Practice makes imperfect: Working memory training can harm recognition memory performance

Laura E. Matzen¹ · Michael C. Trumbo^{1,2} · Michael J. Haass¹ · Michael A. Hunter² · Austin Silva¹ · Susan M. Stevens-Adams¹ · Michael F. Bunting³ · Polly O'Rourke³

Published online: 5 July 2016

© Psychonomic Society, Inc. (outside the USA) 2016

Abstract There is a great deal of debate concerning the benefits of working memory (WM) training and whether that training can transfer to other tasks. Although a consistent finding is that WM training programs elicit a short-term near-transfer effect (i.e., improvement in WM skills), results are inconsistent when considering persistence of such improvement and far transfer effects. In this study, we compared three groups of participants: a group that received WM training, a group that received training on how to use a mental imagery memory strategy, and a control group that received no training. Although the WM training group improved on the trained task, their posttraining performance on nontrained WM tasks did not differ from that of the other two groups. In addition, although the imagery training group's performance on a recognition memory task increased after training, the WM training group's performance on the task *decreased* after training. Participants' descriptions of the strategies they used to remember the studied items indicated that WM training may lead people to adopt memory strategies that are less effective for other types of memory tasks. These results indicate that WM

training may have unintended consequences for other types of memory performance.

Keywords Working memory training · Mental imagery · Recognition memory · Memory strategies

Working memory (WM) refers to the brain system used for storage and manipulation of transitory information necessary for complex tasks such as learning, reasoning, and language comprehension (Becker & Morris, 1999). Recent research has indicated that WM training, where people repeatedly practice increasingly difficult WM tasks, can improve both WM capacity and other aspects of cognitive performance (cf. Jaeggi, Buschkuhl, Jonides, & Perrig, 2008). Improved performance on a trained task should bolster performance for additional domains and tasks to the extent that they rely on overlapping cognitive abilities or share neural systems (Dahlin, Neely, Larsson, Backman, & Nyberg, 2008). Because WM has been identified as a central component of general cognition (Engle, Tuholski, Laughlin, & Conway, 1999), the conjecture is that improvement on the trained WM task will result not only in near transfer (improvement on other WM tasks, such as a spatial WM task following training of a verbal WM task, indicative of heightened WM capacity) but also in far transfer (improvements in other domains, such as fluid intelligence tests).

Fluid intelligence refers to those aspects of intelligence that allow for adaptive reasoning and problem solving (Carpenter, Just, & Shell, 1990), and is a construct predictive of academic achievement (Rohde & Thompson, 2007). Previous research has demonstrated that WM capacity and fluid intelligence are strongly related constructs, sharing approximately 50% of

Electronic supplementary material The online version of this article (doi:10.3758/s13421-016-0629-4) contains supplementary material, which is available to authorized users.

✉ Laura E. Matzen
lematze@sandia.gov

¹ Sandia National Laboratories, Albuquerque, NM, USA

² University of New Mexico, Albuquerque, NM, USA

³ University of Maryland Center for Advanced Study of Language, College Park, MD, USA

their variance (Kane, Hambrick, & Conway, 2005). Fluid intelligence was largely thought to be immutable, but a study by Jaeggi and colleagues (2008) showed that WM training could improve fluid intelligence performance. This finding is significant from a theoretical as well as a practical perspective, and many additional studies have investigated the benefits of WM training for fluid intelligence and other types of cognitive processes. These studies have shown that WM training can improve episodic memory (Rudebeck, Bor, Ormond, O'Reilly, & Lee, 2012), attention (Chein & Morrison, 2010), and can provide general cognitive enhancement for children (Alloway, Bibile, & Lau, 2013). Additionally, WM training has been proposed as a remediating intervention for populations such as adults with amnesic mild cognitive impairment (Carretti, Borella, Fostinelli, & Zavagnin, 2013) and children with dyslexia (Luo, Wang, Wu, Zhu, & Zhang, 2013) or ADHD (Holmes et al., 2010).

However, other research paints a less optimistic view of the benefits of WM training. The idea that WM training can improve fluid intelligence runs contrary to more than a century of research on cognitive training within psychological and educational science. Numerous studies have demonstrated that although task-specific performance commonly increases with training, transfer of this learning to other tasks or domains is rare (Chase & Ericsson, 1981; Ericsson & Delaney, 1998; Healy, Wohldmann, Sutton, & Bourne, 2006; Singley & Anderson, 1989; Thorndike & Woodworth, 1901). Several recent studies of WM training have failed to find near or far task transfer (Chooi & Thompson, 2012; Redick et al., 2013; Thompson et al., 2013), and a recent meta-analysis concluded that while WM training consistently produces near-transfer effects, these effects tend to be short-lived, and improvement fails to generalize to other domains (Melby-Lervåg & Hulme, 2013). Furthermore, researchers have cited a number of methodological concerns within the WM training field, including use of single tasks to define WM change, inconsistent use of valid WM tasks, comparison of trained groups to a no-contact control group, and subjective measurements of change (Shipstead, Redick, & Engle, 2012).

Given the mixed results regarding the transfer of WM training to other tasks and the ability of WM training to improve fluid intelligence, it seems prudent to approach this topic with a dose of skepticism. In this study, we sought to examine the effects of WM training on both untrained WM tasks and on a verbal recognition memory task. Although WM training has been touted as improving academic success, there has been little research on how WM training impacts other types of memory that are also crucial for learning, such as recognition memory. In one of the few studies to test the effects of WM training on recognition memory, Rudebeck and colleagues (2012) showed that spatial WM training can benefit performance on visual recognition memory tasks. However, it remains unclear whether WM training can

improve verbal recognition memory, a type of memory that is particularly important in most educational settings.

With this study, we also sought to address some of the methodological issues that have made it difficult to interpret the results of prior WM training studies (cf. Redick et al., 2013; Shipstead, Redick, & Engle, 2010). Specifically, this study used both a no-contact control group and an active control group in which participants were trained to use mental imagery as a memory strategy. Participants were assigned semirandomly to the three experimental groups (the groups were balanced by gender and age). We selected mental imagery training as the active control condition because mental imagery is long-established as an effective technique for improving recognition memory performance (Paivio, 1971; Prestianni & Zacks, 1974). In theory, both training techniques should improve recognition memory performance, but it was unclear how they would stack up against one another in practice. The two types of training are fundamentally different in terms of the role of memory strategies in the training tasks. Using mental imagery as a memory aid is clearly a strategy, whereas there is debate about the role of strategy in WM tasks, which are intended to be process training, rather than strategy training (Lövdén, Bäckman, Lindenberger, Schaefer, & Schmiedek, 2010). Although several studies have found that strategy use improves performance on WM tasks (McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003), the fundamental goal of WM training is to enhance a basic cognitive ability that can translate to improved performance in tasks that were not trained. Thus, adaptive WM training regimens intentionally discourage participants from developing task-specific memory strategies (e.g., Jaeggi et al., 2008). In this study, we sought to explore hypotheses about transfer of training from strategy-based versus process-based memory training regimens.

