

# Psychology and Aging

## **Working Memory Training and Transfer in Older Adults**

Lauren L. Richmond, Alexandra B. Morrison, Jason M. Chein, and Ingrid R. Olson

Online First Publication, June 27, 2011. doi: 10.1037/a0023631

### CITATION

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011, June 27). Working Memory Training and Transfer in Older Adults. *Psychology and Aging*. Advance online publication. doi: 10.1037/a0023631

# Working Memory Training and Transfer in Older Adults

Lauren L. Richmond, Alexandra B. Morrison, Jason M. Chein, and Ingrid R. Olson  
Temple University

There has been a great deal of interest, both privately and commercially, in using working memory training exercises to improve general cognitive function. However, many of the laboratory findings for older adults, a group in which this training is of utmost interest, are discouraging due to the lack of transfer to other tasks and skills. Importantly, improvements in everyday functioning remain largely unexamined in relation to WM training. We trained working memory in older adults using a task that encourages transfer in young adults (Chein & Morrison, 2010). We tested transfer to measures of working memory (e.g., Reading Span), everyday cognitive functioning [the Test of Everyday Attention (TEA) and the California Verbal Learning Test (CVLT)], and other tasks of interest. Relative to controls, trained participants showed transfer improvements in Reading Span and the number of repetitions on the CVLT. Training group participants were also significantly more likely to self-report improvements in everyday attention. Our findings support the use of ecological tasks as a measure of transfer in an older adult population.

*Keywords:* working memory, plasticity, training, attention, ecologically valid

Pop culture games such as those included in Nintendo's BrainAge™ and on the website Lumosity™, encourage the claim that mental abilities, such as working memory (WM), can be improved through repetitive mental exercise. Researchers are interested in exercises that might improve WM because this central mental capacity is closely tied to fluid intelligence, reasoning, and other aspects of higher cognition. WM has been defined as a mental workspace that is capacity-limited (Morrison & Chein, 2011). Early models of WM posited the existence of material-specific slave systems serving as dedicated short-term storage structures under the control of an active manipulation component (i.e., the central executive; Baddeley & Hitch, 1974), while more recent models propose a single memory store in which specific information can be brought up into working memory, and even more centrally into the focus of attention (Cowan, 1999).

Several recent scientific investigations indicate that WM can be improved through particular types of "mental exercise" (Chein & Morrison, 2010; Schmiedek, Lovden, & Lindenberger, 2010; Thorell, Lindqvist, Nutley, Bohlin, & Klingberg, 2009). Working memory training studies typically follow a similar paradigm: beginning with an initial assessment of cognitive ability, followed by

a period of cognitive training, and ending with a posttraining cognitive assessment. Using this type of design, several studies have reported that WM can be improved through repetitive daily practice (reviewed by Lustig, Shah, Seidler, & Reuter-Lorenz, 2009; Morrison & Chein, 2011). Moreover, these studies demonstrate that certain WM training paradigms can impact other cognitive abilities, such as those assessed in tests of general fluid intelligence (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Schmiedek et al., 2010; but see Chein & Morrison, 2010; Dahlin, Nyberg, Backman, & Neely, 2008), cognitive control (Chein & Morrison, 2010; Klingberg et al., 2005; Klingberg, Forssberg, & Westerberg, 2002), and reading comprehension (Chein & Morrison, 2010).

The fact that WM training can result in such "far transfer" is both theoretically and practically significant. Theoretically, these findings lend backing to an extensive psychometric literature suggesting a relationship between WM capacity and individual differences in fluid intelligence and executive functioning (Engle, Kane, & Tuholski, 1999), language acquisition (see Baddeley, 2003 for a review), reading comprehension (Daneman & Carpenter, 1980), problem solving (Logie, Gilhooly, & Wynn, 1994), and reasoning (e.g., Kane et al., 2004). These findings suggest that WM may serve as a domain-general cognitive resource that modulates ability in a number of seemingly disparate areas of cognitive performance. Practically, transfer effects are important because they raise the possibility that training designed to target this general cognitive resource could produce gains in a potentially vast spectrum of cognitive functions.

While some studies have demonstrated far transfer from WM training, a number of studies have also failed to obtain significant far transfer benefits (see Table 1). A recent large-scale study conducted in the United Kingdom has received particularly broad attention for demonstrating no significant transfer effects following a daily regimen of WM training (Owen et al., 2010). Although it has been suggested that the study has several methodological shortcomings (for a review of criticisms, see Katsnelson, 2010),

---

Lauren L. Richmond, Alexandra B. Morrison, Jason M. Chein, and Ingrid R. Olson, Department of Psychology, Temple University.

This research was supported by NIH Grants RO1MH071615 and RO1MH091113 to I. Olson. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health. Additionally, we thank Sheetal Amin, Ashley Selig, and Katelyn Welshans for their help in testing participants and organizing data, and Tania Giovannetti for helpful advice.

Correspondence concerning this article should be addressed to Lauren L. Richmond, Department of Psychology, Temple University, 1701 North 13th Street, Philadelphia, PA 19122. E-mail: lauren.richmond@temple.edu

Table 1  
Peer-Reviewed Studies of the Effects of Core WM Training in Older Adults

Reference	Training task	Control task	Amount of training	Near transfer	Far transfer
Buschkuehl et al. (2008)	Spatial WM ( $N = 13$ )	Aerobic exercise ( $N = 19$ )	720 min	Block span	Visual free recall
Dahlin et al. (2008a; 2008b)	Letter memory, WM updating ( $N = 11$ )	Non-active control ( $N = 8$ )	675 min	None	None
Li et al. (2008)	Spatial n-back ( $N = 21$ )	Non-active control ( $N = 20$ )	675 min	Variation of n-back	None
Mahnacke et al. (2006)	Auditory discrimination tasks ( $N = 53$ )	Non-active control/ educational DVD ( $N = 56; 53$ )	2700 min	None	Global auditory memory, DSF
Schmiedek et al. (2010)	Processing speed, episodic memory, non-adaptive WM ( $N = 103$ )	Non-active control ( $N = 39$ )	6060 min	Verbal WM span	Rotation span, Raven's IQ, word pairs

*Note.* These studies attempt to increase the central capacity of WM. Studies encouraging strategy training in older adults are not included. The  $N$ 's represent older participants, excluding any young cohorts tested, that completed the study and whose data was included for analysis. Note that all studies that are listed showed an increase in WM capacity after training, but only a small number showed transfer effects. DSF = Digit Span Forward.

their results have raised doubts regarding the efficacy of WM training as a tool for general cognitive enhancement.

