# Metacognitive monitoring and strategic behaviour in working memory performance

By: Dayna R. Touron; Natalie Oransky; Matthew E. Meier; and Jarrod C. Hines

Touron, D. R., Oransky, N., Meier, M. E., & Hines, J. C. (2009). Metacognitive monitoring and strategic behavior in working memory performance: An examination of adult age differences. *Quarterly Journal of Experimental Psychology*, First published on: 29 December 2009 (iFirst).

Made available courtesy of Taylor & Francis:
http://www.informaworld.com/smpp/section?content=a918124031&fulltext=713240928

*****Note: Figures may be missing from this format of the document**

## Abstract:

Research indicates that cognitive age differences can be influenced by metacognitive factors. This research has generally focused on simple memory tasks. Age differences in working memory (WM) performance are pronounced, but are typically attributed to basic cognitive deficits rather than metacognitive factors. However, WM performance can be influenced by strategic behaviour that might be driven by metacognitive monitoring. In the current project, we attempted to connect these lines of research by examining age differences in metacognitive WM monitoring and strategies. In Experiment 1, younger and older adult participants completed a computerized operation span task in conditions that either required or did not require monitoring reports. Participants in the monitoring condition predicted and postdicted global performance for each block and rated their responses following each trial within a block. In Experiment 2, participants also reported their trial-level strategic approach. In contrast to the age equivalence typically found for simple memory monitoring, results demonstrated age differences in WM monitoring accuracy. Overall age differences in strategy use were not found, but using effective strategies benefited older adults' performance more than younger adults'. Furthermore, age-related differences in the WM task appear to be mediated by the accuracy of performance monitoring.

**Keywords:** Working memory; Metacognition; Monitoring; Strategies; Ageing

## Article:

The ability to simultaneously store and manipulate information appears to be a critical element of cognitive performance. Working memory (WM) has been depicted in separate accounts as a limited cognitive fuel (Craik & Byrd, 1982), as a central processing location with limited workspace (Baddeley, 1986), and as an index of the health of the neurological architecture (Conway, Kane, & Engle, 2003; Miyake & Shah, 1999).

Most research agrees that the capacity of WM is a fundamental determinant of age differences in complex cognition (i.e., Park et al., 1996; Salthouse & Babcock, 1991). This fuel, workspace, or health appears to decline with age, impacting the performance of complex cognitive tasks such as organization and decision making (e.g., Cherry & Park, 1993; Phillips, Gilhooly, Logie, Dela Salla, & Wynn, 2003; Zwahr, Park, & Shifren, 1999). For tasks that are cognitively undemanding, however (such as automatic processes or those that include external support), WM ability appears to be less consequential to age differences in performance (see Frieske & Park, 1993; Park, Hertzog, Kidder, Morrell, & Mayhorn, 1997).

Reductions in WM capacity with age have been ascribed to various general cognitive factors, including speed of processing (Salthouse, 1996), inhibitory control (Hasher & Zacks, 1988), and dual-task coordination (Verhaeghen, Kliegl, & Mayr, 1997). New research also suggests that the ability to switch attentional focus is critical to differences in WM (Cowan, 2001; McElree, 2001), and that attentional capacity expands with practice (Verhaeghen & Basak, 2005; Verhaeghen, Cerella, & Basak, 2004).

Performance on WM span tasks can be affected by strategic behaviour (McNamara & Scott, 2001). Span tasks are frequently used to measure WM capacity (Daneman & Carpenter, 1980; Turner & Engle, 1989). Participants solve multiple arithmetic equations or process multiple sentences while also retaining information for later recall, such as unrelated letters or words, with WM span gauged by recall performance. Notably, McNamara and Scott (2001) showed that simple training on an effective memory strategy markedly improved WM span.

Spontaneous strategic behaviour can also account for individual differences in WM scores. WM span is higher for individuals who spend more time encoding the to-be-recalled information (Engle, Cantor, & Carullo, 1992; Friedman & Mikaye, 2004) as well as for those who explicitly report strategic rehearsal (Dunlosky & Kane, 2007; Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001; Turley-Ames & Whitfield, 2003). Behaviour that influences WM task performance, such as strategy recognition and strategy implementation, may be guided by metacognitive monitoring and control. For example, some participants may learn from monitoring task performance that certain encoding strategies reinforce memorization and recall and then selectively implement these strategies to improve performance. Whether metacognitive processes are normatively and effectively utilized during WM tasks is not currently understood.

Metacognitive research has most prominently focused on judgements that predict associative memory performance at study, or that rate confidence in performance at test. Memory monitoring is generally accurate in terms of both relative accuracy—discrimination of learning level between items—and absolute accuracy—rating of overall performance level—with better accuracy when more proximal to the test experience (Benjamin, Bjork, & Schwartz, 1998; Nelson, Dunlosky, Graf, & Narens, 1994; Nelson & Narens, 1990).

Such monitoring does not decline dramatically with increasing age (Dunlosky & Hertzog, 2000; Hertzog & Dixon, 1994), despite declines in memory itself (see Kausler, 1994; Light, 1996). Age-equivalent discrimination is typically obtained for simple memory monitoring, although older adults' judgements are sometimes overconfident (Connor, Dunlosky, & Hertzog, 1997). Such outcomes are inconsistent with a *monitoring deficit hypothesis* (Hertzog & Dixon, 1994), which argues that age differences in memory monitoring account for differences in memory performance. Despite the preservation of memory monitoring ability, however, older adults might less optimally translate monitoring into metacognitive control (Bieman-Copland & Charness, 1994; Dunlosky & Connor, 1997), reducing the adaptability of strategic behaviour.

Most previous research on age differences in monitoring ability examined basic memory tasks rather than complex tasks that involve WM. Age differences in monitoring and control might be more pronounced with increasing task complexity, in accordance with the *complexity hypothesis* of cognitive ageing (e.g., Cerella, 1990). Older adults have shown deficient monitoring in complex memory tasks, such as in the learning of “deceptive” paired associates (Kelley & Sahakyan, 2003). Age differences in monitoring may influence complex memory performance, since the monitoring deficit hypothesis has not been evaluated for complex memory tasks.

Age differences in metacognitive monitoring and control could impact strategy production, effective strategy use, and task approach (see Hertzog, Vernon, & Rympa, 1993; Thapar, Ratcliff, & McKoon, 2003). Age differences in strategy recognition and implementation affect performance of tasks including associative learning, arithmetic computation, inductive reasoning, episodic memory, and mnemonic techniques (Dunlosky, Kubat-Silman, & Hertzog, 2003; Hertzog & Dunlosky, 2005; Lemaire, Arnaud, & Lecacheur, 2004; Saczynski, Willis, & Schaie, 2002; Verhaeghen & Marcoen, 1996).

The current article presents two experiments that examine metacognitive monitoring and strategic behaviour for WM task performance and whether WM monitoring and strategy selection differ for younger and older adults. We hypothesize that metacognitive monitoring, while age invariant for basic memory tasks, does decline with age for a more complex WM task. We also hypothesize that age and individual differences in monitoring accuracy influence subsequent metacognitive control strategies and, by extension, differences in WM task performance.

## EXPERIMENT 1

Experiment 1 examines metacognitive monitoring of WM performance for younger and older adults. Participants completed four blocks of a WM span task. Experimental conditions either did or did not require explicit performance monitoring; those in the monitoring condition reported global and trial-level judgements of performance. Older adults may not spontaneously monitor to enhance subsequent task performance (see Hertzog & Hultsch, 2000). If monitoring bolsters performance, smaller age differences in WM may be obtained for a condition that requires performance monitoring. However, explicit monitoring behaviour could draw resources from the task (e.g., Stine-Morrow, Shake, Miles, & Noh, 2006), producing lower performance and larger age differences in WM for a monitoring condition.

Groups were compared on WM performance, and those in the monitoring conditions were also compared on the relative accuracy (trial-level discriminability) and absolute accuracy (overall level of accuracy) of judgements. We anticipate the typically obtained age differences in WM performance. Generally for simple memory tasks, monitoring judgements are age equivalent; we expect that older adults' monitoring of the WM task will instead show lower relative accuracy as well as poorer absolute accuracy, in accordance with a complexity hypothesis. We also anticipate that monitoring accuracy will predict recall (i.e., span) in the WM task. Finally, we expect that monitoring accuracy will mediate age-related differences in WM span, consistent with a monitoring deficit hypothesis for age differences in complex memory task performance.

### Method

#### *Design*

The experiment had a 2 (age: young, old) ×2 (monitoring: yes, no) ×4 (block: 1-4) mixed factorial design, with age and monitoring as between-subjects independent variables and block as the within-subjects variable.

#### *Participants*

Younger adults were undergraduates who received course credit. Older adults were from the nearby community and were compensated $10 per hour. A total of 70 younger adults between 18 and 25 years of age (*M* = 18.76) and 56 older adults between 60 and 75 years of age (*M* = 66.86) participated. All were prescreened for basic health issues and had good corrected visual acuity (20/50 or better). None had previously completed a similar task (i.e., one that required metacognitive judgements of performance or a WM span task). Data for 13 participants (7 younger adults and 6 older adults) in the control condition were lost due to computer error.