Task selection and hypotheses

All participants completed the same battery of memory tasks before and after training. The battery included a verbal WM task (listening span), a spatial WM task (rotation span), and a verbal recognition memory task. Participants were semirandomly assigned to one of three memory training groups. The participants in the control group received no memory training. The participants placed in the WM training group completed a series of training sessions that consisted of an adaptive *n*-back task and an adaptive symmetry span task. These tasks are similar to those that have been used in prior WM training studies (cf. Melby-Lervåg & Hulme, 2013) and were intended to target both verbal and spatial WM. In accordance with past research and with commercial memory training programs, the tasks adapted their difficulty based on the participant's performance.

The participants in the mental imagery training group were trained to create vivid mental images as an aid for memorizing lists of words. They were shown examples using both concrete and abstract words, and practiced using a mental imagery strategy on a series of short recall tasks. The imagery training was intended to improve participants' memory strategies by teaching them to associate the to-be-remembered items with concrete, vivid, and meaningful or bizarre mental images, all qualities that should improve subsequent memory performance (Baddeley & Andrade, 2000; Nelson & Schreiber, 1992; Paivio, 1965; Paivio, Walsh, & Bons, 1994; West & Holcomb, 2000).

We hypothesized that the participants in the control group would not exhibit any significant differences in performance between the pre- and posttraining baseline tasks. For the participants in the imagery training group, we expected that using a mental imagery memory strategy would lead to a general improvement on the recognition memory task after training. In addition, because strategy use has been shown to improve performance on WM tasks (McNamara & Scott, 2001; St. Clair-Thompson, Stevens, Hunt, & Bolder, 2010; Turley-Ames & Whitfield, 2003), we hypothesized that the imagery training group's performance on the WM baseline tasks would also improve after training. Because the imagery training involved memorizing lists of words, the recognition memory task can be thought of as a near-transfer task and the WM baseline tasks as far-transfer tasks for participants in the imagery training group. For the WM training group, we expected to see near transfer in which participants improved their performance on the WM baseline tasks after training. We also predicted that there would be far transfer of the WM training to the recognition memory task. Specifically, if WM training improves WM capacity, we would expect to see higher performance for the once-presented and repeated words, which may be encoded better after spending more time in WM (Braun & Rubin, 1998).

The recognition memory task included several conditions that allowed us to investigate the impact of WM training on repetition effects, spacing effects, and testing effects. Analyses of these effects before and after training are presented in the [Supplemental Materials](#).

Method

Each participant in the experiment completed tasks over the course of a 5-week period. During the first week, participants completed pretraining baseline memory tasks that included a verbal WM task (listening span), a spatial WM task (rotation span), and a verbal recognition memory task. During the next 3 weeks, participants completed memory training sessions that differed based on the training group to which they were assigned. Participants assigned to the mental imagery training

group completed three training sessions (one per week), and participants in the WM training group completed 14 training sessions (four to five per week) during the 3-week training period. Participants assigned to the control group did not complete any tasks during the training period. At the end of the training period, all participants completed the same baseline tasks for a second time. Each of the baseline and training tasks is described in detail.

Participants

Eighty-six participants recruited from the employee population of Sandia National Laboratories participated in this experiment and were paid for their time. All were right-handed, had no early exposure to languages other than English, and had no history of neurological disease or defect. Participants were assigned semirandomly to one of the three training groups (each participant's age and gender were taken into account in group assignment to keep the groups as demographically balanced as possible). Eight participants dropped out of the study before completing all of the sessions (one from the control group, four from the mental imagery group, and three from the WM training group), and four additional participants (two each from the control group and WM training group) failed to follow instructions and were excluded from the data analysis. Of the remaining 74 participants, 25 (12 female) were in the control group, 24 (10 female) were in the imagery training group, and 25 (13 female) were in the WM training group. The mean age for all of the participants was 37 years (range 18–63 years). The mean ages for each group were 37 years for the control group (range 18–61 years), 39 years for the imagery training group (range 18–63 years), and 35 years for the WM training group (range 20–63 years). Figure S1 in the Supplemental Materials shows the distributions of age and educational background for the participants in this study.

Baseline tasks

Listening span task

Based on Daneman and Blennerhassett (1984), the listening span task required participants to recall a sequence of symbols in the order in which they were presented. The presentation of the symbols was interleaved with the auditory presentation of sentences. Participants had to indicate whether the sentences made sense or not. Participants practiced the two tasks separately, and then performed both in the dual-task phase.

Materials Materials for the memory task were nine black Wingdings symbols in size 24 font presented against a white background:

The secondary task was comprised of 110 spoken sentences, half of which were sensible (“They gave the waiter a tip even though he was rude”), and half of which were not (“The children were summer and wanted their parents to come home”). The same speaker was used for all sentence recordings.

Procedure The listening span task had three phases, the first of which was a memory task involving the sequences of symbols. Participants saw a series of symbols that were presented for 1,000 ms each in the center of a computer screen. They were then shown a recall screen that displayed all of the symbols and were asked to select the symbols in the order in which they had appeared. Participants could edit their selections and insert “blanks” into the sequence in place of symbols that they could not recall. Following each response screen, participants saw feedback on their performance (i.e., “You recalled X of Y items correctly”).

The second task required participants to listen to sentences and judge their sensibility. During sentence presentation, participants received instructions to click the mouse once they could tell whether the sentence made sense or not, at which point the sentence stopped playing and they responded by clicking either the yes or no button on the screen. Participants received feedback on their accuracy after each response.

During the dual-task phase, participants saw a new symbol after judging each sentence. The recall screen appeared after a sequence of four to eight symbols had accrued. Each participant saw two sequences of four and five symbols, and three sequences of six, seven, and eight symbols. The different sequence lengths were randomly ordered for each participant and for each session (pre- and posttraining).

Rotation span task

Based on Shah and Miyake (1996), this task required participants to recall sequences of arrows of varying length and orientation. The presentation of the arrows was interleaved with the presentation of letter characters. Participants had to make a judgment as to whether the letters appeared normally or backward. After each block, participants were asked to recall the sequence of arrows. As in the listening span task, participants practiced the two tasks separately, and then performed both in the dual task phase.