Moreover, transfer findings are especially unimpressive in studies conducted among older adults. Indeed, trends in the extant literature suggest that far transfer effects may be age-dependent, with robust transfer effects apparent in young adults and weak transfer effects observed in older adults (see Table 1 for a tabulation of the literature on WM training in older adults). For instance, Klingberg and colleagues showed that spatial WM training in children leads to marked gains in a wide range of other cognitive skills, such as general intelligence and Stroop speed and accuracy, that persist at a 3-month follow-up visit (Klingberg et al., 2005). In a contrasting example, one study attempting to train spatial WM processes in adults aged 80 years and older showed transfer effects to block span and visual free recall in the training group, and at a 1-year follow-up assessment, no reliable differences persisted between groups (Buschkuehl et al., 2008). Another study looking at improving multiple areas of cognition in older adults, including processing speed and phoneme discrimination, found little evidence of skill transfer to nontrained domains (Mahnacke et al., 2006). Perhaps the most encouraging findings are found in a recent study in which far transfer was reported for adults who engaged in executive function training. However, the training benefits were smallest in older adults (Karbach & Kray, 2009). Similarly, Schmiedek and colleagues (2010) executed the same training protocol in groups of younger and older adults, and found that far transfer from training was of a substantially smaller magnitude in the older cohort. Together, these results illustrate the status quo in the field of WM training with regard to the age dependence of transfer effects. Dishearteningly, a recent review concluded that in older adults "few studies report transfer effects for tasks dissimilar enough from the trained tasks to suggest transfer at the broad level of abilities" (Noack, Lovden, Schmiedek, & Lindenberger, 2009).

As adults age, they typically experience cognitive decline. Often these cognitive complaints are related to episodic memory function, but they may also take on the form of difficulties with executive functioning, of which WM is a key component (Salt-house, Atkinson, & Berish, 2003). It is thus ironic that the transfer findings are so modest in studies of older adults, for whom effective WM training interventions could be most meaningful. Techniques that delay or reduce age-related cognitive decline

could improve older individuals' capacity to lead independent lives, and decrease the social burden of caring for those with diminished cognitive capacity (Hertzog, Kramer, Wilson, & Lindenberger, 2009).

While prior findings suggest that age may be an important variable in obtaining positive far transfer from WM training, there are also important methodological limitations that might account for the poor record of transfer in older populations. First, there are a number of studies of WM training conducted in older adults that have used training programs that are not adaptive (e.g., Li et al., 2008). Adaptive training may be important because it ensures that each successive day of training presents an increasing challenge, thus curtailing a slide into automatic, habitual response patterns. In fact, nonadaptive versions of training tasks have at times been used as control tasks (Holmes, Gathercole, & Dunning, 2009). Second, since many older adults spend little time in front of a computer and have been removed from testing environments (e.g., school) for a long time, this cohort may simply be less adept with the computerized laboratory tests that have been previously used to assess performance pre- and posttraining. This fact may explain why older adults can exhibit substantial gains in the computerized training tasks that they complete repetitively and with intensive practice over the course of training, but not on the unfamiliar and sparsely sampled transfer measures. Either of these issues (the use of nonadaptive training regimes, less familiarity or comfort with the procedures used to assess transfer) might reduce the ability to detect training-related transfer gains. Thus, it is plausible that important transfer benefits are present for older adults, but have been overlooked simply because the prior literature has not assessed generalization from training using measures that are pertinent (in terms of the implications for everyday functioning) or well-suited to this age group.

In order to investigate whether WM can improve via training in older adults, and whether such improvements can elicit near or far transfer to other measures, we trained older adults on a WM training task that was previously validated in young adults (Chein & Morrison, 2010). This approach to training involves repetition of verbal and spatial variants of an adaptive, complex WM span task. Complex span tasks, characterized by interweaving a storage task and an unrelated processing task, are thought to place a premium on domain-general attention control mechanisms (see

Engle & Kane, 2004). We accordingly hypothesized that strengthening of these mechanisms through repetition of adaptive complex WM span tasks would yield generalizable cognitive benefits. Encouragingly, Chein and Morrison (2010) found that younger adults who trained with these tasks showed significantly improved WM task performance, and that the gains in WM performance transferred to improvements on a composite of short-term memory measures (letter, location span), controlled attention (Stroop task) and reading comprehension (Nelson Denny).

In our study, older adults were assigned to one of two groups: a training group or an active control group. Both groups were administered a battery of tests pre and post intervention, including assessments of everyday functioning. We predicted that: (a) older adults would show improved WM span after training, and (b) older adults would show far transfer to assessments of everyday functioning. In addition, we predicted a replication of the finding in Chein and Morrison (2010) that this particular WM training paradigm does not produce far-transfer to a common measure of general intelligence, Raven's Progressive Matrices.

An additional aim of this study was to address a general criticism of the WM training literature - the absence of comparison to an active control group that engages in an alternate activity during the training interim. Unfortunately, many studies of WM training use only a nonactive, no-contact control group, (for examples of nonactive control groups, see Chein & Morrison, 2010; Dahlin, Neely, Larsson, Backman, & Nyberg, 2008a; Li et al., 2008; Schmiedek et al., 2010). It is well known that participants can improve or modify their performance by simply reacting to the fact that they are being studied, rather than to an experimentally manipulated variable (e.g., the *Hawthorne Effect*, discussed by Green & Bavler, 2008). A no-contact control group will account for practice effects in posttest performance, but may be insufficient to clarify whether findings are due to the Hawthorne Effect. In the present study, the control group completed a trivia training program, which was designed to engage and motivate the participants without targeting WM mechanisms.