Participants completed a brief cognitive battery. Age group characteristics differed in expected directions and are reported in Table 1. Via random assignment, 37 younger adults and 28 older adults were tested in the monitoring condition. Importantly, no reliable differences or interactions with the monitoring variable were found for participant characteristics.

**Table 1. *Means of participant characteristics for Experiment 1 and Experiment 2 by age group***

| | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|
| **Measure** | **Young** | **Old** | **Young** | **Old** |
| Age in years | 18.8 (0.12) | 66.9 (0.53) | 21.3 (0.52) | 67.3 (0.58) |
| Education[a,b] | 12.6 (0.11) | 16.3 (0.36) | 14.7 (0.93) | 15.6 (0.32) |
| Medications[a,b] | 0.84 (0.14) | 2.5 (0.26) | 1.2 (0.24) | 2.1 (0.27) |
| Vocabulary[a,b] | 14.3 (0.45) | 21.5 (1.04) | 13.6 (0.94) | 20.0 (1.05) |
| F/L Names[a,b] | 14.5 (0.73) | 10.0 (0.76) | 14.3 (0.29) | 11.9 (1.04) |
| Digit Symb[a,b] | 62.4 (1.33) | 48.9 (1.54) | 63.5 (1.52) | 50.1 (1.72) |
| DS Memory[a,b] | 7.9 (0.17) | 6.3 (0.29) | 7.1 (0.32) | 5.3 (0.37) |

*Note:* Education = number of years of education completed. Medications = self-reported number of daily medications taken. Vocabulary = number correct out of 40 on the Shipley Vocabulary Test (Zachary, 1986). F/L Names = ETS First and Last Names test of associative memory (Ekstrom, French, & Harman, 1976). Digit Symb = Wechsler Adult Intelligence Scale (WAIS) Digit Symbol subtest of perceptual speed (Wechsler, 1981). DS Memory = symbol recall memory following the WAIS Digit Symbol subtest (Wechsler, 1981), an index of incidental memory. Standard errors in parentheses.

[a] Experiment 1 age comparison $p < .05$. For Experiment 1, no comparisons of or interactions with the monitoring variable were significant. [b] Experiment 2 age comparison $p < .05$.

### *Materials and procedure*

Participants completed four blocks of a computerized operation span WM task. Participants verified the solutions for a series of arithmetic equations—for example, (2 ×4) - 3 = 7—while remembering a series of letters (one letter following the solution of each equation) for later recall. Equations were randomly selected from the full population of equations with a first digit less than 20, second digit less than 10, and third digit less than 10. The first operator was multiplication or division, and the second operator was addition or subtraction. A population of 100 letter series each of the lengths 2 to 6 was randomly generated; stimuli were selected from a subpopulation of the least memorable and pronounceable. Stimuli were shown in 15-point Arial font on a 15-inch 4:3 LCD monitor with a resolution of 1,024 ×768. Participants sat at a height and distance that optimized screen viewing and comfort.

Self-paced instructions and practice preceded the task. Equation practice included 50 trials. Letter recall practice included 1 trial for each list length 2 to 6. Combined practice included 1 trial for each length with both equation and letter components.

In the task, participants viewed each equation separately and judged whether (or not) the provided solution was correct by pressing keys marked “Y” for yes and “N” for no on the keypad. Equations were presented until response and were followed by letter presentation for 800 ms to younger adults and for 1,300 ms to older adults (to compensate for general slowing; e.g., Salthouse, 1996). After each trial, participants recalled the letter series by typing in correct serial order with forgotten letters keyed with the spacebar. Each of the four blocks contained 15 trials (3 trials each for memory span 2 to 6 in random order). Participants were instructed to maintain at least 85% equation accuracy and were given feedback and warned at the end of a trial when accuracy fell below 85%. All participants were offered a brief break after each block.

Participants in the monitoring condition provided the following judgements. (a) Before each block, participants provided a global prediction of how many letters they would correctly recall for each list length 2 to 6 (between 0 and 6). (b) After each trial, participants rated confidence in their recall performance (continuous from 0 = no confidence to 100 = full confidence). (c) After each block, participants provided a global postdiction of how many letters they correctly recalled for each list length 2 to 6 (between 0 and 6).

The task was followed by a survey of performance judgements. Participants rated both the equation and letter components of the task in terms of: (a) perceived importance of accuracy versus speed, (b) ease of understanding the task instructions, and (c) task difficulty. Participants in the monitoring condition also rated each of the monitoring requirements in terms of difficulty. Ratings were made on a 0-100 scale with higher ratings indicating greater values.

## Results

Data were analysed as follows unless otherwise specified. To evaluate task approach and performance, survey and WM task data were compared using a 2 (age: young, old) ×2 (condition: control, monitoring) ×4 (block) repeated measures analyses of variance (ANOVAs). To evaluate WM monitoring for those in the monitoring condition, comparisons of judgements and monitoring accuracy used 2 (age: young, old) ×4 (block) repeated measures ANOVAs. Alpha level for all tests reported was set to .05, and partial eta-squared was included to measure effect size. The monitoring deficit hypothesis was examined using mediation analyses described below.

### Posttask judgements

Appropriately, both younger and older adults believed that accuracy was more important than speed in this task, for both the equations ($M_{\text{accuracy}}$ = 91.8, $M_{\text{speed}}$ = 68.3) and recall ($M_{\text{accuracy}}$ = 92.3, $M_{\text{speed}}$ = 69.2) . Ratings for task understanding were generally above 95% ($M$ = 98.4) and did not vary by component or group. Older adults rated both the equation and letter recall tasks as more difficult than did younger adults, $F(1, 125) = 5.29$, $p$ = .023 ($M_{\text{young}}$ = 10.5, $M_{\text{old}}$ = 21.0), and $F(1, 125) = 5.95$, $p$ = .016 ($M_{\text{young}}$ = 30.3, $M_{\text{old}}$ = 42.8), respectively. Both young and older adults found it minimally difficult to provide predictions ($M$ = 31.8), postdictions ($M$ = 34.4), and trial-level confidence judgements ($M$ = 25.4).

### Working memory task performance

Expected age differences in WM were obtained. As noted earlier, participants were warned to be more accurate when their equation accuracy fell below 85%; such warnings were infrequent (0.03% of trials).1 Letter recall performance was defined as the percentage of presented letters correctly recalled for a trial regardless of serial order. This definition corresponded to the wording of metacognitive judgements. When data were otherwise analysed (such as with correct serial order or binary coding), outcomes were qualitatively similar. Note that our measure is less conservative than alternatives, leading to higher reported performance; extended practice might also have elevated performance.

Younger adults were more accurate in recalling letters—our measure of WM span ($M_{\text{young}}$ = 96.6, $M_{\text{old}}$ = 92.2), $\eta_p^2$ = .14, $F(1, 112) = 5.94$, $MSE$ = 343, $p$ = .02. Participants in the control condition were more accurate than were participants in the monitoring condition ($M_{\text{control}}$ = 96.2, $M_{\text{monitoring}}$ = 92.5), $\eta_p^2$ = .10, $F(1, 112) = 5.94$, $MSE$ = 343, $p$ = .02. Performance did not vary reliably with experience, and no further interactions approached statistical significance.

Because performance and monitoring might vary with task difficulty, we also considered age and group differences by recall list length (see Table 2). As expected, performance was lower for larger trial lengths, $\eta_p^2$ = .26, $F(4, 448) = 34.9$, $MSE$ = 75, $p < .01$, and this length effect was more pronounced for older adults (e.g., $M_{\text{young,length6-length2}}$ = -2.6, $M_{\text{old,length6-length2}}$ = -9.7), $\eta_p^2$ = .05, $F(4, 448) = 3.18$, $MSE$ = 75, $p$ = .01. Further comparisons were not significant.