Materials For the memory task, the item set was comprised of pictures of long and short arrows at eight orientations, (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). The secondary task used five letters, (R, L, J, G, and F), at eight different orientations, (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), with both normal and backward versions (flipped along the vertical axis). All of the stimuli were white and were presented on a black background.

Procedure On each trial, participants saw a sequence of rotated letters, each of which was followed by an arrow. For each letter, the participants had to press a key on the keyboard to indicate whether the letter was presented normally or backward. The letter remained on the screen until the participant made a response. Then an arrow was presented for 1,000 ms. The trials varied in length and contained between two and five arrows. After the last arrow was presented, participants were asked to recall the sequence of arrows they had seen in that trial. The recall screen showed all 16 possible arrows (long or short arrows at each of eight orientations). Participants clicked on the arrows to indicate which arrows had appeared in the previous sequence, in the order they appeared. As in the listening span task, participants could edit their response and insert blanks in the place of arrows they had forgotten. When the participants completed their responses for this trial, they clicked the “next” button to advance to the next trial.

Recognition memory task

In the recognition memory task, participants were shown a list of common English nouns and were asked to memorize them for a subsequent recognition test. The task was designed to include several conditions with varying levels of difficulty. There were words that were studied only once, words that were repeated at short (one intervening item) and long (nine intervening items) lags, and words that were quizzed at short or long lags within the study block. We expected that the testing effect (Karpicke & Roediger, 2008) would make the quizzed words easiest to remember on the subsequent memory test, whereas the words that were repeated but not quizzed would be more difficult to remember. For the repeated words, we expected that the spacing effect (Melton, 1970) would lead to better performance for the words that were repeated after a longer lag. We expected the poorest memory performance for the words that were studied only once.

Materials The recognition memory task used a list of 1,344 words, all of which were common English nouns. The average length of the nouns was five letters, and their average frequency was 55.67 (based on the Kucera & Francis, 1967, norms included in Balota et al., 2002). The words were assigned to counterbalanced experimental lists such that every word appeared in every study and test condition across lists.

The experimental lists were divided into six study-test blocks with equal numbers of each item type in each block. The words were placed in a pseudorandom order within the blocks such that no more than three items in the same condition appeared in sequence. Within each study-test block, there were 28 words that were studied once, 14 words that were studied twice with a short lag between repetitions, 14 words that were studied twice with a long lag between repetitions, 14 words that were studied and then quizzed after a short lag, and 14 words that were studied and then quizzed after a long lag. In addition to the studied items, there were 28 words that served as new items for the quizzes within the study blocks (these words were quizzed but had not been studied) and 84 words that served as new, unstudied items in the subsequent recognition test. In total, each study block contained 112 study words (including repeated study words) and 56 quizzed words. Each test block contained 168 test words, half of which had been studied and half of which were new.

Three of the study-test blocks were presented to each participant during the pretraining baseline session, and the other three were presented during the posttraining session. The placement of the blocks (pre or posttraining) was counterbalanced across participants.

Procedure The participants were instructed that they would be tested on their memory for a list of study words. Throughout each task, a fixation cross was shown in the center of the screen. Prior to the presentation of each word, a yellow or red dot appeared on the screen immediately above the fixation cross. Participants were instructed that the yellow dot indicated that the next word was a study word and that they should silently read that word and try to remember it for later. They were told that the red dot indicated that the next word was a quiz word, and that following the word they should press a button to indicate whether or not they had studied that word earlier in the session. The study or quiz word was presented 600–800 ms after the dot disappeared and remained on the screen for 1 second. The words were presented immediately above the fixation cross. If the word was a quiz word, it was followed by a red question mark that remained on the screen until the participant pressed a response button. Participants pressed one of two buttons on a game controller, labeled *yes* and *no*, to indicate whether they remembered studying that word. They did not receive feedback about the accuracy of their responses.

At the end of each study block, participants took a short break before beginning the test block. All of the words in the test block were presented in the same way as the quizzed items from the study block. Each test word was preceded by a red dot and followed by a question mark. While the question mark was on the screen, participants pressed the *yes* or *no* button to indicate whether they remembered studying that word during the study block. As in the study block, they did not receive

feedback about the accuracy of their responses. It took participants approximately 10 minutes to complete each study block and 12 minutes to complete the corresponding test block.

Training tasks

Mental imagery training

In the three weeks in between the pretraining and posttraining baseline sessions, 24 of the participants completed three memory training sessions in which they practiced using a mental imagery strategy to remember word lists for a free recall test. The training sessions became more difficult as the participants progressed by using longer word lists, shorter encoding times, and more words with low imaginability.

Materials The memory tests used in the mental imagery training consisted of 168 nouns. Care was taken to ensure that none of the words used in the training sessions appeared in any of the pretraining or posttraining baseline tasks. Of the 168 nouns, 49 had low imaginability (ratings below 400 in the norms included in the MRC Psycholinguistic Database; Wilson, 1988) and the remainder had high imaginability (ratings above 550).

Procedure The mental imagery training consisted of three sessions, and participants were asked to complete a session once a week for 3 weeks, for a total of 90 minutes of training (assuming completion of all three sessions). Each session took approximately half an hour to complete. During the training, the participants were then given examples of mental imagery. The examples, which included both concrete and abstract concepts, explained that creating detailed and unusual mental images could be helpful for remembering information. After a short practice session in which participants were asked to generate and describe mental images for a short list of words, the training provided examples of grouping several mental images into one scene to increase their memorability. The participants were then asked to practice the mental imagery strategy by memorizing two lists of words, each of which was followed by a recall test. During the first practice list, the participants controlled how long the study words were presented. For the second practice list, each word was presented for 3 seconds. Each practice list contained 10 words, and participants had 10 chances to enter the words during the recall test. Participants received feedback after each entry. After the memory test, the participants were asked to describe the mental images that they had generated for the word list and to rate the effectiveness of the mental imagery strategy.

In the second and third training sessions, participants saw a brief review of the examples of mental imagery that were presented in the first session and were then asked to practice

the mental imagery strategy while completing memory tests with the same format as the tests used in the first training session. As the training progressed, the study lists became longer and included more words with low imaginability. The encoding time per word also decreased to 2 seconds. The structure of each study list is shown in Table S1 in the Supplemental Materials.

Working memory training

The 25 participants in the WM training group were trained on two tasks, the adaptive n -back task and the symmetry span task. Participants were loaned a laptop containing the two tasks and were asked to do each task once on every business day for 3 weeks. They were told that they could skip the training on one day of their choice, for a total of 14 sessions. Each training session lasted approximately 25 minutes, for a total of 350 minutes of training (assuming completion of all 14 sessions).