## Method

### Participants

A group of older (age range 60–80) participants were recruited from the community and screened for neurological and psychiatric disorders via a self-report questionnaire. Participants with a minimal state exam (MMSE; Folstein, Folstein, & McHugh, 1975) score of 26 or below were excluded from further testing. Forty participants (eight male, 32 female) entered and completed the study, with a mean age of 66 years, mean education of 17 years, and mean MMSE score of 29. Six additional participants initiated the longitudinal phase of the study and subsequently withdrew their participation. Participants were randomly assigned to the active control group ( $n = 19$ ) or training group ( $n = 21$ ). The groups did not differ in age, education, or MMSE score (all  $ps > 0.20$ ). All procedures were reviewed and approved by the Temple University Institutional Review Board.

### Design

The complete study consisted of three parts: (a) a pretest assessment, (b) a complex WM training regime (experimental group)

or trivia learning regime (active control group), and (c) a posttest assessment.

### Pretest Assessment

Participants completed two hours of on-site testing administered by a trained experimenter. Pretest materials consisted of several standardized neuropsychological tests. The MMSE was administered to allow screening for dementia. Two tests were given to assess the presence of “near” transfer effects after WM training: reading span, a test of complex working memory span (Unsworth, Heitz, Schrock, & Engle, 2005), and digit span. Reading span in particular was included because of the similarity to the training task in terms of possessing process-and-storage components. Digit span was chosen because it is a widely used simple short-term memory span task that may be positively affected by training on a complex WM paradigm. Three tests were given to assess “far” transfer effects after WM training: Raven's Standard Progressive Matrices (Raven's; Raven, 1976), portions of the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994), and the California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987). Raven's, a measure of nonverbal fluid intelligence, was administered because some prior studies have found that WM training benefits transfer to measures of fluid intelligence (Jaeggi et al., 2008; Karbach & Kray, 2009; Schmiedek et al., 2010). However, other studies have failed to find this effect (Chein & Morrison, 2010; Dahlin et al., 2008a), so the issue remains contentious. The TEA is an ecological measure of attention that engages participants in a scenario which participants feel is “relevant to their . . . everyday life” (Robertson et al., 1994). Furthermore, portions of the TEA have been shown to better predict functional status in individuals with potential cognitive impairment (multiple sclerosis patients) than traditional laboratory tasks (Higginson, Arnett, & Voss, 2000). This task was chosen because attention and working memory are closely related cognitive constructs (for reviews, see Engle & Kane, 2004; Engle et al., 1999). The subtests administered were: (a) Elevator Counting: counting a series of tones; (b) Elevator Counting with Distraction: counting low-pitched tones and ignoring higher-pitched tones; (c) Visual Elevator: counting pictures of elevator doors as floors and reversing the count as pictures of arrows pointing up or down are encountered; (d) Elevator Counting with Reversal: the auditory correlate of the Visual Elevator subtest; (e) Telephone Search: quickly searching through a page in a telephone book for target shapes; and (f) Telephone Search while Counting: searching for the target shapes, with the additional contingency of counting a series of tones and later reporting the total. The CVLT is a 16-item list-learning task with a subsequent distracter 16-item list, testing short- and long-term verbal learning and memory. Participants are required to recall items from the target list (a) directly after hearing the target list (Trials 1–5), (b) after an intervening distracter list without presentation of the target list, and (c) after a 20-min delay without presentation of the target list. This test was chosen because it addresses a main concern of cognitive aging in older adults, namely problems with episodic memory (Small, 2001). There were no significant group differences at pretest (all  $ps > .08$ ) for any of the assessed measures.

## Training Regime

The WM training regime was modeled after that of Chein and Morrison (2010) (see Figure 1), with minor modifications. Participants in the experimental condition participated in 20, 30-min. long training sessions on a complex WM span task over a span of 4–5 weeks, with most participants completing their sessions in a 5-day a week (e.g., weekday) regimen. To accommodate missed days, some participants continued training into the fifth week. All participants completed posttesting no more than 40 days from training Day 1. Within each 30-min training session, approximately 15 minutes were devoted to a spatial WM task and approximately 15 minutes were devoted to a verbal WM task. Trials were blocked by material type, but randomly sampled each day so that there was an equal likelihood of beginning each training session with the spatial or verbal subtest. The spatial subtest involved making symmetry decisions (symmetry) about a series of partially filled (black and white squares) matrices while intermittently encoding a sequence of highlighted locations on a  $4 \times 4$  grid for later recall. Locations were recalled by mouse clicking on locations within the  $4 \times 4$  grid in the order that they were seen. The verbal subtest required participants to make a series of word/nonword decisions (lexicality) while intermittently encoding a sequence of letters for later recall. The letters were sampled from a pool of 16 consonants, evenly sampled throughout the task. The probe screen consisted of a  $4 \times 4$  grid containing all of the possible letters. The letter sequence was recalled by mouse clicking on the appropriate letters (shown in shifted positions from trial to trial) in the order that they were seen. After each trial was completed, correct/incorrect feedback for both the decision and recall portions were presented on the computer screen. Successive trials were initiated by participant keypress.