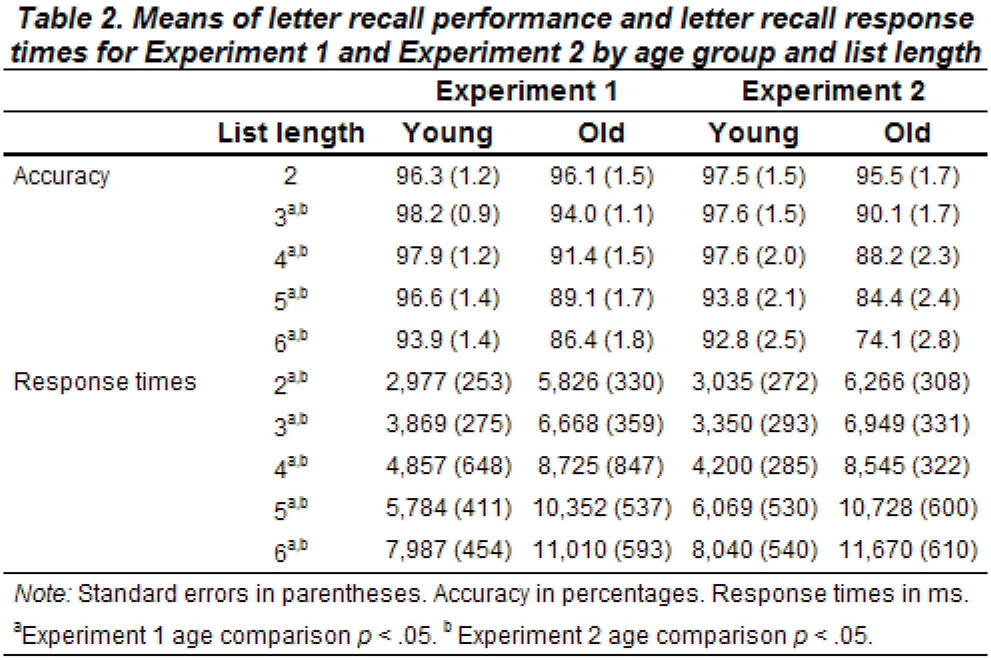


***Table 2. Means of letter recall performance and letter recall response times for Experiment 1 and Experiment 2 by age group and list length***

| | | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|---|
| | **List length** | **Young** | **Old** | **Young** | **Old** |
| Accuracy | 2 | 96.3 (1.2) | 96.1 (1.5) | 97.5 (1.5) | 95.5 (1.7) |
| | 3[a,b] | 98.2 (0.9) | 94.0 (1.1) | 97.6 (1.5) | 90.1 (1.7) |
| | 4[a,b] | 97.9 (1.2) | 91.4 (1.5) | 97.6 (2.0) | 88.2 (2.3) |
| | 5[a,b] | 96.6 (1.4) | 89.1 (1.7) | 93.8 (2.1) | 84.4 (2.4) |
| | 6[a,b] | 93.9 (1.4) | 86.4 (1.8) | 92.8 (2.5) | 74.1 (2.8) |
| Response times | 2[a,b] | 2,977 (253) | 5,826 (330) | 3,035 (272) | 6,266 (308) |
| | 3[a,b] | 3,869 (275) | 6,668 (359) | 3,350 (293) | 6,949 (331) |
| | 4[a,b] | 4,857 (648) | 8,725 (847) | 4,200 (285) | 8,545 (322) |
| | 5[a,b] | 5,784 (411) | 10,352 (537) | 6,069 (530) | 10,728 (600) |
| | 6[a,b] | 7,987 (454) | 11,010 (593) | 8,040 (540) | 11,670 (610) |

*Note:* Standard errors in parentheses. Accuracy in percentages. Response times in ms.

[a]Experiment 1 age comparison $p < .05$. [b] Experiment 2 age comparison $p < .05$.

Letter recall reaction time (RT) was recorded from presentation of the recall input box until the participant pressed the "enter" key. Participant medians for correct responses were analysed to reduce the influence of positive skew and outliers that occur infrequently; we report group means of participant medians. Younger adults were faster to recall letter series than were older adults, $\eta_p^2 = .82$, $F(1, 112) = 101.64$, $MSE = 8{,}267{,}824$, $p < .01$. Participants in the control condition were faster to recall than were participants in the monitoring condition, $\eta_p^2 = .21$, $F(1, 112) = 6.11$, $MSE = 8{,}267{,}824$, $p = .02$, but no interactions with the condition variable were significant. Recall RTs improved with experience, $\eta_p^2 = .30$, $F(3, 336) = 37.65$, $MSE = 702{,}622$, $p < .01$, and improvements were greater for older than for younger adults (e.g., $M_{\text{young,Block1-Block4}} = 725$, $M_{\text{old,Block1-Block4}} = 1615$), $\eta_p^2 = .06$, $F(3, 336) = 5.80$, $MSE = 702{,}622$, $p < .01$.

### *Monitoring judgements*

Predictions of letter recall performance were higher for younger than for older adults ($M_{\text{young}} = 94.1$, $M_{\text{old}} = 83.5$), $\eta_p^2 = .26$, $F(1, 63) = 8.79$, $MSE = 774$, $p < .01$. Postdiction data revealed a similar age difference ($M_{\text{young}} = 94.7$, $M_{\text{old}} = 84.1$), $\eta_p^2 = .31$, $F(1, 63) = 6.36$, $MSE = 1{,}081$, $p = .01$. Neither global rating varied by or interacted with block.

Younger adults also reported more confidence in their recall performance than did older adults ($M_{\text{young}} = 95.2$, $M_{\text{old}} = 83.0$), $\eta_p^2 = .27$, $F(1, 63) = 7.89$, $MSE = 1{,}154$, $p < .01$. Although confidence did not change overall, the age by block interaction was significant, $\eta_p^2 = .02$, $F(1, 63) = 4.25$, $MSE = 32$, $p < .01$, with slight but reliable decreases by younger adults and increases by older adults (e.g., $M_{\text{young,Block1-Block4}} = 3.01$, $M_{\text{old,Block1-Block4}} = -3.55$).

### *Relative accuracy of confidence judgements*

Research in metacognition typically relies on gamma correlations between performance and judgements to measure monitoring accuracy (see Nelson, 1984). To increase gamma stability we collapsed the four blocks into two parts; note that gamma is indeterminate for a participant if performance or judgements are invariant. Discriminability of recall confidence judgements was better for younger than for older adults, $\eta_p^2 = .22$, $F(1, 38) = 8.58$, $MSE = .07$, $p < .01$ (see Table 3). The main effect of block was nonsignificant, but the age by block interaction indicated that monitoring improved somewhat for older adults while decreasing for young, $\eta_p^2 = .06$, $F(1, 35) = 4.45$, $MSE = .03$, $p = .04$.

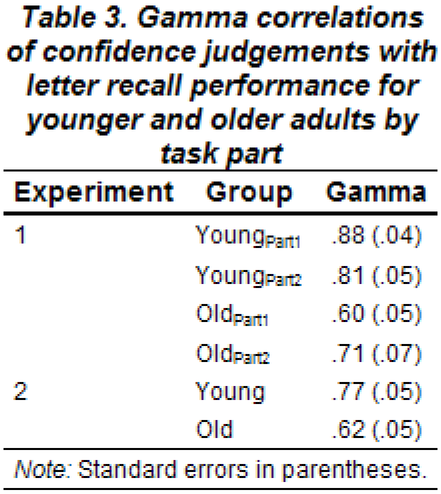


***Table 3. Gamma correlations of confidence judgements with letter recall performance for younger and older adults by task part***

| Experiment | Group | Gamma |
|---|---|---|
| 1 | Young$_{Part1}$ | .88 (.04) |
| | Young$_{Part2}$ | .81 (.05) |
| | Old$_{Part1}$ | .60 (.05) |
| | Old$_{Part2}$ | .71 (.07) |
| 2 | Young | .77 (.05) |
| | Old | .62 (.05) |

*Note:* Standard errors in parentheses.

To consider the effects of task difficulty on relative accuracy, we compared younger and older adult gamma correlations on short (2-3) and long (4-6) trial lengths. Discrimination of confidence judgements was more accurate for short lengths ($M = .84$) than for long lengths ($M = .65$), $\eta_p^2 = .24$, $F(1, 22) = 12.15$, $MSE = .03$, $p < .01$, but this did not vary by age ($p = .9$).

### *Absolute accuracy of monitoring judgements*

Absolute accuracy was computed as the difference between provided ratings and actual performance; values greater than zero are overconfident, and values less than zero are underconfident (see Figure 1). Prediction accuracy did not vary with the main effects of age or block, but the reliable interaction, $\eta_p^2 = .20$, $F(3, 189) = 2.60$, $MSE = 169$, $p = .05$, indicated that initial underconfidence resolved toward greater accuracy for younger but not older adults. Although the accuracy of letter recall confidence did not reliably differ overall by age group or block, a significant interaction, $\eta_p^2 = .01$, $F(3, 189) = 2.82$, $MSE = 15$, $p = .04$, showed slightly decreasing underconfidence by older adults and slightly increasing underconfidence by younger adults. Accuracy of postdictions for letter recall performance did not differ by age, by block, or with the interaction, but trends of greater underconfidence by older adults were supported post hoc in Blocks 2 ($p = .05$) and 3 ($p = .03$).