Adaptive n -back task In the adaptive n -back task, sequences of 25 single letters appeared one at a time on the screen, and participants were asked to indicate with a button press whether the current letter matched the letter that appeared n items earlier. For example, if subjects were shown the sequence A-B-C-B in a two-back task, they would indicate that the second “B” was a target because it matched the letter that had appeared two letters back. They would respond “nontarget” to the other items in the sequence. Within each sequence of 25 items, five were targets and up to five items were lures, whereas the rest of the items were nontargets (see also Sprenger et al., 2013). The lures were letters that matched a letter that had appeared in position $n - 1$ or $n + 1$ (Kane, Conway, Miura, & Colflesh, 2007). For example, in the sequence A-B-A-C-D in a three-back task, the second A is a lure (an $n - 1$ lure) because it repeats a letter that appeared recently, but it is not a correct match for the three-back task. There were three lure conditions that corresponded with three levels of difficulty (no lures, $n + 1$ lures only, and both $n + 1$ and $n - 1$ lures). The difficulty level of the task changed based on the participant’s performance. When participants achieved at least 85% accuracy for one sequence of 25 letters, difficulty was increased for the next sequence by increasing the lure difficulty level. Once participants achieved at least 85% accuracy for a given n at the highest level of lure difficulty, n was increased by one, and the level of lure difficulty was reset to the lowest level (no lures). If accuracy fell below 65%, task difficulty decreased for the next sequence, first by decreasing lure difficulty and then by decreasing n . Task difficulty, therefore, represented both the value of n and the lure difficulty level. Each training session consisted of nine sequences of letters and lasted for approximately 10 minutes. Performance was scored by computing the response accuracy for each level of n and the mean

level of n achieved during the training session. All participants started the first training session at a difficulty level of $n = 2$ with no lures.

Symmetry span task The symmetry span task required participants to remember the locations of a sequence of blocks that appeared in a 4×4 grid, in the order in which they were presented. The blocks were presented serially, for 650 ms each. The presentation of the blocks was interleaved with the presentation of a design on an 8×8 grid. Participants had to determine if the design was symmetrical across the vertical axis. At the end of a series of these presentations, participants saw a blank 4×4 grid and were asked to recall the positions and order of the to-be-remembered blocks by clicking on the grid. The score was the number of blocks recalled in the correct serial order. The difficulty of this task was adjusted by changing the number of blocks that the participant needed to remember. Performance was evaluated after four sets of memory responses. If the participant got three or more correct, the sequence length increased by one for the next set of four. Conversely, if performance fell below two correct, the sequence length decreased by one for the next set. All participants started with a sequence length of three blocks. Each training session lasted for 15 minutes, and the number of sets included in each session varied depending on the sequence length and how many sequences the participants were able to complete in 15 minutes.

Memory strategy survey

At the end of their participation in the study, the participants were asked to complete a follow-up questionnaire about their use of various memory strategies in the baseline memory tasks. Fifty-six of the participants completed the survey (18 from the control group, 19 from the mental imagery training group, and 19 from the WM training group). The first part of the survey asked participants to describe their memory strategy for each task. The second part gave examples of different memory strategies (mental imagery, generating sentences or stories, linking items to one another, rehearsal and self-quizzing) and participants were asked if they had used those strategies on the pre- and posttraining memory tasks.

Results

Mental imagery training

Twenty-three of the 24 participants completed all three of the imagery training sessions, and one participant completed only two of the training sessions. The average number of words recalled by the participants remained fairly consistent across the 14 memory tests used in the imagery training session, even

as the encoding task became more difficult (longer word lists, shorter encoding times, more abstract words). Figure 1 shows the average number of words recalled on each memory test. Participants recalled an average of 8.33 words per test list during the first training session, 7.42 words per list during the second training session, and 8.04 words per list during the third training session. A repeated-measures ANOVA showed that there was a significant effect of test number on the number of words recalled on each test, $F(13, 284) = 3.68$, $p < .001$, $\eta_p^2 = 0.14$. Paired t tests comparing average performance in each of the three training sessions showed that participants performed significantly worse during the second training session relative to the first session, $t(23) = 2.87$, $p < .01$, Hedges's $g_{av} = 0.63$, and significantly better during the third training session relative to the second training session, $t(22) = 2.67$, $p < .01$, Hedges's $g_{av} = 0.38$. As the memory tests became more difficult, participants reported that it became more difficult to create mental images for the word lists, and they felt that the imagery strategy was less effective for the more difficult lists. Additional analyses are presented in the [Supplemental Materials](#).

Working memory training

Twenty-four of the 25 participants in the WM training group completed at least 12 of the 14 WM training sessions, and one participant completed nine of the training sessions. The participants who completed at least 12 of the WM training sessions were included in the analysis of the WM training. On average, the participants' performance improved across the training sessions for both training tasks. During the first training session, the participants had an average n -back level of 1.81 and an average symmetry span difficulty level of 3.77. On the 12th training session, the participants' average n level was 4.23 and their average symmetry span difficulty level was 5.43. However, there was a great deal of variability across participants. The average n level achieved by each participant on the 12th training session ranged from 1.0 to 8.83. Similarly, the average level of difficulty achieved by each participant for the 12th session of the symmetry span task ranged from 3.18 to 7.33. Figure 2 shows the changes in performance across the WM training sessions. Repeated-measures ANOVAs showed that there were significant effects of training session on scores for both the symmetry span task, $F(11, 242) = 14.76$, $p < .001$, $\eta_p^2 = 0.40$, and the n -back task, $F(11, 231) = 32.74$, $p < .001$, $\eta_p^2 = 0.61$. Paired t tests comparing the first and 12th training sessions showed that participants scored significantly higher on the 12th session in both the symmetry span task, $t(22) = 8.56$, $p < .001$, Hedges's $g_{av} = 1.50$, and the n -back task, $t(21) = 7.64$, $p < .001$, Hedges's $g_{av} = 1.93$.

Baseline memory tasks

Although participants generally improved their performance on the tasks on which they were trained, the key question was whether their training would affect their performance on untrained memory tasks. To address this question, we compared the three training groups' changes in performance on the three pre- and posttraining baseline tasks. Given the large age range of the participants, the statistical tests were run with and without including age as a covariate. Including age did not change the results of the tests unless otherwise noted.