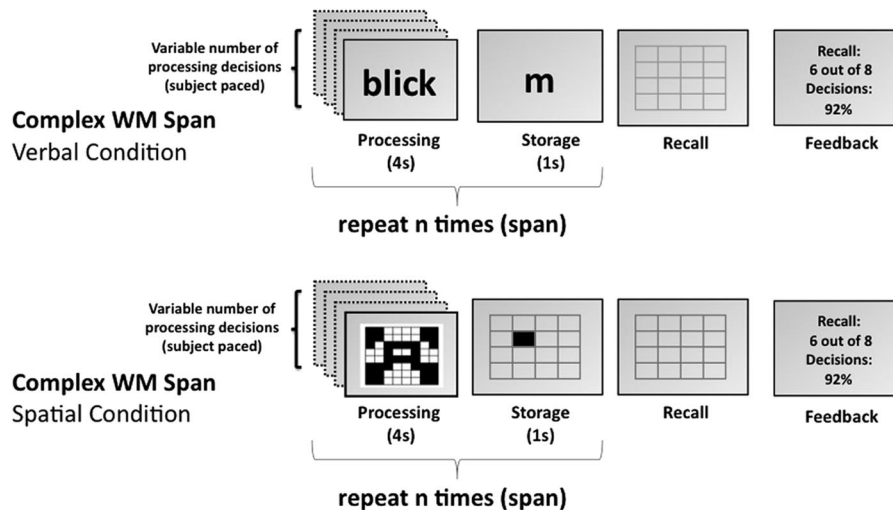
All training participants began the program with a span of two recall items. Participants were told to work as quickly and as accurately as possible, and that both the decision and recall portions of the task were important. The training task was adaptive to the participant's performance level such that the number of recall items increased or decreased based on performance. If both items were correctly recalled for two trials in a row and at least 75% of the lexicality or symmetry decisions were made correctly, the list length was increased to three items, and so on. Likewise, two successive incorrect trials caused the set size to be reduced by one item.

## Trivia Regime

Participants in the active control condition completed 20, 30-min. long, sessions of trivia "training". This task was chosen because it was presumed to have a low WM load while still being cognitively engaging. Participants were given a username and password for funtrivia.com and a list of quizzes that they could choose to complete. Some quizzes on funtrivia.com were targeted at improving memory or IQ, and these were excluded from the list of permissible quizzes to avoid potential confounds. Participants recorded their progress on a log sheet upon completion of each session.

## Posttest Assessment

Upon completion of the training sessions, participants from both groups returned to perform a 2-hr posttest assessment that mirrored the pretest (excluding the MMSE), using alternate forms for the TEA, CVLT and Raven's. Alternate forms were not available for digit span and reading span. Participants also completed a posttest



*Figure 1.* A schematic of the training task. Both portions included four seconds of processing decisions (verbal: word/nonword, spatial: symmetry) and one second of storage. After the appropriate span was reached (indicated by the number of items to be stored), participants engaged in recall of the stimuli (verbal: letters, spatial: locations) and then saw a feedback screen. Each portion of the training (verbal and spatial) lasted for 15 minutes. (Breaks taken by subjects during the training period were not counted toward the 15-minute total.) Adapted from Expanding the mind's workspace: Training and transfer effects with a complex working memory span task by J. Chein and A. Morrison, 2010, *Psychonomic Bulletin and Review*, Copyright 2010 by The Psychonomic Society.

questionnaire, created for and tailored to our training study, which probed subjective reports about the cognitive effects of training. Effect sizes for within-subjects data were computed using methods prescribed by Morris and DeShon (2002).

## Results

We investigated two primary questions regarding the impact of WM training on older adults: (a) Does complex WM span task training lead to an increase in WM task performance; and (b) Do the benefits of WM training transfer to untrained tasks?

### Effects of WM Training

The session-to-session effects of WM training on our older adult sample population are depicted in Figure 2a, which shows that older adults improved their WM performance (WM span) with training. To assess the statistical significance of training-related gains obtained in the spatial and verbal tasks, we compared older adults' WM span on the first (Session 1) to that in the last (Session 20) training session using a 2-factor (session and memory type: spatial or verbal) repeated measures ANOVA. A significant main effect of session was obtained,  $F(1, 19) = 19.72, p < .001$ . Over training, average verbal memory span improved by 51%, from a span of 3.70 to 5.57,  $t(19) = 3.10, p = .006, d = 1.107$ . Average spatial memory span improved by 46%, from a span of 2.80 to 4.10,  $t(19) = 3.79, p = .001, d = .844$ . Although WM performance was generally higher on the verbal subtest as shown by the main effect of memory type,  $F(1, 19) = 23.089, p < .001$ , the interaction of session and memory type was not significant,  $F(1,$

19) = 1.83,  $p = .192$ , thus indicating that verbal and spatial WM showed similar levels of improvement with training.

Additionally, we looked at speed of processing decisions as a potential variable. Participants were faster at making lexicality and symmetry decisions at the end of the training than at the beginning, as measured by the number of processing decisions made throughout the course of one training session—processing speed data for three participants were not available for Session 1 and for one participant for Session 20;  $F(1, 16) = 26.796, p < .001, d = 1.342$ . These results suggest that participants were able to improve WM by increasing the number of to-be-remembered items that were held in WM as well as increasing the speed with which they were able to complete processing decisions correctly. This result provides evidence that WM span improvements did not arise from subjects strategically trading off processing (decision) accuracy to increase storage (recall) performance.

### Age Differences in Training

For comparison, effects of a nearly identical WM training paradigm on young adult performance, as assessed by Chein and Morrison (2010), are depicted in Figure 2b. Figure 2 shows that older adults' performance began at a relatively lower level than younger adults, and reached a similar level on the verbal subtest by the end of training. Using a mixed effects ANOVA, treating age as a between-subjects factor and memory type and session as within-subject factors, we observed a main effect of memory type [verbal better than spatial for both young and old,  $F(1, 33) = 28.318, p < .001$ ], a main effect of time [Session 20 span greater than Session 1 span for both young and old,  $F(1, 33) = 27.240, p < .001$ ] and a main effect of age [younger adults performed better than older adults,  $F(1, 33) = 18.628, p < .001$ ]. By the end of training, younger adults maintained their superior WM span on the spatial subtest [at Session 20: spatial,  $t(34) = 2.674, p = .012$ ], but group differences on the verbal test only approached significance [verbal  $t(34) = 1.822, p = .077$ ]. We did not observe any two- or three-way interactions between these factors, but the interaction of memory type by time did reach marginal significance [verbal WM improved more than spatial WM over time, ( $F(1, 33) = 3.086, p = .070$ , all other  $ps > .15$ ).