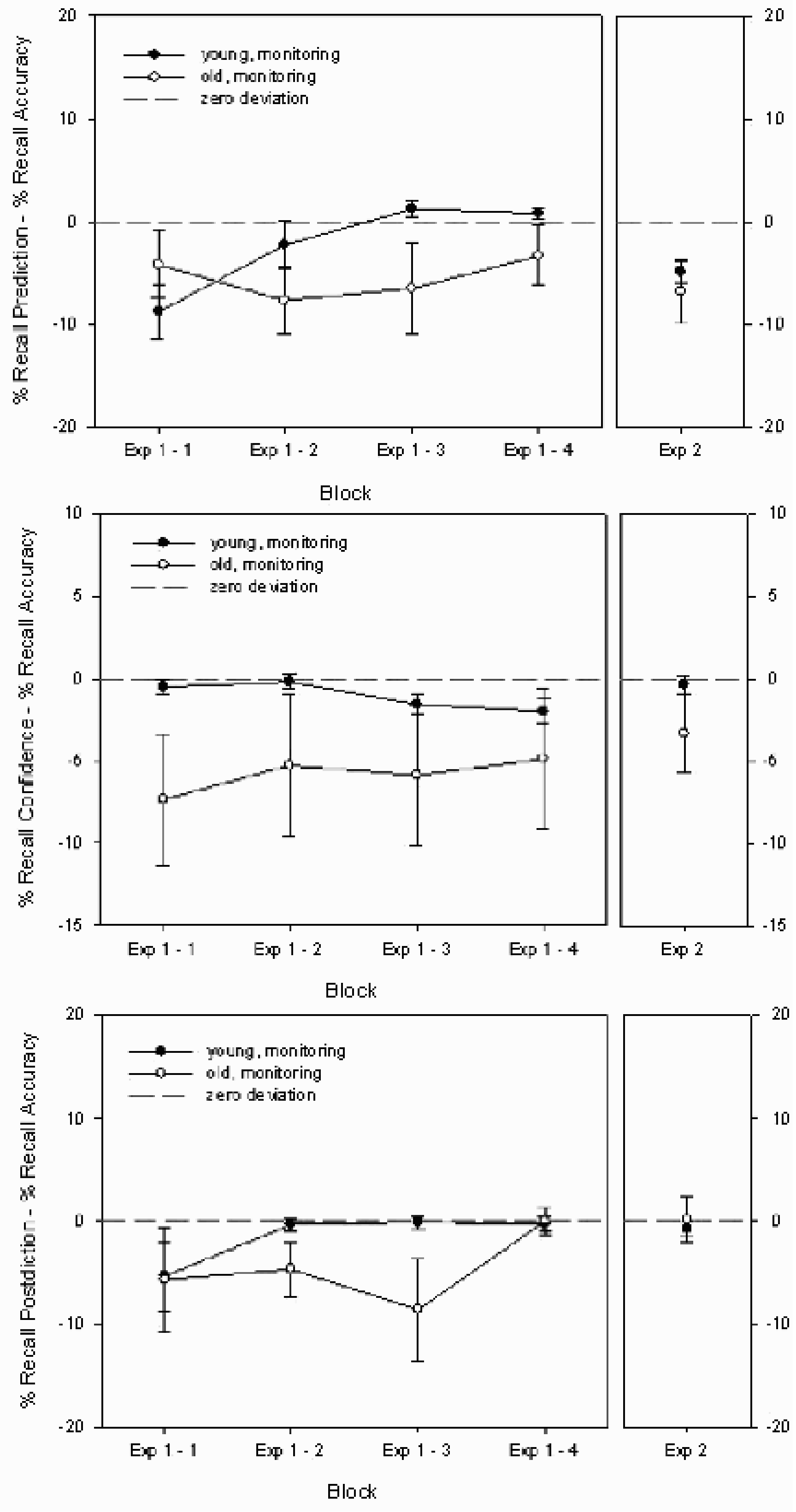


Figure 1. Absolute accuracy of global letter recall predictions (top), letter recall confidence judgements (middle), and global postdictions (bottom) by age and block for Experiment 1 and by age only for Experiment 2.

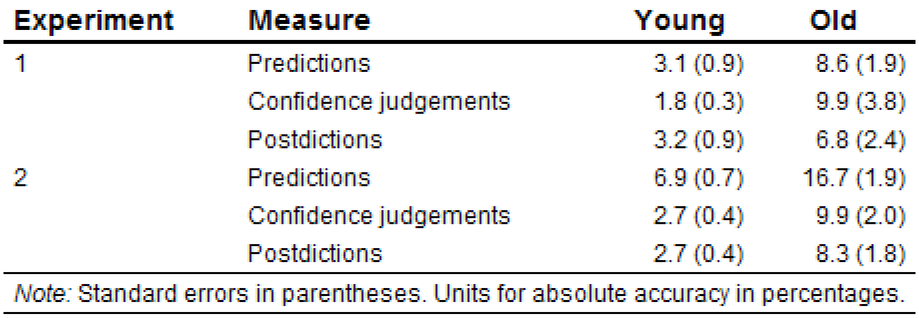
To follow up on the age by length interaction noted for recall performance, we considered age and group differences in monitoring accuracy across recall list lengths. Confidence judgements were only marginally more underconfident for long lengths than for short lengths (e.g., $M_{\text{length2}}$ = -2.0, $M_{\text{length6}}$ = -4.8), $\eta_p^2$ = .06, *F*(4, 252) = 2.21, *MSE* = 66, *p* = .07. In contrast to the performance data, there was no length by age interaction (*p* = .17).

### *Relationships between measures*

To test the monitoring deficit hypothesis—that monitoring accuracy mediates age-related WM declines—we used mediation analysis to assess direct and indirect relationships between age, monitoring, and letter recall performance. We followed the approach outlined by Baron and Kenny (1986): An initial variable is mediated when reliable relationships occur between (a) the initial variable and the outcome variable, (b) the initial variable and the mediator variable, and (c) the mediator variable and the outcome variable, such that the relationship between the initial variable and the outcome variable is reduced (partial mediation) or eliminated (full mediation) when controlling on the mediator. Each step requires estimation of the resulting regression equation. Below we report the standardized estimates and significance tests as well as an index of model fit ($R^2$, adjusted for model complexity).

We included age as a categorical variable in these regressions, while recall performance and indices of monitoring accuracy were continuous variables. Absolute monitoring accuracy for each judgement type (predictions, confidence judgements, and postdictions) was entered in absolute value to aid interpretability, such that the direction of errant calibration—underconfidence versus overconfidence—was irrelevant, and relationships between the simple magnitude of calibration error and performance were evaluated. Directional data were presented earlier to facilitate comparison with previous work; means for absolute values are given in Table 4. The translation to absolute values magnified age differences in absolute accuracy; older adults were more poorly calibrated in general (i.e., when direction is disregarded) for their predictions, *F*(1, 63) = 9.27, *MSE* = 49, *p* < .01, and postdictions differed with marginal reliability (*p* = .08).

***Table 4. Means of absolute values for absolute accuracy of judgements by younger and older adults***



| Experiment | Measure | Young | Old |
|---|---|---|---|
| 1 | Predictions | 3.1 (0.9) | 8.6 (1.9) |
| | Confidence judgements | 1.8 (0.3) | 9.9 (3.8) |
| | Postdictions | 3.2 (0.9) | 6.8 (2.4) |
| 2 | Predictions | 6.9 (0.7) | 16.7 (1.9) |
| | Confidence judgements | 2.7 (0.4) | 9.9 (2.0) |
| | Postdictions | 2.7 (0.4) | 8.3 (1.8) |

*Note:* Standard errors in parentheses. Units for absolute accuracy in percentages.

Including accuracy in absolute value also aided data transformation, as common transformation methods require positive values. Multiple regression analyses assume variable normality. Some degree of skew was seen for all continuous variables other than absolute prediction accuracy (all Shapiro-Wilk *W*s < .80, *p*s < .01). Each was transformed toward normality prior to the regression analyses (all Shapiro-Wilk *W*s > .95, *p*s > .05).

In the first step of the mediation analysis, we examined a model including age as the predictor variable and recall performance as the criterion variable. As expected, age was a reliable predictor of recall performance ($adjR^2$ = .06, $\beta$ = .24), *t*(1) = 2.61, *p* = .01.

In the second step, we examined models including age as the predictor variable and indices of monitoring accuracy as criterion variables. Age was a reliable predictor of absolute prediction accuracy ($adjR^2$ = .19, $\beta$ = .44), *t*(1) = 2.45, *p* = .02, the absolute accuracy of confidence judgements ($adjR^2$ = .29, $\beta$ = .54), *t*(1) = 3.28, *p* <

.01, and the relative accuracy of confidence judgements ($adjR^2 = .07$, $\beta = -.30$), $t(1) = -2.17$, $p = .03$, but age was not a reliable predictor of postdiction accuracy ($adjR^2 = -.03$, $\beta = -.02$), $t(1) < 1$.

In the third step, we examined a model of recall performance including (as predictors) the age variable as well as potential mediating variables—absolute prediction accuracy and the relative accuracy of confidence judgements. According to Baron and Kenny (1986), multiple mediators can be tested simultaneously to assess independent effects as long as they are not highly correlated. We therefore included variables that were not reliably correlated ($r = -.11$, $p = .4$) and are fairly conceptually distinct. The third condition of mediation was met; the age effect was no longer reliable with an estimate close to zero and therefore appears to be fully mediated by metacognitive monitoring ($adjR^2 = .50$, $\beta = -.04$), $t(1) < 1$. Absolute prediction accuracy again contributed reliable variance in WM performance ($\beta = -.76$), $t(1) = 4.31$, $p < .01$, but the relative accuracy of confidence judgements did not contribute an independent effect ($\beta = -.13$), $t(1) < 1$. There is a risk that obtained relationships between recall performance and judgement accuracy might reflect the fact that indices of monitoring accuracy are derived from the performance measure. Given that obtained relationships with performance vary among monitoring measures, it does not appear that this accounts for our findings. Sobel's $z$-test confirmed the interpretation of absolute prediction accuracy as a mediator of age differences in recall, $z = 2.13$, $p = .03$.