Listening span task Five participants (four from the control group and one from the WM training group) were excluded from the analysis of the listening span task due to a problem with the presentation of the sound files during the pretraining session. The average total scores from the remaining participants in each training group are shown in Table 1. A one-way ANOVA showed that the performance of the three groups did not differ significantly on the pretraining test, $F(2, 66) = 1.33$, $p = .27$, $\eta_p^2 = 0.04$.¹ Paired t tests were used to assess each group's change in performance between the pretraining and posttraining sessions. All three groups performed significantly better during the posttraining session, control group: $t(20) = 3.23$, $p < .01$, Hedges's $g_{av} = 0.57$; imagery training group: $t(23) = 3.36$, $p < .01$, Hedges's $g_{av} = 0.43$; WM training group: $t(23) = 1.87$, $p < .04$, Hedges's $g_{av} = 0.24$. However, a one-way ANOVA showed that the three groups were not significantly different in terms of how much their performance improved, $F(2, 66) = 1.69$, $p = .19$, $\eta_p^2 = 0.05$.

Rotation span task Two participants were excluded from the analysis of the rotation span task (one from the control group and one from the WM training group) due to failure to complete the posttraining test. The mean accuracy for the remaining participants in each training group are shown in Table 1. A one-way ANOVA showed that the performance of the three groups did not differ significantly on the pretraining test, $F(2, 69) = 0.26$, $p = .77$, $\eta_p^2 = 0.01$. Paired t tests were used to assess each group's change in performance between the pretraining and posttraining sessions. All three training groups performed significantly better during the posttraining session, control group: $t(23) = 3.79$, $p < .001$, Hedges's $g_{av} = 0.59$; imagery training group: $t(23) = 3.61$, $p < .001$, Hedges's $g_{av} = 0.41$; WM training group: $t(23) = 4.09$, $p < .001$, Hedges's $g_{av} = 0.56$. However, a one-way ANOVA comparing the change in performance across all three training groups showed that there were no significant differences between the groups, $F(2, 69) = 0.49$, $p = .61$, $\eta_p^2 = 0.01$.

¹ Effect sizes were calculated using the materials provided by Lakens (2013).

Fig. 1 Average number of words recalled on each memory test during the mental imagery training sessions

Recognition memory task The participants' recognition memory performance (average proportion correct and standard deviation for each condition) is shown in Table S2 in the Supplemental Materials. The bulk of our analysis focused on using d' as a measure of recognition memory performance. The hits and false alarm rates were used to calculate d' for each test condition, and the average d' values are shown in Table 2.

A 3×7 repeated-measures ANOVA (Training Group $\times$ Memory Test Condition) was used to compare the memory performance of the three groups before and after training. Prior to training, the d' scores of the three groups did not differ significantly, $F(2, 420) = 2.19$, $p = .12$, $\eta_p^2 = 0.19$. However, after training, there was a significant main effect of training group for the d' scores, $F(2, 420) = 6.77$, $p < .01$, $\eta_p^2 = 0.48$. One-way ANOVAs comparing the performance of the three groups on each condition of the posttraining recognition memory test showed that the performance of the groups differed significantly on all seven conditions (all F s > 4.08 , all p s $< .02$, all $\eta_p^2 > 0.10$). In other words, although the three groups of participants had similar recognition memory performance prior to training, their d' scores were significantly different for every condition after training. (See the Supplemental Materials for additional analysis of the participants' pretraining performance on the recognition memory test.)

The crucial comparison for examining the effects of the memory training techniques on recognition memory performance was the difference between pretraining and posttraining performance across the three training groups. For each test condition, one-way ANOVAs were used to compare the change in hit rates (posttraining hit rates minus pretraining hit rates) and the change in d' scores (posttraining d' minus pretraining d' , or Δ

d') across the three groups. The average d' difference scores are shown in Fig. 3. The changes in hit rates for the three groups were significantly different for every test condition where participants were tested on old words, once-presented words: $F(2, 70) = 3.50$, $p = .04$, $\eta_p^2 = 0.09$; short-lag repeated words: $F(2, 70) = 3.33$, $p = .04$, $\eta_p^2 = 0.09$; long-lag repeated words: $F(2, 70) = 3.60$, $p = .03$, $\eta_p^2 = 0.09$; short-lag quizzed words: $F(2, 70) = 4.23$, $p = .02$, $\eta_p^2 = 0.11$; long-lag quizzed words: $F(2, 70) = 8.12$, $p < .01$, $\eta_p^2 = 0.19$. However, the three groups did not differ in terms of the change to their false-alarm rates in response to new, unstudied words, $F(2, 70) = 0.22$. The changes in d' for the three groups were significantly different for long-lag quizzed words, $F(2, 70) = 5.67$, $p < .01$, $\eta_p^2 = 0.15$, and marginally significant for the short-lag repeated words, $F(2, 70) = 2.63$, $p = .07$, $\eta_p^2 = 0.07$, and the short-lag quizzed words, $F(2, 70) = 2.92$, $p = .06$, $\eta_p^2 = 0.08$.

Post hoc paired t tests were used to assess each group's change in recognition memory performance between the pre- and posttraining sessions. The control group's performance in the pretraining and posttraining sessions did not differ significantly for any of the test conditions, whether measured by hit rates (all t s < 1.03 , all p s $> .16$), false-alarm rates, $t(24) = 0.60$, or by d' scores (all t s < 0.44 , all p s $> .33$). The participants in the mental imagery training group had nearly identical false-alarm rates before and after training, $t(23) = 0.27$, but significantly better hit rates in the posttraining session for the once-presented words, $t(23) = 1.80$, $p = .04$, Hedges's $g_{av} = 0.39$, the short-lag repeated words, $t(23) = 2.40$, $p = .01$, Hedges's $g_{av} = 0.49$, and the short-lag quizzed words, $t(23) = 2.17$, $p = .02$, Hedges's $g_{av} = 0.35$. When measured by d' scores, the participants in the imagery training

Fig. 2 Individual and average working memory training results for the *n*-back and symmetry span tasks. Error bars for the averaged training results show mean standard error

group had significantly better performance after training for the words in the short-lag repetition condition, $t(23) = 1.74$, $p < .05$, Hedges's $g_{av} = 0.40$, but no significant differences in d' scores for any of the other test conditions, once-presented

Table 1 Means and (Standard Deviations) of Baseline WM Task Scores for Each Training Group

Baseline Task	Training Group	Pretraining Performance	Posttraining Performance	Change in Performance (Post–Pre)
Listening Span Task	Control Group	34.67 (14.10)	42.76 (13.88)	8.10 (11.38)*
	Mental Imagery	37.75 (13.61)	43.88 (14.48)	6.13 (8.92)*
	Working Memory	41.04 (11.63)	44.04 (13.09)	3.00 (7.88)*
Rotation Span Task	Control Group	0.60 (0.16)	0.70 (0.18)	0.10 (0.13)*
	Mental Imagery	0.57 (0.19)	0.64 (0.19)	0.08 (0.11)*
	Working Memory	0.57 (0.21)	0.68 (0.19)	0.12 (0.14)*