### Transfer Effects

Our training regime involved practice with tasks that tested memory for letters and locations in the presence of distraction from a secondary processing task. Thus, one might expect there to be near transfer to short-term memory tasks using similar materials. Indeed, this was previously observed in young adults, using the same training task as used here (Chein & Morrison, 2010). To assess transfer effects in the present study, we used a mixed effects factorial ANOVA with group as the between subjects factor and testing occasion (pre- or posttest) as the within subjects factor. We found that older adults who underwent WM training exhibited improved reading span performance compared to control participants,  $F(1, 38) = 7.325, p = .010, d = 1.078$ , but did not increase their forward or backward digit span. Participants in the trivia group also did not show significant improvement on this measure (see Table 2).

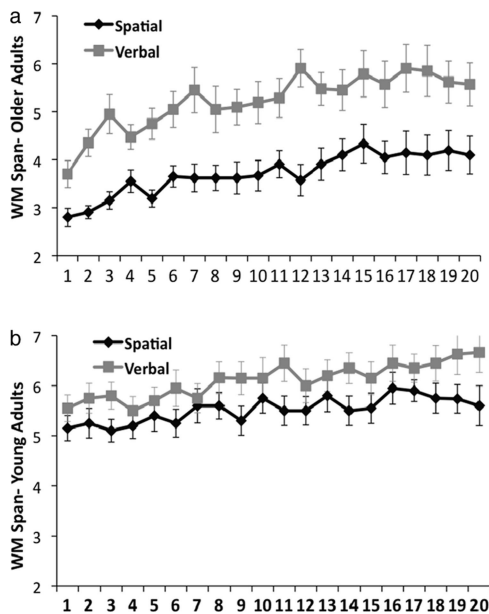


Figure 2. (a) Working memory training performance in older adults across time. (b) Working memory training performance in young adults (M age 20 years), replotted from Chein and Morrison (2010). Young adults began the study with a span of four for each verbal and spatial, while older adults began with a span of two items for each.

Table 2

*Task Performance Scores in Cognitive Measures Before and After the Training Period*

	Control Group			Training Group			Mixed ANOVA interaction
	<i>M</i> Before (SEM)	<i>M</i> After (SEM)	<i>p</i> -value	<i>M</i> Before (SEM)	<i>M</i> After (SEM)	<i>p</i> -value	<i>p</i> -value
Reading span	26.47 (4.94)	29.89 (4.32)	.21	20.10 (2.57)	35.48 (4.26)	<b>&lt;.01</b>	<b>.01</b>
Digit span forward	10.53 (.51)	10.84 (.55)	.33	11.52 (.48)	11.52 (.46)	1.00	.58
Digit span backward	6.63 (.47)	7.47 (.54)	.08	7.19 (.48)	6.95 (.42)	.68	.15
Raven's	15.32 (.95)	15.74 (1.02)	.66	17.43 (.72)	16.29 (.88)	.32	.30
TEA: ECR	9.74 (.94)	11.21 (1.05)	.18	11.71 (.98)	11.38 (.92)	.62	.15
CVLT Total Correct	99.47 (5.97)	104.58 (4.44)	.14	107.38 (2.95)	109.24 (3.79)	.59	.51
CVLT Repetitions	4.53 (1.18)	6.05 (1.47)	.19	5.10 (.84)	2.67 (.44)	<b>&lt;.01</b>	<b>&lt;.01</b>
CVLT Intrusions	3.89 (1.14)	3.11 (.86)	.36	3.14 (.91)	2.19 (.49)	.27	.89

*Note.* There were no significant group differences at pre-test. Significant effects are highlighted with bold font. Digit span measures report number of correct trials, at 2 trials per length. TEA: ECR represents scaled scores. For the CVLT Total Correct, a higher score is better while CVLT Repetitions and Intrusions, a lower score is better.

Of greater interest is whether there is evidence of transfer from WM training to measures that assess more disparate cognitive abilities. As shown in Table 2, trained participants exhibited statistically significant improvements in the number of verbal items repeated compared to controls, as measured by the CVLT,  $F(1, 38) = 9.413, p = .004, d = .876$ . In the CVLT, participants were instructed to recall as many items as they could from a 16-word list, without repeating already-recalled words in a given trial (repetitions), hence decreased repetitions indicate better performance on the task. Intrusions were demonstrated by including words not given in the target list in the recall portion of the task. Short-term (recall directly after presentation of a target list) and long-term (recall after a 20 minute delay) memory were assessed by the CVLT. Control participants' performance on CVLT repetitions were of a nonsignificant magnitude (see Table 2).

Turning to the TEA, the pattern of results suggested that there was a ceiling effect (discussed in more detail below) present for all but one subtest (elevator counting with reversal; one participant in each group could not discriminate between the tones used in this subtest and as such did not complete the task). There were no significant improvements exhibited on any subtest by either group.

Across transfer measures, the overall group data suggested that there was meaningful transfer from WM training to some untrained transfer measures. However, individuals in the WM training group varied greatly in the size of their training effect (WM span improvements), ranging from no increase for some participants to an increase of WM span by nine items for others. To further mine the data for evidence of positive transfer, we accordingly examined the transfer performance among only the subset of participants for whom WM training produced an observable WM span improvement.

Specifically, we examined transfer performance in a subset of 'well-trained' participants who demonstrated training-related improvements for WM span in the spatial domain (the more stringent of the two tasks, and the one that is least likely to benefit from rehearsal strategies). We defined this group as participants who showed an improvement of two span items or better in the spatial portion of the WM training task (final spatial set size 4 or greater,  $n = 11$ ). Extending the earlier transfer findings, we found that the well-trained group recalled a greater number of words at posttest

on the CVLT following a 20 minute delay, compared to pretest,  $t(10) = 4.10, p = .002, d = 1.183$  and made fewer intrusion errors,  $t(10) = 2.80, p = .017, d = .870$ . The trivia group did not demonstrate similar improvements (all  $ps > .40$ ), and differences between groups at posttest on the CVLT long-delay were significant,  $t(29) = 2.102, p = .044$ . The subset of well-trained participants also showed increased speed in completing the Telephone Search subtest of the TEA,  $t(10) = 2.42, p = .036, d = .764$ . However, differences between the experimental and active control groups at posttest did not reach significance ( $p > .60$ ).