## Discussion

As expected, older adults performed more poorly than younger adults in the WM operation span task. It does not appear that age differences reflect general task approach; younger and older adults showed a comparable mental task model. For both young and older adults, letter recall performance was poorer in the condition that required monitoring judgements. This suggests that the requirement to explicitly monitor and report judgements detracts from rather than supporting the WM task; spontaneous monitoring might be less demanding.

It is also noteworthy that the outcomes from Experiment 1 demonstrate age-related declines in metacognitive monitoring for a WM span task. Monitoring judgements did reflect age differences in performance. However, general and trial-level performance ratings were less accurate by older adults than by younger adult participants. As hypothesized, older adults were less able to discriminate trial-level performance with recall confidence judgements—the preferred index of monitoring ability. Older adults were more underconfident in their letter recall performance; confidence judgements were underconfident in the first block, while global judgements were underconfident only in later blocks. When translated into absolute values, age differences in the absolute accuracy of judgements were magnified. Importantly, these outcomes contrast with age equivalence in monitoring associative memory, where older adults generally show no age deficits in trial-level discrimination and produce judgements that are well calibrated or overconfident (e.g., Connor et al., 1997). It seems that age deficits in monitoring can occur when performing complex cognitive tasks.

Mediation analyses support a monitoring deficit hypothesis for age differences in WM. The relationship between age and WM task performance becomes nonsignificant when the absolute accuracy of performance predictions is considered. Age differences in WM may be influenced by differences in monitoring accuracy. Note, however, that given the nature of correlation analyses, it is also possible that this relationship is reversed, with differences in monitoring accuracy driven by WM. Further research is needed to further define this causal direction. Note, however, that the mediation of WM performance by predictions prior to the task rather than posttask judgements is causally consistent with a framework that assumes that metacognitive monitoring influences strategic task approach and thereby task performance.

## EXPERIMENT 2

Experiment 2 examines the implications of metacognitive monitoring of WM performance for strategic task approach. Encoding strategies can bolster WM task performance, and age differences in strategy use influence age-related performance differences in various cognitive tasks. It follows that older adults may use less effective WM performance strategies.

Participants completed the WM task with metacognitive judgements and then reported strategy use. We used a method adopted by Dunlosky and Kane (2007) and by Bailey, Dunlosky, and Hertzog (2009) in an extension with older adults (also see Dunlosky & Hertzog, 2001; Hertzog & Dunlosky, 2005). Use of retrospective strategies and a single task block avoided problems of reactivity to the strategy report procedure. Participants indicated the memory strategy used on a given recall trial by selecting from a list of options ranging in normative effectiveness (see Craik, 2002). We expect to replicate the outcomes above supporting age-related declines in both the relative and absolute accuracy of WM monitoring, as well as a mediator relationship between monitoring and performance. We further anticipate that differences in strategy use relate to task performance and monitoring, such that greater use of effective task strategies will result from better monitoring and lead to better performance.

## Method

### Participants

Participants were recruited and compensated as in Experiment 1. A total of 45 younger adults between 18 and 25 years of age and 36 older adults between 60 and 75 years of age participated. Group characteristics again differed in expected directions (see Table 1).

### Procedure and materials

Participants completed one block of the task with monitoring judgements from Experiment 1. After the main task, participants provided computerized strategy reports. Letter sequences were shown in the original presentation order with seven strategy choices given concurrently, as follows: “How did you originally try to remember the letters from the series above? 1 = read each letter as it appeared, 2 = repeated the letters as much as possible, 3 = linked the letters together using sounds, 4 = linked the letter together using words, 5 = developed mental images of words linked to the letters, 6 = grouped the letters in a meaningful way, 7 = did something else.” The task was followed by a survey; in addition to the judgements in Experiment 1, participants also rated the difficulty of providing strategy reports.

## Results

Analysis of Experiment 2 data was generally consistent with the approach taken above. Strategy reports were compared across groups and related to monitoring and performance.

### Posttask judgements

Both younger and older adults again appropriately believed that accuracy was more important than speed in both the equation ($M_{\text{accuracy}} = 90.9$, $M_{\text{speed}} = 67.6$) and letter ($M_{\text{accuracy}} = 90.8$, $M_{\text{speed}} = 65.7$) components of the task. Ratings for task understanding were again high; understanding of the equation task did not vary by group ($M = 94.1$), but older adults reported slightly lower understanding of the letter recall task, $F(1, 79) = 8.26$, $p = .005$ ($M_{\text{young}} = 98.7$, $M_{\text{old}} = 92.2$). Older adults did not rate the equation and letter recall tasks as more difficult than did younger adults ($M_{\text{equation}} = 22.4$, $M_{\text{recall}} = 43.4$). Ratings for the monitoring requirements generally did not differ by age, but older adults did rate providing letter confidence judgements as more difficult, $F(1, 79) = 8.01$, $MSE = 7{,}227$, $p = .006$ ($M_{\text{young}} = 28.0$, $M_{\text{old}} = 45.0$). Older adults did not report more difficulty with strategy reporting ($M$ = 27.8). Some age differences vary from Experiment 1, probably due to the additional blocks completed in that study.

### Working memory task performance

Younger adults were again more accurate ($M_{young}$ = 95.8, $M_{old}$ = 86.1), $\eta_p^2$ = .19, $F(1, 80)$ = 18.23, $MSE$ = 1,063, $p < .01$, and faster ($M_{young}$ = 4,939, $M_{old}$ = 8,832), $\eta_p^2$ = .50, $F(1, 80)$ = 79.98, $MSE$ = 3,825,875, $p < .01$, in recalling letters.

We again considered age and group differences in performance across stimulus trial lengths (lengths 2-6; see Table 2). As in Experiment 1, performance accuracy was lower for larger trial lengths (e.g., $M_{length2}$ = 95.7, $M_{length6}$ = 84.6), $\eta_p^2$ = .24, $F(4, 320)$ = 20.86, $MSE$ = 90, $p < .01$, and the length effect was again more pronounced for older adults (e.g., $M_{young,length6-length2}$ = -4.8, $M_{old,length6-length2}$ = -19.4), $\eta_p^2$ = .09, $F(4, 320)$ = 6.66, $MSE$ = 90, $p < .01$.

### Monitoring judgements

Outcomes were consistent with those of Experiment 1. Predictions of letter recall performance were higher for younger than for older adults ($M_{young}$ = 91.1, $M_{old}$ = 79.4), $\eta_p^2$ = .25, $F(1, 80)$ = 27.22, $MSE$ = 101, $p < .01$. Postdictions of letter recall performance also varied by age group ($M_{young}$ = 95.0, $M_{old}$ = 86.2), $\eta_p^2$ = .13, $F(1, 80)$ = 11.77, $MSE$ = 134, $p < .01$. Younger adults were more confident in their recall performance than were older adults ($M_{young}$ = 95.5, $M_{old}$ = 82.8), $\eta_p^2$ = .23, $F(1, 80)$ = 24.56, $MSE$ = 134, $p < .01$.

### Relative accuracy of confidence judgements

Discriminability of recall confidence judgements was again better for younger adults than for older adults, $\eta_p^2$ = .08, $F(1, 58)$ = 4.85, $MSE$ = .09, $p$ = .03 (see Table 3). We again compared old and younger adult gamma correlations on short (2-3) and long (4-6) trial lengths. The discrimination of confidence judgements was again more accurate for short lengths ($M$ = .75) than for long lengths ($M$ = .58), $\eta_p^2$ = .04, $F(1, 12)$ = 7.36, $MSE$ = .3, $p$ = .02, but this effect did not interact with age ($p$ = .6).

### Absolute accuracy of monitoring judgements

Absolute accuracy was computed as above (see Figure 1), and outcomes were generally consistent with Experiment 1 Block 1. Neither prediction nor postdiction accuracy differed by age. Confidence judgement accuracy did not vary by age, in contrast to the first block of Experiment 1. However, the obtained pattern of more underconfidence by older adults was consistent, $\eta_p^2$ = .04, $F(1, 80)$ = 1.77, $MSE$ = 101, $p$ = .19.

We again examined age and group differences in monitoring accuracy by stimulus trial length. The main effect of length was nonsignificant, but a length by age interaction indicated that younger adults were more underconfident for large lengths while older adults were more underconfident for small lengths (e.g., $M_{young,length6-length2}$ = -3.5, $M_{old,length6-length2}$ = 5.5), $\eta_p^2$ = .05, $F(4, 320)$ = 2.68, $MSE$ = 88, $p$ = .03.

### Relationships between measures

We again assessed the monitoring deficit hypothesis using mediation analysis. Absolute accuracy was again translated into absolute values, magnifying age differences in absolute accuracy. Variables were again not initially normally distributed (before transformation Shapiro-Wilk $W$s < .85, $p$s < .01; after transformation all Shapiro-Wilk $W$s > .94, $p$s > .05).