Statistically significant results are noted with an asterisk and bold font

Table 2 Mean and (Standard Deviation) d' Scores for Each Condition of the Recognition Memory Task for Each Training Group

	Condition	Training Group	Pretraining d'	Posttraining d'	Change in Performance (Post–Pre)
Words quizzed during study session	Short-lag quizzes	Control Group	2.89 (1.03)	3.35 (0.94)	0.46 (0.97)*
		Mental Imagery	3.51 (0.80)	3.84 (0.84)	0.33 (0.93)*
		Working Memory	3.24 (0.71)	3.15 (0.79)	-0.09 (0.71)
	Long-lag quizzes	Control Group	1.99 (0.88)	2.27 (0.89)	0.28 (0.79)*
		Mental Imagery	2.39 (1.05)	2.91 (1.21)	0.53 (1.25)*
		Working Memory	2.18 (0.69)	2.05 (0.87)	-0.13 (0.58)
	Once-studied words	Control Group	1.00 (0.55)	1.02 (0.66)	0.02 (0.47)
		Mental Imagery	1.27 (0.75)	1.51 (0.84)	0.24 (1.00)
		Working Memory	0.94 (0.46)	0.79 (0.60)	-0.15 (0.45)
Words tested during subsequent memory test	Short-lag repetitions	Control Group	1.33 (0.66)	1.35 (0.76)	0.02 (0.56)
		Mental Imagery	1.59 (0.76)	1.94 (0.94)	0.35 (0.98)*
		Working Memory	1.31 (0.46)	1.21 (0.52)	-0.09 (0.40)
	Long-lag repetitions	Control Group	1.35 (0.70)	1.38 (0.92)	0.03 (0.83)
		Mental Imagery	1.83 (0.82)	2.06 (1.03)	0.23 (1.18)
		Working Memory	1.41 (0.49)	1.16 (0.69)	-0.25 (0.50)*
	Short-lag quizzes	Control Group	1.47 (0.64)	1.43 (0.66)	-0.03 (0.56)
		Mental Imagery	1.70 (0.77)	1.96 (0.82)	0.26 (0.89)
		Working Memory	1.50 (0.54)	1.31 (0.56)	-0.19 (0.41)*
	Long-lag quizzes	Control Group	1.56 (0.65)	1.51 (0.80)	-0.05 (0.59)
		Mental Imagery	1.94 (0.83)	2.20 (0.95)	0.25 (0.94)
		Working Memory	1.86 (0.57)	1.43 (0.67)	-0.43 (0.52)*

Statistically significant results are noted with an asterisk and bold font

words: $t(23) = 1.19$, $p = .12$, Hedges's $g_{av} = 0.30$; long-lag repetitions: $t(23) = 0.94$, $p = .18$, Hedges's $g_{av} = 0.24$; short-lag quizzes: $t(23) = 1.43$, $p = .08$, Hedges's $g_{av} = 0.32$; long-lag quizzes: $t(23) = 1.33$, $p = .10$, Hedges's $g_{av} = 0.28$.

In contrast to the other two groups, the participants in the WM training group had significantly *lower* hit rates in the

posttraining session relative to the pretraining session for the long-lag repeated items, $t(24) = 2.81$, $p < .01$, Hedges's $g_{av} = 0.46$, the short-lag quizzed words, $t(24) = 2.26$, $p = .02$, Hedges's $g_{av} = 0.40$, and the long-lag quizzed words, $t(24) = 4.55$, $p < .01$, Hedges's $g_{av} = 0.72$. Their performance was marginally worse for the once-presented words, $t(24) = 1.65$, p

Fig. 3 Difference between posttraining memory performance and pretraining memory performance for each condition on the recognition memory test, compared across training groups. Error bars show mean standard error

= .06, Hedges's g_{av} = 0.21, but the difference was significant when controlling for age, p = .01. Their false alarm rates did not differ significantly before and after training, $t(23)$ = 0.52. The participants in the WM training group also had significantly worse d' scores after training for the long-lag repetition, short-lag quiz, and long-lag quiz conditions, long-lag repetitions: $t(23)$ = 2.48, p < .02, Hedges's g_{av} = 0.41; short-lag quizzes: $t(23)$ = 2.21, p < .02, Hedges's g_{av} = 0.33; long-lag quizzes: $t(23)$ = 4.04, p < .001, Hedges's g_{av} = 0.68. Their posttraining performance was marginally worse for the once-presented words, $t(23)$ = 1.61, p = .06, Hedges's g_{av} = 0.27, and there was no significant difference in performance for the short-lag repetition condition, $t(23)$ = 1.10, p = .14, Hedges's g_{av} = 0.18.

Memory strategy survey

The results of the memory strategy survey were analyzed to assess whether the participants in the imagery training group were more likely to report using a mental imagery strategy. Based on their survey responses, participants were placed into one of three categories: those who reported using mental imagery, those who reported using a semantic memory strategy (such as generating sentences using the words or forming associations between items), and those who reported using a shallow memory strategy, such as repetition. Representative examples of each category are shown in Table S3 in the Supplemental Materials. Among participants in the control group who completed the memory strategy survey, five reported using mental imagery, six reported using a semantic strategy, and seven reported using a shallow strategy. In the WM training group, four participants reported using mental imagery, four reported using a semantic strategy, and 11 reported using a shallow strategy. In the imagery training group, 12 participants reported using mental imagery, four reported using a semantic strategy, and three reported using a shallow strategy. A chi-square test showed that participants in the imagery training group reported using an imagery strategy significantly more often than would be expected by chance (χ^2 = 7.93, p = .02).