### Subjective Reports

Beyond assessment of task performance, we sought to determine whether older adults subjectively felt like they were achieving cognitive benefit from the training exercises. To assess subjective experience, a questionnaire probing general cognitive changes that might be attributed to the training regime was administered at posttest to all study participants.

On this questionnaire, all participants, regardless of training group, reported a positive change in their memory compared to before beginning the training ( $p > .20$ ). Probing further, we asked participants what specific aspect of their cognition they felt was being improved. Compared to the trivia control subjects, a significantly greater number of participants in the training group self-reported an increase in attention when queried about general cognitive improvements they thought may have been affected by training (seven training participants and two trivia participants freely responded that they observed an increase in their attention [one-tailed  $\chi^2(1, n = 9) = 2.78, p = .05$ ]). Participants also were asked to rate their level of commitment to the daily training exercises, as well as their overall enjoyment. Ratings were given on a 5-point Likert-type scale, 5 being high and 1 being low. Self-reported commitment to daily exercises did not differ by group,  $t(36) = 0.875, p = .387$ , nor did reported enjoyment of the training exercises,  $t(36) = 0.876, p = .387$ .

### Discussion

The present study applied a WM training task, which has successfully produced far transfer in younger adults (Chein & Mor-

rierson, 2010), to a sample of older adults. By using a complex span task, which was designed to target domain general WM mechanisms, we aimed to demonstrate transfer from the trained task to other cognitive measures tapping similar resources. Key limitations of the existing literature were also addressed. First, performance by the trained group was compared to that of an active control group who engaged in a trivia program. Second, ecological measures were used in the pre- and posttraining assessment in an effort to assess performance in tasks indicative of everyday functioning.

As a nearly identical training protocol was used in younger adults (Chein & Morrison, 2010), the current study afforded us the opportunity to look at training gains across age groups. WM training gains were similar in our younger and older cohorts (see Figures 2a and 2b). Attaining similar levels of performance in our older cohort in comparison to a college-aged sample may be due to the fact that our population was extremely compliant and highly motivated, or it may be explained by the fact that older adults began training with lower memory spans, so they had more room to improve. In both younger and older adults, training led to improvements on the trained task as well as near transfer to other short term and working memory tasks (i.e., letter span in Chein & Morrison, 2010; reading span in the present study). Notably, neither study demonstrated transfer to measures of general fluid intelligence (Ravens Advanced Progressive Matrices in Chein & Morrison, 2010; Ravens Standard Progressive Matrices in the present study).

The positive transfer findings partially dispel the notion that older adults' cognitive skills are less malleable than those of younger adults. Instead, our results suggest that the lack of transfer observed in the WM training literature for older adults may in fact be more related to the selection of measures that are unsuitable for assessing this group's performance.

Perhaps the most exciting and novel finding from the present study was that older trained participants showed a reduced number of repetitions on the CVLT, a test of verbal learning and memory. In order to reduce the number of repetitions made throughout the CVLT, participants need to be able to recall what they have already said, trial-by-trial. This involves constantly "updating" the list of words that have already been recalled, and comparing this list to all of the words that have been encoded. Interestingly, a prior study showed that although older adults can improve their WM with training, there was no evidence of transfer to an updating task (*N*-back task; Dahlin et al., 2008a; Li et al., 2008). We interpret our successful demonstration of significant transfer to CVLT updating as resulting from the comparatively increased suitability of this assessment task among older adult participants. However, specific shared aspects of our training paradigm and the repetition component of CVLT performance (both can be thought of as a divided attention task with elements of inhibition) may account for the CVLT repetition findings.

Transfer effects in older populations are notoriously difficult to obtain (Noack et al., 2009). In our review of the relevant literature (see Table 1), three WM training studies reported little or no transfer in older adults (Dahlin et al., 2008a; Dahlin et al., 2008b; Li et al., 2008) and three studies, not including our own, reported far transfer (Buschkuhl et al., 2008; Mahncke et al., 2006; Schmiedek et al., 2010). Our study contributes two important findings to the extant literature on WM training in older adults;

transfer was found in participant's ability to inhibit repeating already-recalled items from memory (CVLT repetitions) and in one measure of attention (self-report), in comparison to a cognitively active control group that was subjected to the same demands in terms of time commitment and contact with the research team.

## Negative Findings

The field of WM training is relatively new. Because of this, investigators such as ourselves are still in search of appropriate cognitive tasks that will reveal transfer effects, if they exist. Additional variables such as age make it even more difficult to identify appropriate transfer measures. In choosing our transfer tasks, we relied on the rule that a key prerequisite of transfer is that the training task and transfer task tap similar underlying cognitive operations (Thorndike & Woodworth, 1901). However, we did not observe transfer effects on every measure that we administered. First, as we predicted, there was no evidence of transfer from our specific training regime to the measure of general intelligence, the Raven's. Much has been made of study results showing that WM training can improve one measure of *IQ*, the Bochumer Matrizen Test (BOMAT; Jaeggi et al., 2008), due to the widely held belief that intelligence is relatively stable over the life span. However, this finding has proven difficult to replicate (Chein & Morrison, 2010; Dahlin et al., 2008b) and has become controversial due to the unconventional manner in which the BOMAT was administered (Moody, 2009). Specifically, Moody points out that in this case, the time restriction on the task (reduced from 45 minutes to 10 minutes) effectively eliminated the most difficult items from the test set insofar as participants were not able to progress through the test fast enough to be administered these items (i.e., the majority of subjects answered less than 14 out of 29 items correctly at posttest, 2009). Furthermore, this finding does not speak to the fact that participants who had received eight days of training show virtually no improvement on the Raven's, purportedly the easier of the two tests (Moody, 2009). Nevertheless, it will be important for future researchers to identify the key features of the training paradigm used by Jaeggi et al. (2008) that led to improvements on their IQ test, and to assess whether these effects generalize to other measures of intelligence, such as the Kaufman's Brief Intelligence Test (K-BIT) or Wechsler's Adult Intelligence Scale (WAIS). Some possible explanations for these differential findings with regard to IQ include differences in statistical power, length of training, or requirements of the training task. In regards to power, it is also possible that our sample size was too small to detect an improvement in IQ. However we think this unlikely given that WM trained older adults were numerically worse on the Raven's after training, and also, because we previously found no effect of WM training on IQ in young adults (Chein & Morrison, 2010). Also, the length of our training regime was similar to that used by Jaeggi and colleagues (2008), so this variable may not be critical. We feel that the most fruitful variable to investigate is the type of training task, given the recent finding that older adults who underwent executive control training showed improvements in fluid intelligence (Karchach & Kray, 2009).

