology: General, 131*(2), 147–162. https://doi.org/10.1037//0096-3445.131.2.147.

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. *A Handbook of Memory and Metamemory*, 117–136.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171.

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(4), 1011–1023.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 851–861.

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition, 29*, 131–140.

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*(2), 200–219. https://doi.org/10.1037/a0039923.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the" labor-in-vain effect.". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 676–686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bowers (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). Academic Press.

Nelson, H. E., & Willison, J. (1991). National Adult Reading Test (NART). Nfer-Nelson Windsor. http://www.academia.edu/download/31611053/NART_MANUAL.pdf

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004a). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004b). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*(1), 53–69. https://doi.org/10.1037/1082-989X.9.1.53.

Price, J. (2017). The impact of presentation format on younger and older adults' self-regulated learning. *Experimental Aging Research, 43*(4), 391–408. https://doi.org/10.1080/0361073X.2017.1333835.

Price, J., & Murray, R. G. (2012). The region of proximal learning heuristic and adult age differences in self-regulated learning. *Psychology and Aging, 27*(4), 1120–1129.

Price, J., Hertzog, C., & Dunlosky, J. (2010). Self-regulated learning in younger and older adults: Does aging affect metacognitive control? *Aging, Neuropsychology, and Cognition, 17*(3), 329–359. https://doi.org/10.1080/13825580903287941.

Rast, P., & Zimprich, D. (2009). Age differences in the Underconfidence-with-practice effect. *Experimental Aging Research, 35*(4), 400–431. https://doi.org/10.1080/03610730903175782.

Robey, A. M., Dougherty, M. R., & Buttaccio, D. R. (2017). Making retrospective confidence judgments improves learners' ability to decide what not to study. *Psychological Science, 28*(11), 1683–1693.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003.

Salthouse, T. (2010). *Major issues in cognitive aging*. Oxford University Press.

Schwartz, B. L., & Metcalfe, J. (1994). Methodological Probems and pitfalls in the study of human metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing (pp. 93–113)*. MIT Press.

Schwartz, B. L., Boduroglu, A., & Tekcan, A. İ. (2016). Methodological concerns: The feeling-of-knowing task affects resolution. *Metacognition and Learning, 11*(3), 305–316. https://doi.org/10.1007/s11409-015-9152-4.

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*(2), 553–558. https://doi.org/10.1037/a0038388.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition, 33*(6), 1116–1129.

Souchay, C., & Isingrini, M. (2012). Are feeling-of-knowing and judgment-of-learning different? Evidence from older adults. *Acta Psychologica, 139*(3), 458–464. https://doi.org/10.1016/j.actpsy.2012.01.007.

Tauber, S. K., & Rhodes, M. G. (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. *Psychology and Aging, 27*(2), 474–483. https://doi.org/10.1037/a0025246.

Thomas, A. K., Lee, M., & Balota, D. A. (2013). Metacognitive monitoring and dementia: How intrinsic and extrinsic cues influence judgments of learning in people with early-stage Alzheimer's disease. *Neuropsychology, 27*(4), 452–463. https://doi.org/10.1037/a0033050.

Tse, C.-S., Balota, D. A., & Roediger, H. L. (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging, 25*(4), 833–845. https://doi.org/10.1037/a0019933.

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review, 19*(4), 743–749.

Vaughn, K. E., Hausman, H., & Kornell, N. (2017). Retrieval attempts enhance learning regardless of time spent trying to retrieve. *Memory; Hove, 25*(3), 298–316. https://doi.org/10.1080/09658211.2016.1170152.

Vernon, D., & Usher, M. (2003). Dynamics of metacognitive judgments: Pre-and postretrieval mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3): 339.

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x.

West, R. L., Thorn, R. M., & Bagwell, D. K. (2003). Memory performance and beliefs as a function of goal setting and aging. *Psychology and Aging, 18*(1), 111–125. https://doi.org/10.1037/0882-7974.18.1.111.

Zachary, R. A. (1991). *Shipley institute of living scale*. Western Psychological Services: WPS.

Zakay, D., & Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychologica, 98*(1), 103–115.