In the first step of the mediation analysis, we examined a model including age as the predictor variable and monitoring accuracy as the criterion variable. As expected, age was a reliable predictor of recall performance ($adjR^2 = .18$, $\beta = .44$), $t(1) = 3.8$, $p < .01$.

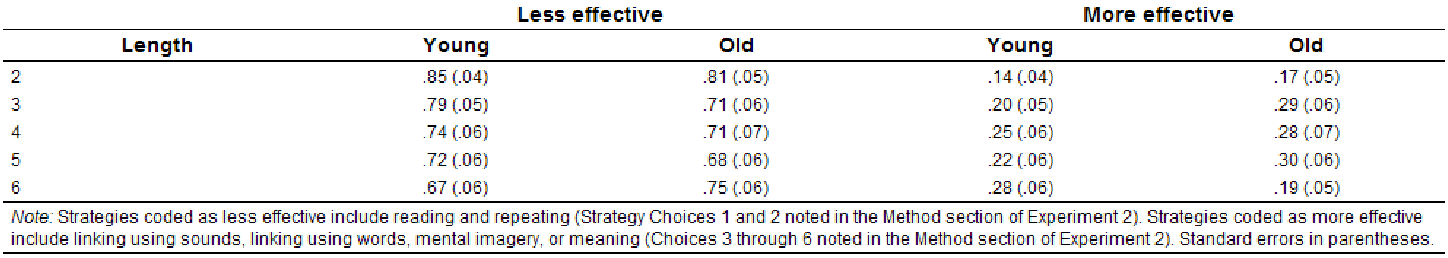
In the second and third steps, we tested the mediation model obtained in Experiment 1. In Step 2, we included age as the predictor variable and absolute prediction accuracy as a criterion variable. Age was again a reliable predictor of absolute prediction accuracy ($adjR^2 = .11$, $\beta = .35$), $t(1) = 3.28$, $p < .01$. In Step 3, we included recall performance as the criterion variable predicted by age and absolute prediction accuracy as a potential mediator. The third condition of mediation was met; the age effect was reduced and in this experiment appears to be partially mediated by metacognitive monitoring ($adjR^2 = .24$, $\beta = .29$), $t(1) = 2.4$, $p = .02$. Absolute prediction accuracy again contributed reliable variance in WM performance ($\beta = .32$), $t(1) = 2.53$, $p = .02$. Sobel's $z$-test confirmed this interpretation, $z = 2.11$, $p = .03$.

### *Strategy reports*

Proportions of strategy reports by age, normative effectiveness, and list length are given in Table 5. No age differences in strategy reports were noted, and age did not interact with any other variable. Participants reported strategies with low effectiveness more often than strategies with high effectiveness, $\eta_p^2 = .84$, $F(1, 79) = 66.96$, $MSE = .78$, $p < .01$. The list length comparison was also significant, $\eta_p^2 = .01$, $F(4, 316) = 3.65$, $MSE = .01$, $p = .02$, and was qualified by an effectiveness by length interaction, $\eta_p^2 = .13$, $F(4, 316) = 3.28$, $MSE = .11$, $p = .02$, indicating that participants were more likely to report effective strategies at long list lengths and more likely to report less effective strategies at short list lengths.

We also examined recall accuracy and monitoring accuracy by age, list length, and normative strategy effectiveness (see Figure 2). Because few participants used each type of strategy at each list length, we collapsed lengths into short (2-3) and long (4-6) and compared age (young, old), effectiveness (high, low) and length (short, long). We used SAS PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 1996) to include all available data.

***Table 5. Mean proportions of strategy report for younger and older adults by normative effectiveness and list length***



| | Less effective | | More effective | |
|---|---|---|---|---|
| **Length** | **Young** | **Old** | **Young** | **Old** |
| 2 | .85 (.04) | .81 (.05) | .14 (.04) | .17 (.05) |
| 3 | .79 (.05) | .71 (.06) | .20 (.05) | .29 (.06) |
| 4 | .74 (.06) | .71 (.07) | .25 (.06) | .28 (.07) |
| 5 | .72 (.06) | .68 (.06) | .22 (.06) | .30 (.06) |
| 6 | .67 (.06) | .75 (.06) | .28 (.06) | .19 (.05) |

*Note:* Strategies coded as less effective include reading and repeating (Strategy Choices 1 and 2 noted in the Method section of Experiment 2). Strategies coded as more effective include linking using sounds, linking using words, mental imagery, or meaning (Choices 3 through 6 noted in the Method section of Experiment 2). Standard errors in parentheses.

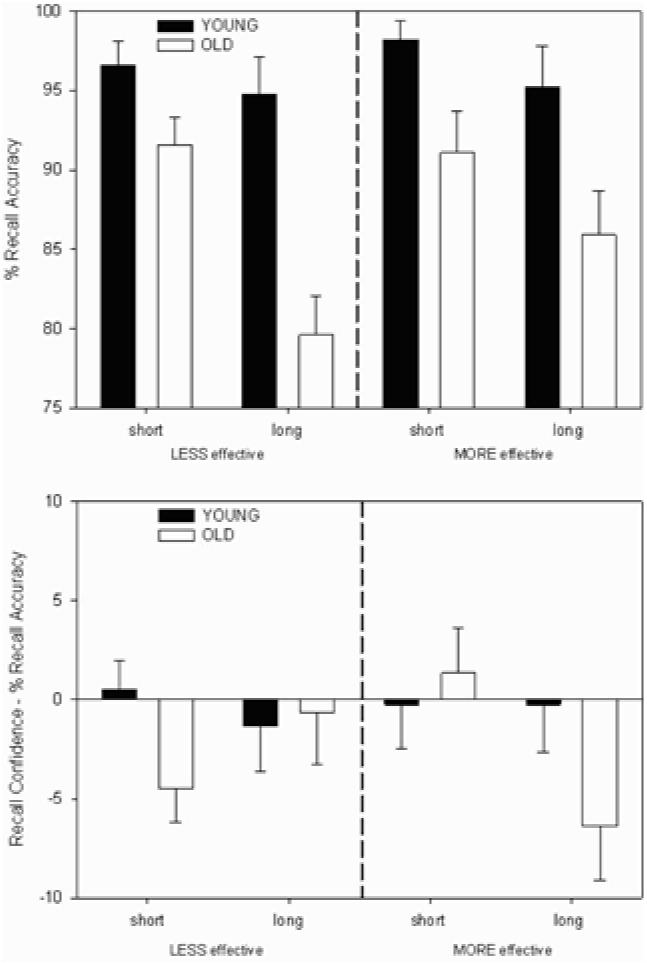


Figure 2. Experiment 2 letter recall performance (top) and absolute accuracy of confidence judgements (bottom) by age and list length (short or long) for normatively less effective strategies (left) and normatively more effective strategies (right).

For recall performance, in addition to the age difference described above, we found significant main effects of strategy effectiveness and list length as well as a significant three-way interaction. Participants were more accurate when using normatively effective strategies, $\eta_p^2 = .05$, $F(1, 46) = 3.93$, $p = .05$, and for short list lengths, $\eta_p^2 = .13$, $F(1, 94) = 15.09$, $p < .01$. Furthermore, for older but not younger adults, the list length difference was reduced when more effective strategies were used, $\eta_p^2 = .06$, $F(1, 73) = 4.15$, $p = .05$.

For the absolute accuracy of recall confidence judgements, the interaction of strategy effectiveness and length was significant, $\eta_p^2 = .03$, $F(1, 81) = 3.97$, $p = .05$, as was the three-way interaction with age, $\eta_p^2 = .04$, $F(1, 81) = 7.46$, $p < .01$. Older adults were more underconfident at long lengths than at short lengths when using normatively effective strategies, but more underconfident at short lengths when using less effective strategies; younger adult monitoring was not substantially impacted by strategy use and list length.

## Discussion

The outcomes from Experiment 2 replicated findings from Experiment 1; age differences were noted in WM span and monitoring accuracy, and WM performance was related to the absolute accuracy of metacognitive predictions. Experiment 2 also extended Experiment 1 to more directly consider the relationship between strategic behaviour and age differences in WM task performance. Comparable to findings in the associative memory literature as well as recent WM research (Bailey et al., 2009), older adults were as likely as young adults to report the use of normatively effective memory strategies for the operation span task.

Both younger and older adult participants were more likely to use normatively effective strategies to recall long lists than to recall short lists. Although younger adult performance was near ceiling, complicating interpretation, it appears that the use of effective strategies was particularly beneficial to older adults for long recall lengths. However, whereas younger adults' monitoring accuracy did not vary by strategy report, older adults showed pronounced underconfidence in long list recall when using effective strategies. This might suggest that older adults did not fully appreciate the benefits of effective strategy use or their ability to implement them, although the relationship might also be considered adaptive if low confidence leads older adults to adopt an effective strategy. Effective strategy use improved older adults' WM performance. Furthermore, mediation analyses indicated that participants with more accurate monitoring judgements also produced superior WM recall. These outcomes are consistent with the possibility that monitoring is an adaptive means for individuals to select appropriate task strategies and bolster performance in complex cognitive tasks. As in Experiment 1, absolute prediction accuracy mediated age differences in WM performance. Note that the mediation was partial rather than full in this Experiment, and that the causal possibility remains that, alternatively, it is age differences in WM that influence age differences in monitoring accuracy.