We also obtained null results on subtests of the TEA, likely due to ceiling effects. Although the TEA is purportedly sensitive to attentional differences that accompany normal aging, our participants scored at, or near ceiling on three of the six TEA subtests at

pretest: elevator counting ( $M = 6.95$  out of 7 possible), visual elevator ( $M = 8.93$  out of 10 possible), and elevator counting with distraction ( $M = 8.83$  out of 10 possible). Note that mean raw scores collapsed over all participants' performance at pretest are reported here. Ceiling effects were difficult to assess on both telephone search subtests because the outcome measures consist of timing scores. Future researchers may want to consider including only the Elevator Counting with Reversal subtest of the TEA to elucidate the extent to which the skills used in this subtest may be influenced by WM training.

Training did not affect digit span performance. Digit span is known to be a relatively insensitive measure (Dickson et al., 1991) that has persisted in neuropsychology due to its ease of bedside administration. To our knowledge, only one study has reported that WM training specifically improves digit span in older adults (Mahncke et al., 2006), whereas others have reported null results on this measure (Buschkuhl et al., 2008; Dahlin et al., 2008b; Li et al., 2008). We would advise future investigators to think carefully before choosing to include this measure in their training studies.

### Optimizing Transfer in WM Training Studies

Much of the WM training literature has focused on transfer of training to other tasks rather than training gains in and of themselves. However, we are compelled to point out that "in the field of skill learning, transfer . . . from the trained task to even other very similar tasks is generally the exception rather than the rule" (Green & Bavlier, 2008). Of the training paradigms where transfer effects are frequently seen, training is typically complex, such as with video game training or athletic training (Green & Bavlier, 2008). While it is the belief of the research team that our training task was complex in nature, the intent when developing the task was to specifically train WM capacity. Perhaps in focusing on core features of WM, other cognitive processes that are encouraged by more complex tasks such as video game training, for example, were omitted from the design. If one's goal is rehabilitation of cognitive aging, we would recommend that complex, process-impure training procedures be used (such as in Schmiedek et al., 2010). It is obvious that training on complex tasks would bear favorably to Thorndike's rule of training on tasks with cognitive constructs that overlap with the transfer task (Thorndike & Woodworth, 1901).

It is also important to take into account that the training regime, although adaptive, was repetitive. While the to-be-remembered items increased or decreased as a function of participant performance, the tasks themselves were the same over each of the 20 sessions. The Yerkes-Dodson law (Yerkes & Dodson, 1908) predicts learning as a U-shaped function of arousal. It is possible that due to the repetitive nature of the task, participant arousal over the 20 sessions waned. Several participants reported that they disliked the WM training sessions, but they were simultaneously motivated and excited about participating in this study.

Related to this topic, there are a number of problems that may arise when running a study of this length. There is often severe subject attrition (Owen et al., 2010), which adds to the time and cost of conducting such an experiment, and can diminish statistical power. Several of the studies listed in Table 1 used relatively small

sample sizes and, accordingly, may have been underpowered (Buschkuhl et al., 2008; Dahlin et al., 2008a; Dahlin et al., 2008b), which could have contributed to the meager transfer effects. The small Ns reflect the fact that studies with training regimes are expensive in terms of subject payment and personnel time.

Participants may experience exhaustion and inattentiveness during the testing session and 'burn-out' across sessions. The ideal length of training is currently unknown and as can be seen in Table 1, varies widely across laboratories. A recent study in an older population by Schmiedek and colleagues (2010) used a lengthy (100 days) and complex training regime, and reported the most impressive transfer effects with regard to scope that have been demonstrated to date. It is, therefore, possible that additional training would have led to more robust transfer effects.

Many of our subjects were notably concerned about how they were performing. Research shows that participants who make self-evaluative thoughts in the midst of a task have more difficulty than those who report on-task thoughts (McVay & Kane, 2009). Minimizing the number of self-evaluative thoughts throughout testing may help participants perform at a higher level both on the trained task as well as transfer measures.

Lastly, due to the fact that the literature on WM training regimes and assessment of transfer effects is fairly new, a large number of transfer measures were included. Unfortunately, correcting for multiple comparisons in this and many other training studies would make it virtually impossible to detect improvement. However, our results are supported by the interaction effects on our ANOVAs and moderate effect sizes, which lead us to believe that these are true improvements. As the body of literature on WM training grows, future researchers should be able to choose transfer measures more carefully and hold their data to more stringent significance cut-offs.

In sum, there are several factors that potentially moderate the impact of WM training on transfer measures. In regards to the training regime, success is intertwined with participants' level of commitment and the cognitive challenge of the task. The former factor can be directly linked to the number of training sessions and the experiments' 'pleasure quotient,' which if correctly titrated, will sustain involvement and attention. More subjects and more training is always a safe bet, but the realities of conducting longitudinal research make this goal impractical. Also, older adults may require more training than younger adults to achieve the same levels of transfer. Limitations like these may impact the data insofar as small transfer benefits may be undetectable, and potentially larger transfer effects reduced. Nevertheless, we have shown here that transfer effects are observable in older adults after a modest amount of training (20 sessions) in a modestly sized training group ( $N = 21$ ).