The recognition memory test results indicated that strategy-based training led to improved performance whereas process-based WM training led to worse performance. However, within each training group, participants reported using a variety of memory strategies. To further investigate the impact of memory strategies on the recognition memory task, we used the data from the survey to assess the relationships between the participants' self-reported memory strategies and their performance on the recognition memory test. In this analysis, rather than grouping the participants according to their training group, we grouped them based on the memory strategies that they reported using for the recognition memory test. The

average d' scores for each group are shown Figure S2 in the Supplemental Materials. For each of the test conditions in the posttraining recognition memory task, one-way ANOVAs were used to compare the d' scores of participants in the imagery, semantic, and shallow strategy groups. The groups' scores were significantly different for every test condition (all F s > 15.95, all p s < .001, all η_p^2 > 0.38). Post hoc t tests showed that participants who reported using an imagery strategy had higher d' scores than did participants in the other two groups for all test conditions (all t s > 3.44, all p s < .001). Participants who reported using a semantic strategy performed better than participants who reported using a shallow strategy for the once-presented items, $t(18)$ = 2.02, p = .03, Hedges's g_s = 0.77, and for the long-lag repetitions, $t(18)$ = 1.95, p = .03, Hedges's g_s = 0.74, but the two groups did not differ significantly on any of the other conditions (all t s < 1.67, all p s > .06).

We also assessed participants' change in recognition memory performance from the pretraining session to the posttraining session using the groupings based on self-reported memory strategy. The average d' difference scores are shown in Fig. 4. A one-way ANOVA showed that the average $\Delta d'$ scores (posttraining d' minus pretraining d') of participants in the imagery, semantic, and shallow strategy groups were significantly different for every recognition test condition (all F s > 5.96, all p s < .01, all η_p^2 > 0.18). Post hoc t tests showed that the participants who reported using mental imagery had a bigger improvement in d' scores than did the other two groups for all conditions (all t s > 2.02, all p s < .03). The participants who reported using a semantic strategy had a bigger change in performance than participants who reported using a shallow strategy for the once-presented items, $t(22)$ = 2.16, p = .02, Hedges's g_s = 0.77, but the two groups did not differ significantly for any other conditions (all t s < 1.67, all p s > .05).

Discussion

The results of this study add to the growing body of literature that suggests that WM training does not necessarily transfer to other memory tasks. Although the participants in the WM training group did improve their performance on the trained tasks, we failed to find near transfer to other WM tasks. The participants in the WM training group did perform better on the untrained baseline WM tasks after training, but their performance did not improve any more than that of the participants in the other groups, indicating that the improvement was not related to the WM training. This study also revealed that WM training can have negative effects on other types of memory tasks. The results showed that the participants in the WM training group performed worse on the recognition memory task after training, whereas participants in the no-contact

Change in d' Scores for Each Self-reported Memory Strategy

Fig. 4 Difference between posttraining memory performance and pretraining memory performance for each condition on the recognition memory test, compared across participants who reported using imagery,

semantic, or shallow memory strategies during the posttraining session. Error bars show mean standard error

control group maintained the same level of performance, and participants in the mental imagery training group improved their performance.

One of the key differences between the mental imagery training group and the WM training group is that the former was trained to use a memory strategy and the latter was not. Differences in use of memory strategies between the two groups could account for their performance changes on the recognition memory test. The memory survey revealed that participants within each training group reported using a variety of memory strategies on the posttraining recognition memory test. However, participants in the imagery training group were more likely to report using mental imagery as a memory strategy than participants in the other two groups, and participants in the WM training group were more likely to report using a shallow memory strategy (repetition or no strategy at all) than were participants in the other groups. The participants who reported using mental imagery outperformed the other groups on the recognition memory test and showed the biggest gains in performance after training. We did not collect data about the participants' strategy use prior to training, so it is not possible to know precisely how each individual's use of memory strategies changed over the course of the experiment. However, the distribution of self-reported memory strategies across the three training groups after training, in combination with the changes in the training groups' performance on the recognition memory task, suggests that changes in memory strategy during training may have had an important impact on posttraining performance. Presumably, each participant

approached the baseline recognition memory task with some sort of memory strategy. The participants in the no-training control group would be likely to use the same strategy for the task both times they completed it, leading to similar performance in both instances. The participants who were trained to use a mental imagery strategy tried to apply that strategy to the recognition memory task and their performance improved. The WM training group's decline in performance suggests that the memory strategies they used after training may have been less effective than the strategies they used prior to training.

Prior research on the effects of strategy use on WM task performance has indicated that asking participants to use a rehearsal strategy can improve WM span, particularly for participants with low span scores prior to strategy training (Turley-Ames & Whitfield, 2003). Other strategies, such as imagery and semantic elaboration, were not as helpful, perhaps because enacting those strategies placed too much of a burden on limited WM resources. McNamara and Scott (2001) also found that participants who completed WM and short-term memory tasks tended to use a rehearsal strategy unless they were trained to use a semantic chaining strategy. In the present study, the intensive WM training may have led participants to favor rehearsal strategies in order to improve their performance during the training sessions. However, rehearsal strategies were not very effective for the recognition memory task, which had over 100 words per block. The results of the memory strategy survey show that use of a rehearsal-based

strategy (or no strategy at all) was associated with lower performance on the recognition memory test.

Past studies have noted that measures of WM capacity are seldom free from strategy use (McNamara & Scott, 2001) and that an accurate assessment of WM capacity requires control over variability in strategy use (Turley-Ames & Whitfield, 2003). Although the goal of assessing WM capacity is to predict higher order cognitive functioning, the goal of WM training is to enhance higher order cognitive functioning. WM training is intended to transfer to tasks beyond the trained tasks so that it can provide more general cognitive benefits to the trainees. In order to foster transfer to other tasks, researchers have used adaptive training tasks to minimize the formation of task-specific memory strategies during training (e.g. Jaeggi et al., 2008). If task-specific strategies are minimized, improvements during training should reflect enhancement of the underlying construct rather than task-specific improvement. However, this study indicates that adaptive WM training tasks may discourage people from adopting memory strategies that would be beneficial when applied to other types of memory tasks. When focusing on WM enhancement, it may be more useful to think of memory strategies as complementary interventions. Development of task-specific strategies remains a critical concern, but rather than discouraging the formation of strategy use via adaptive tasks, it may be beneficial to explicitly teach participants strategies that are applicable to a wide variety of tasks.

One of the limitations of this study is that the total memory training time differed between groups. The WM training group spent more time training, and their training tasks were quite difficult. This may have resulted in a loss of motivation that contributed to their relatively poor performance on the posttraining recognition memory task. Prior research has indicated that motivation can be a key factor in performance (Melby-Lervåg & Hulme, 2013; Redick et al., 2013) and a potential mechanism of task transfer, because improvement on a trained task may motivate participants to perform well on tasks completed following training. This could result in improved performance in the absence of WM enhancement, *per se* (Hayes, Petrov, & Sederberg, 2015). Conversely, in the present study, participants who were discouraged by the WM training regimen may have experienced negative effects. However, the length and number of WM training sessions used in this study were comparable to those used in prior WM training studies that showed a dose–response relationship between number of sessions and positive task transfer (Jaeggi et al., 2008), indicating that the length of the WM training was not necessarily demotivating. In addition, the WM training group's performance on the training tasks improved significantly over the course of the training (see Fig. 2). Because their performance on the training tasks did improve, we would expect them to be motivated to perform well on the posttraining tasks (Hayes et al., 2015). Notably,

the performance decrement for the WM training group was specific to the recognition memory task. Like the other groups, their performance improved on the other posttraining assessments. Finally, the mental imagery and WM training groups had similar rates of attrition. If the additional time spent on the WM training was demotivating, we would expect to see a higher rate of attrition in that training group. Overall, the pattern of results suggests that the participants remained motivated throughout the study.