It is important to note that older adults' equivalent reporting of strategies in the operation span task does not indicate whether strategies are as effectively used by older adults. Should older adults use less effective strategic mediators (a utilization deficit), or forget attempted mediators more often (a retrieval deficit), simple strategy reporting would not reflect important age differences. Older adults have equivalent mediator quality but impaired mediator recall for an associative learning task (Dunlosky, Hertzog, & Powell-Moman, 2005), and the same variables might influence strategic behaviour in the operation span task as well.

Finally, it is notable that use of effective strategies was unrelated to processing latency (as measured by equation RTs) in the current study ($r = -.05$, $p = .64$), given that earlier suggestions of a link between performance and strategic behaviour were in part driven by the finding that individuals with long latencies (thought to be a reflection of strategy use) show high performance. Individuals might use normatively less effective strategies for long periods and still reap benefits. The finding of no relationship between latency and strategy in the current study, coupled with having obtained a typical latency-performance relationship ($r = -.40$, $p < .01$), suggests that both the effectiveness and duration of strategy use influence WM performance.

## GENERAL DISCUSSION

The present research supports the position that metacognitive monitoring and strategic behaviour can influence WM task performance. We reinforce findings that more effective encoding strategies lead to higher accuracy in a WM span task (e.g., Dunlosky & Kane, 2007; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003). We demonstrate that WM recall is better for individuals who more accurately monitor performance, perhaps supporting such strategic metacognitive control.

It is particularly notable that current findings of adult age differences in monitoring for a complex cognitive task contrast with typical findings for simple memory tasks (see Dunlosky & Hertzog, 2000). Older adults were poorer in WM monitoring in terms of relative monitoring accuracy and absolute monitoring accuracy. Comparison of gamma correlations, the standard index of metacognitive monitoring ability, indicated that older adults were considerably less capable of distinguishing between high and low levels of recall performance in the WM task than were younger adults, especially early in practice. Although absolute monitoring accuracy generally showed underconfidence by older adults, age differences were magnified when examined as absolute values, indicating that older adults' judgements are also less well calibrated in a more general sense. This suggests that while mean older adult monitoring was underconfident, some older adults were miscalibrated in the direction of overconfidence, and older adults were overall more likely to show errant monitoring.

Monitoring deficits appear to substantially influence performance of the operation span task, consistent with a monitoring deficit hypothesis. Although age differences in strategic approach are well documented in various other task domains (see Hertzog & Dunlosky, 2005), it appears that metacognitive monitoring deficits might have a more profound influence on cognitive ageing than previously believed. It appears that deficits in metacognitive monitoring account in part for age-related declines in WM performance. It is possible, for example, that underconfident prediction leads older adults to terminate item encoding prematurely, out of a sense that further encoding would be futile. The potential for flexibility in WM performance is particularly important given that age differences in WM appear to impact older adults' performance of many complex cognitive tasks (see Salthouse & Babcock, 1991). Various everyday cognitive tasks involve complex processing, such as balancing a cheque book, navigating an automated telephone menu, or driving in a new city. Future research should consider whether performance monitoring might improve older adults' performance of such fundamental life tasks.

Age differences in WM monitoring might have various underlying causes. Older adults might rely on different information in making metacognitive judgements, due either to selection biases or to availability. It is also possible that older adults' WM performance or monitoring performance is impacted by the build-up of proactive interference (e.g., Eakin & Hertzog, 2006; Lustig, May, & Hasher, 2001). Older adults' WM performance and monitoring accuracy might also be influenced by general beliefs about cognitive performance declines, or even by an implicit stereotype threat. It does not appear that declines in WM monitoring reflect broad cognitive ability, as correlations with cognitive pretest scores were generally nonsignificant.2

An alternate explanation of the obtained relationships between monitoring and WM performance is that individuals with greater WM capacities can devote more resources to performance monitoring. If task difficulty were responsible for the obtained age differences in monitoring accuracy, we should expect to see larger age differences in monitoring accuracy with large list lengths. Older adults' WM task performance was indeed more impacted by long recall lengths than was that of younger adults, but analogous age differences in the length effect did not occur for monitoring, arguing against a resource account. Future research should evaluate potential mechanisms for age-related declines in WM performance monitoring.

Much is also still unknown regarding the nature of WM strategy use by young and older adults. Although we did not obtain age differences in the proportional adoption of normatively effective strategies (consistent with a new study by Bailey et al., 2009), further research should examine whether older adults' behaviour within a given strategy is less optimal than that of young adults. Given the present outcomes, it appears possible that older adults' WM task performance might benefit from training that improves monitoring accuracy and strategy implementation.

The inclusion of explicit monitoring requirements impacted letter recall performance in terms of both speed and accuracy. Performance often suffers when tasks are switched or performed concurrently (see Pashler, 1998). Most generally, this outcome suggests that when participants engage in spontaneous performance monitoring during WM task performance, it might differ markedly from the explicit monitoring performed here. It is noteworthy, however, that the monitoring variable did not interact with age or block for task performance in Experiment 1. Although performance monitoring can be particularly detrimental to older adults' learning in a sentence-reading task (Stine-Morrow et al., 2006), monitoring did not differentially impair older adults' WM performance.

To summarize, two experiments compared WM performance, metacognitive monitoring, and reports of strategic behaviour by younger and older adults. Results support previous work demonstrating that WM performance can be influenced by strategy use and implicate metacognitive monitoring as a possible influence on strategy recognition and adoption. Outcomes also demonstrate age deficits in WM monitoring that mediate age differences in recall performance in the operation span task. It appears that metacognitive monitoring, while age equivalent for simple memory tasks, declines with age for more complex cognition and has functional consequences for complex task performance.

## Acknowledgments

This research was supported by a grant from the Appalachian State University Research Council, as well as by a grant from the National Institute on Aging, one of the National Institutes of Health (R01 AG0248).

## REFERENCES

- 1. Baddeley, A. D. (1986) *Working memory* Clarendon Press/Oxford University Press , Oxford, UK
- 2. Bailey, H. , Dunlosky, J. and Hertzog, C. (2009) Does differential strategy use account for age-related deficits in working-memory performance?. *Psychology & Aging* **24** , pp. 82-92.
- 3. Baron, R. M. and Kenny, D. A. (1986) The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology* **51** , pp. 1173-1182.
- 4. Benjamin, A. S. , Bjork, R. A. and Schwartz, B. L. (1998) The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General* **127** , pp. 55-68.
- 5. Bieman-Copland, S. and Charness, N. (1994) Memory knowledge and memory monitoring in adulthood. *Psychology and Aging* **9** , pp. 287-302.
- 6. Cerella, J. Birren, J. E. and Schaie, K. W. (eds) (1990) Aging and information-processing rate. *Handbook of the psychology of aging* pp. 201-221. 3, Academic Press , New York
- 7. Cherry, K. and Park, D. (1993) Individual differences and contextual variables influence spatial memory in younger and older adults. *Psychology and Aging* **8** , pp. 517-526.
- 8. Connor, L. T. , Dunlosky, J. and Hertzog, C. (1997) Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging* **12** , pp. 50-71.
- 9. Conway, A. R. A. , Kane, M. J. and Engle, R. W. (2003) Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* **7** , pp. 547-552.
- 10. Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brains Sciences* **24** , pp. 87-185.
- 11. Craik, F. I. M. (2002) Levels of processing: Past, present, and … future?. *Memory* **10** , pp. 305-318.
- 12. Craik, F. I. M. and Byrd, M. Craik, F. I. M. and Trehub, S. E. (eds) (1982) Aging and cognitive deficits: The role of attentional resources. *Aging and cognitive processes* pp. 191-211. Plenum Press , New York
- 13. Daneman, M. and Carpenter, P. A. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* **19** , pp. 450-466.
- 14. Dunlosky, J. and Connor, L. T. (1997) Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition* **25** , pp. 691-700.
- 15. Dunlosky, J. and Hertzog, C. (2000) Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging* **15** , pp. 462-474.
- 16. Dunlosky, J. and Hertzog, C. (2001) Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition* **29** , pp. 247-253.
- 17. Dunlosky, J. , Hertzog, C. and Powell-Moman, A. (2005) The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology* **41** , pp. 389-400.
- 18. Dunlosky, J. and Kane, M. J. (2007) The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology* **60**:9 , pp. 1227-1245.