### References

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839. doi:10.1038/nrn1201
- Baddeley, A., & Hitch, G. (1974). Working Memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.
- Buschkuhl, M., Jaeggi, S., Hutchinson, S., Perrig-Chiello, P., Sapp, C.,

- Muller, M., . . . Perrig, W. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging, 23*, 743–753. doi:10.1037/a0014342
- Chein, J., & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review, 17*, 193–199. doi:10.3758/PBR.17.2.193
- Cowan, N. (1999). An embedded process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). New York: Cambridge University Press.
- Dahlin, E., Neely, A., Larsson, A., Backman, L., & Nyberg, L. (2008a). Transfer of learning after updating training mediated by the striatum. *Science, 320*, 1510–1512. doi:10.1126/science.1155466
- Dahlin, E., Nyberg, L., Backman, L., & Neely, A. (2008b). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer and long-term maintenance. *Psychology and Aging, 23*, 720–730. doi:10.1037/a0014296
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Verbal Learning and Verbal Behavior, 19*, 450–466. doi:10.1016/S0022-5371(80)90312-6
- Delis, D., Kramer, J., Kaplan, E., & Ober, B. (1987). *California verbal learning test: Adult version manual*. San Antonio, TX: Psychological Corporation.
- Dickson, D., Crystal, H., Mattiace, L., Masur, D., Blau, A., Davies, P., . . . Aronson, M. (1991). Identification of normal and pathological aging in prospectively studied nondemented elderly humans. *Neurobiology of Aging, 13*, 179–189. doi:10.1016/0197-4580(92)90027-U
- Engle, R., & Kane, M. (2004). Executive attention, working memory capacity and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 102–134). New York: Academic Press.
- Engle, R., Kane, M., & Tuholski, S. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.
- Folstein, M., Folstein, S., & McHugh, P. (1975). Mini-Mental State: A practical method for grading cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198. doi:10.1016/0022-3956(75)90026-6
- Green, C., & Bavelier, D. (2008). Exercising your brain: A review of human brain plasticity and training-induced learning. *Psychology and Aging, 23*, 692–701. doi:10.1037/a0014345
- Hertzog, C., Kramer, A., Wilson, R., & Lindenberger, U. (2009). Enrichment effects on adult cognitive development. *Psychological Science in the Public Interest, 9*, 1–65.
- Higginson, C., Arnett, P., & Voss, W. (2000). The ecological validity of clinical tests of memory and attention in multiple sclerosis. *Archives of Clinical Neuropsychology, 15*, 185–204.
- Holmes, J., Gathercole, S., & Dunning, D. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*, F9–F15. doi:10.1111/j.1467-7687.2009.00848.x
- Jaeggi, S., Buschkuhl, M., Jonides, J., & Perrig, W. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, USA of the United States of America, 105*, 6829–6833. doi:10.1073/pnas.0801268105
- Kane, M., Hambrick, D., Tuholski, S., Wilhelm, O., Payne, T., & Engle, R. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217. doi:10.1037/0096-3445.133.2.189
- Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*, 978–990. doi:10.1111/j.1467-7687.2009.00846.x
- Katsnelson, A. (2010). No gain from brain training. *Nature, 464*, 1111–1111. doi:10.1038/4641111a
- Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlstrom, K., . . . Westerberg, H. (2005). Computerized training of working memory in children with ADHD—A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*, 177–186. doi:10.1097/00004583-200502000-00010
- Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24*, 781–791. doi:10.1076/jcen.24.6.781.8395
- Li, S., Huxhold, O., Smith, J., Schmiedek, F., Rocke, C., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer and maintenance. *Psychology and Aging, 23*, 731–742. doi:10.1037/a0014343
- Logie, R., Gilhooly, K., & Wynn, V. (1994). Counting on working memory in mental arithmetic. *Memory & Cognition, 22*, 395–410. doi:10.3758/BF03200866
- Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. (2009). Aging, training and the brain: A review and future directions. *Neuropsychology Review, 504*–522. doi:10.1007/s11065-009-9119-9
- Mahncke, H., Connor, B., Appelman, J., Ahsanuddin, O., Hardy, J., Wood, R., . . . Merzenich, M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences, USA of the United States of America, 103*, 12523–12528. doi:10.1073/pnas.0605194103
- McVay, J., & Kane, M. (2009). Conducting the train of thought: Working memory capacity, goal neglect and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 196–294. doi:10.1037/a0014104
- Moody, D. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37*, 327–328. doi:10.1016/j.intell.2009.04.005
- Morris, S., & DeShon, R. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*, 105–125. doi:10.1037/1082-989X.7.1.105
- Morrison, A., & Chein, J. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18*, 46–60. doi:10.3758/s13423-010-0034-0
- Noack, H., Lovden, M., Schmiedek, F., & Lindenberger, U. (2009). Cognitive plasticity in adulthood and old age: Gauging the generality of cognitive intervention effects. *Restorative Neurology and Neuroscience, 27*, 435–453.
- Owen, A., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A., . . . Ballard, C. (2010). Putting brain training to the test. *Nature, 465*, 775–778. doi:10.1038/nature09042
- Raven, J. (1976). *Standard progressive matrices: Sets A, B, C, D & E*. Oxford: Oxford Psychologists Press.
- Robertson, I., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The test of everyday attention*. Bury St. Edmunds: Thames Valley Test Company.
- Salthouse, T., Atkinson, T., & Berish, D. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132*, 566–594. doi:10.1037/0096-3445.132.4.566
- Schmiedek, F., Lovden, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience, 2*, 1–10.
- Small, S. (2001). Age-related memory decline. *Archives of Neurology, 58*, 360–364. doi:10.1001/archneur.58.3.360
- Thorell, L., Lindqvist, S., Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training

- and transfer effects of executive functions in preschool children. *Developmental Science*, *12*, 106–113. doi:10.1111/j.1467-7687.2008.00745.x
- Thorndike, E., & Woodworth, R. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*, 247–261. doi:10.1037/h0074898
- Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505. doi:10.3758/BF03192720
- Yerkes, R., & Dodson, J. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. doi:10.1002/cne.920180503

Received September 7, 2010

Revision received March 2, 2011

Accepted March 8, 2011 ■