Another limitation of this study was the use of only one recognition memory transfer task. Shipstead and colleagues (2012) recommended that construct assessment should occur via a battery of multiple tests, which allows for creation of latent variables, thereby reducing the potential for task idiosyncrasies driving results. Multiple tests enable creation of a latent difference score model that uses factor-analysis techniques to assess gains at the latent level (Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009). Our focus was on comparing WM training to memory strategy training in addition to a no-training control group. Due to the constraints imposed by our participant pool (professionals whose time is limited and expensive), there was a trade-off between including more tasks or more participants. We chose to use three training groups and a relatively small battery of pre- and posttraining tasks. The recognition memory task was selected because recognition memory is crucially important in real-world learning tasks and because the task could be structured as a near-transfer task for the imagery training group and a far-transfer task for the WM training group. The task included multiple conditions, such as repeated and quizzed words appearing at varying lags within the study list, but the study design did not allow us to take a latent variable approach. Therefore, it is possible that results on the recognition memory task are due to a task-specific quirk rather than reflective of changes to the underlying construct. This should be addressed in future research by incorporating a more comprehensive task battery to allow for a latent variable approach.

In summary, this study was designed in order to evaluate the relative efficacy of strategy-based (imagery training) and process-based (WM training) training programs. The results present a cautionary tale about unintended consequences arising from cognitive training. Even if participants improve on the trained task, the training may impact their performance on untrained tasks in unforeseen ways. If intensive cognitive training changes trainees' use of strategies, it could impact on performance on many other tasks. This is particularly concerning if the mental strategies that are reinforced during training are maladaptive for other common tasks. Many studies have investigated near and far transfer of WM training to WM and fluid intelligence tasks, but few studies have examined the impact of WM training on other types of memory performance. In this study, we found an example of negative transfer in which WM training decreased performance on a

recognition memory task. Given the broad popularity and time-consuming nature of WM training, its effects on other types of memory that are crucial for learning, and the interplay between WM training and memory strategies are clearly areas that warrant additional research.

Author note Laura E. Matzen, Michael C. Trumbo, Michael J. Haass, Austin Silva and Susan Stevens-Adams, Sandia National Laboratories, Michael A. Hunter, University of New Mexico, and Michael Bunting and Polly O'Rourke, Center for the Advanced Study of Language.

This work was funded by the Sandia National Laboratories Laboratory Directed Research and Development Program.

References

- Alloway, T. P., Bibile, V., & Lau, G. (2013). Computerized working memory training: Can it lead to gains in cognitive skills in students? *Computers in Human Behavior*, 29, 632–638.
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, 129, 126–145.
- Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). *The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords*. Retrieved from <http://lexicon.wustl.edu/>
- Becker, J. T., & Morris, R. G. (1999). Working memory(s). *Brain and Cognition*, 41, 1–8.
- Braun, K., & Rubin, D. C. (1998). The spacing effect depends on an encoding deficit, retrieval, and time in working memory: Evidence from once-presented words. *Memory*, 6, 37–65.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- Carretti, B., Borella, E., Fostinelli, S., & Zavagnin, M. (2013). Benefits of training working memory in amnesic mild cognitive impairment: Specific and transfer effects. *International Psychogeriatrics*, 25, 617–626.
- Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.
- Chen, J. M., & Morrison, A. B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193–199.
- Chooi, W., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, 40, 531–542.
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, 320, 1510–1512.
- Daneman, M., & Blennerhassett, A. (1984). How to assess the listening comprehension skills of prereaders. *Journal of Educational Psychology*, 76, 1372–1381.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309–331.
- Ericsson, K., & Delaney, P. (1998). Working memory and expert performance. In R. Logie & K. Gilhooly (Eds.), *Working memory and thinking: Current issues in thinking and reasoning* (pp. 93–114). Hillsdale, NJ: Erlbaum.
- Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence*, 48, 1–14.
- Healy, A. F., Wohldmann, E. L., Sutton, E. M., & Bourne, L. E., Jr. (2006). Specificity effects in training and transfer of speeded responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 534–546.
- Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, 24, 827–836.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *PNAS*, 105, 6829–6833.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, 131, 66–71.
- Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the *n*-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 615–622.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4, 863.
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676.
- Luo, Y., Wang, J., Wu, H. R., Zhu, D. M., & Zhang, Y. (2013). Working-memory training improves developmental dyslexia in Chinese children. *Neural Regeneration Research*, 8, 452–460.
- McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition*, 29, 10–17.
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49, 270–291.
- Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 9, 596–606.
- Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, 31, 237–260.
- Paivio, A. (1965). Abstractness, imagery and meaningfulness in paired associate-learning. *Journal of Verbal Learning and Verbal Behaviour*, 4, 32–38.
- Paivio, A. (1971). *Imagery and mental processes*. New York, NY: Holt, Rinehart and Winston.
- Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1196–1204.
- Prestia, F. L., & Zacks, R. T. (1974). The effects of learning instructions and cueing on free recall. *Memory & Cognition*, 2, 194–200.
- Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142, 359–379.
- Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92.

- Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. H. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLoS ONE*, 7(11), e50431. doi:10.1371/journal.pone.0050431
- Schmiedek, F., Hildebrandt, A., Lövdén, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1089–1096.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, 125, 4–27.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50, 245–276.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, 138, 628–654.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J., Novick, J. M., Chrabaszcz, J. S., . . . & Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, 41, 638–663.
- St. Clair-Thompson, H., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology*, 30, 203–219.
- Thompson, T. W., Waskom, M. L., Garel, K. A., Reynolds, G. O., Winter, R., Chang, P., & Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE*, 8(5), e63614. doi:10.1371/journal.pone.0063614
- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in on mental function upon the efficiency of other functions (I). *Psychological Review*, 8, 247–261.
- Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49, 446–468.
- West, C. W., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, 12, 1024–1037.
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine readable dictionary (Version 2). *Behavioural Research Methods, Instruments and Computers*, 20, 6–11.