- 19. Dunlosky, J. , Kubat-Silman, A. K. and Hertzog, C. (2003) Effects of aging on the magnitude and accuracy of quality-of-encoding judgments. *American Journal of Psychology* **116** , pp. 431-454.
- 20. Eakin, D. K. and Hertzog, C. (2006) Release from implicit interference in memory and metamemory: Older adults know that they can't let go. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences* **61**:6 , pp. P340-P347.
- 21. Ekstrom, R. B. , French, J. W. and Harman, H. H. (1976) *Kit of factor-referenced cognitive tests* Educational Testing Service , Princeton, NJ
- 22. Engle, R. W. , Cantor, J. and Carullo, J. J. (1992) Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **18** , pp. 972-992.
- 23. Friedman, N. P. and Miyake, A. (2004) The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General* **133** , pp. 101-135.
- 24. Frieske, D. A. and Park, D. C. (1993) Effects of organization and working memory on age differences in memory for scene information. *Experimental Aging Research* **19** , pp. 321-332.
- 25. Hasher, L. and Zacks, R. T. Bower, G. H. (ed) (1988) Working memory, comprehension, and aging: A review and a new view. *The psychology of learning and motivation* **22** , pp. 193-225. Academic Press , New York
- 26. Hertzog, C. and Dixon, R. A. Metcalfe, J. and Shimamura, A. P. (eds) (1994) Metacognitive development in adulthood and old age. *Metacognition: Knowing about knowing* pp. 227-251. The MIT Press , Cambridge, MA
- 27. Hertzog, C. and Dunlosky, J. Ross, B. H. (ed) (2005) Aging, metacognition, and cognitive control. *Psychology of learning and motivation* pp. 215-251. Academic Press , San Diego, CA
- 28. Hertzog, C. and Hultsch, D. F. Salthouse, T. A. and Craik, F. I. M. (eds) (2000) Metacognition in adulthood and old age. *The handbook of aging and cognition* pp. 417-466. 2, Lawrence Erlbaum Associates , Mahwah, NJ
- 29. Hertzog, C. , Vernon, M. C. and Rympa, B. (1993) Age differences in mental rotation task performance: The influence of speed/accuracy tradeoffs. *Journal of Gerontology: Psychological Sciences* **48B** , pp. 150-156.
- 30. Kausler, D. H. (1994) *Learning and memory in normal aging* Academic Press , New York
- 31. Kelley, C. M. and Sahakyan, L. (2003) Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language* **48** , pp. 704-721.
- 32. Lemaire, P. , Arnaud, L. and Lecacheur, M. (2004) Adults' age-related differences in adaptivity of strategy choices: Evidence from computational estimation. *Psychology and Aging* **19** , pp. 467-481.
- 33. Light, L. L. Bjork, E. J. and Bjork, R. A. (eds) (1996) Memory and aging. *Memory* pp. 443-490. Academic Press , San Diego, CA
- 34. Littell, R. C. , Milliken, G. A. , Stroup, W. W. and Wolfinger, R. D. (1996) *SAS system for mixed models* SAS Institute , Cary, NC
- 35. Lustig, C. , May, C. P. and Hasher, L. (2001) Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General* **130**:2 , pp. 199-207.
- 36. McElree, B. (2001) Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27** , pp. 817-835.
- 37. McNamara, D. S. and Scott, J. L. (2001) Working memory capacity and strategy use. *Memory & Cognition* **29** , pp. 10-17.
- 38. Miyake, A. , Friedman, N. P. , Rettinger, D. A. , Shah, P. and Hegarty, M. (2001) How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General* **130** , pp. 621-640.
- 39. Miyake, A. and Shah, P. (1999) *Models of working memory: Mechanisms of active maintenance and executive control* Cambridge University Press , New York
- 40. Nelson, T. O. (1984) A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin* **95** , pp. 109-133.
- 41. Nelson, T. O. , Dunlosky, J. , Graf, A. and Narens, L. (1994) Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science* **5** , pp. 207-213.

- 42. Nelson, T. O. and Narens, L. Bower, G. H. (ed) (1990) Metamemory: A theoretical framework and some new findings. *The psychology of learning and motivation* **26** , pp. 125-173. Academic Press , New York
- 43. Park, D. C. , Hertzog, C. , Kidder, D. P. , Morrell, R. W. and Mayhorn, C. B. (1997) Effect of age on event-based and time-based prospective memory. *Psychology and Aging* **12** , pp. 314-327.
- 44. Park, D. C. , Smith, A. D. , Lautenschlager, G. , Earles, J. L. , Frieske, D. Zwahr, M. et al. (1996) Mediators in long-term memory performance across the life span. *Psychology and Aging* **11** , pp. 621-637.
- 45. Pashler, H. E. (1998) *The psychology of attention* MIT Press , Cambridge, MA
- 46. Phillips, L. H. , Gilhooly, K. J. , Logie, R. H. , Della Salla, S. and Wynn, V. E. (2003) Age, working memory, and the Tower of London task. *European Journal of Cognitive Psychology* **15** , pp. 291-312.
- 47. Saczynski, J. S. , Willis, S. L. and Schaie, K. W. (2002) Strategy use in reasoning training with older adults. *Aging, Neuropsychology and Cognition* **9** , pp. 48-60.
- 48. Salthouse, T. A. (1996) The processing-speed theory of adult age differences in cognition. *Psychological Review* **103** , pp. 403-428.
- 49. Salthouse, T. A. and Babcock, R. L. (1991) Decomposing adult age differences in working memory. *Developmental Psychology* **27** , pp. 763-776.
- 50. Stine-Morrow, E. A. L. , Shake, M. C. , Miles, J. R. and Noh, S. R. (2006) Adult age differences in the effects of goals on self-regulated sentence processing. *Psychology and Aging* **21** , pp. 790-803.
- 51. Thapar, A. , Ratcliff, R. and McKoon, G. (2003) A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging* **18** , pp. 415-429.
- 52. Turley-Ames, K. and Whitfield, M. M. (2003) Strategy training and working memory task performance. *Journal of Memory and Language* **49** , pp. 446-448.
- 53. Turner, M. L. and Engle, R. W. (1989) Is working memory capacity task dependent?. *Journal of Memory and Language* **28** , pp. 127-154.
- 54. Verhaeghen, P. and Basak, C. (2005) Aging and switching the focus of attention in working memory: Results from a modified n-back task. *Quarterly Journal of Experimental Psychology* **58A** , pp. 134-154.
- 55. Verhaeghen, P. , Cerella, J. and Basak, C. (2004) A working memory workout: How to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **30** , pp. 1322-1337.
- 56. Verhaeghen, P. , Kliegl, R. and Mayr, U. (1997) Sequential and coordinative complexity in time-accuracy functions for mental arithmetic. *Psychology and Aging* **12** , pp. 555-564.
- 57. Verhaeghen, P. and Marcoen, A. (1996) On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging* **11** , pp. 164-178.
- 58. Wechsler, D. (1981) *Wechsler Adult Intelligence Scale-Revised* Psychological Corporation , New York
- 59. Zachary, R. (1986) *Shipley Institute of Living Scale Revised manual* Western Psychological Services , Los Angeles
- 60. Zwahr, M. D. , Park, D. C. and Shifren, K. (1999) Judgments about estrogen replacement therapy: The role of age, cognitive abilities, and beliefs. *Psychology and Aging* **14** , pp. 179-191.

## Notes

[1] In Experiment 1, equation reaction times (RTs) varied by age ($M_{\text{young}}$ = 4,420 ms, $M_{\text{old}}$ = 5,333 ms), $\eta_p^2 = .30$, $F(1, 112) = 10.01$, $MSE = 9{,}066{,}955$, $p < .01$, and improved with experience (e.g., $M_{\text{Block1}}$ = 5,405 ms, $M_{\text{Block4}}$ = 4,444 ms), $\eta_p^2 = .21$, $F(3, 336) = 42.23$, $MSE = 447{,}552$, $p < .01$, but block did not interact with age group. Equation accuracy also varied by age ($M_{\text{young}} = 96.9$, $M_{\text{old}} = 95.0$), $\eta_p^2 = .15$, $F(1, 112) = 5.64$, $MSE = 76$, $p = .02$, with no main effect of block, and an age by block interaction that indicated improvement by younger but not by older adults (e.g., $M_{\text{young,Block1-Block4}} = 0.83$, $M_{\text{old,Block1-Block4}} = -1.28$), $\eta_p^2 = .03$, $F(3, 336) = 2.89$, $MSE = 8.8$, $p = .04$. No comparisons with the condition variable were significant. In Experiment 2, younger adults

again responded to equations more rapidly ($M_{\text{young}}$ = 5,091 ms, $M_{\text{old}}$ = 6,143 ms), $\eta_p^2 = .07$, $F(1, 80) = 5.8$, $MSE = 3,859,968$, $p = .02$, and were also again more accurate in equation responses ($M_{\text{young}}$ = 96.4, $M_{\text{old}}$ = 94.2), $\eta_p^2 = .06$, $F(1, 80) = 4.7$, $MSE = 22$, $p = .03$.

[2] We examined correlations between older adults' education and cognitive ability pretests (see Table 1) with monitoring accuracy (absolute accuracy of predictions, postdictions and CJs, as well as relative accuracy of CJs) and strategy use. Education was not related to any outcome measure (*p*s > .24). We also did not obtain relationships with any outcome variables for the following abilities: vocabulary (*p*s > .08), perceptual speed (*p*s > .15), and incidental memory (*p*s > .24). Older adults with higher associative memory scores had better absolute accuracy in postdictions ($r = -.42$, $p < .05$), and there was a marginal relationship between postdiction accuracy and vocabulary ($r = -.31$, $p = .08$). None of the analogous relationships were significant for younger adults.