

Do the Effects of Working Memory Training Depend on Baseline Ability Level?

Jeffrey L. Foster
Western Sydney University

Tyler L. Harrison, Kenny L. Hicks, and
Christopher Draheim
Georgia Institute of Technology

Thomas S. Redick
Purdue University

Randall W. Engle
Georgia Institute of Technology

There is a debate about the ability to improve cognitive abilities such as fluid intelligence through training on tasks of working memory capacity. The question addressed in the research presented here is who benefits the most from training: people with low cognitive ability or people with high cognitive ability? Subjects with high and low working memory capacity completed a 23-session study that included 3 assessment sessions, and 20 sessions of training on 1 of 3 training regiments: complex span training, running span training, or an active-control task. Consistent with other research, the authors found that training on 1 executive function did not transfer to ability on a different cognitive ability. High working memory subjects showed the largest gains on the training tasks themselves relative to the low working memory subjects—a finding that suggests high spans benefit more than low spans from training with executive function tasks.

Keywords: working memory capacity, cognitive training, fluid intelligence

It is also possible that training of WM [working memory] will be useful in other conditions in which WM deficits are prominent. (Klingberg et al., 2005, p. 185)

Over the past decade, many studies have been conducted testing whether working memory training can improve broad cognitive abilities such as fluid intelligence (Gf) and attention control (Shipstead, Redick, & Engle, 2010, 2012). One of the reasons behind the interest in this research was that working memory capacity (WMC) and Gf are highly related at the construct level (Kane et al., 2004). Some researchers have inferred a causal relationship with individual differences in WMC being responsible for individual differences in Gf (Jaeggi et al., 2008). According to this logic, if working memory training could improve WMC, this should lead to improvements in Gf and all of the myriad real-world cognitive tasks which have been shown to be strongly related to WMC (Engle & Kane, 2004). The quote at the beginning of this article highlights the potential importance of working memory training benefitting subjects with WMC deficits. Klingberg and colleagues (Klingberg, Forssberg, & Westerberg, 2002; Klingberg

et al., 2005) have suggested that children with ADHD can improve their WMC through a commercial working memory training program (i.e., Cogmed) and that these improvements would ameliorate the symptoms of ADHD. Other studies have suggested that populations of subjects with low levels of WMC such as older adults (Richmond, Morrison, Chein, & Olson, 2011), children with learning disabilities (Alloway, Bibile, & Lau, 2013), and stroke patients (Westerberg et al., 2007) can all benefit from working memory training. Although some previous research suggested that working memory training may be effective in improving performance on higher-order tasks, researchers have not addressed one important question: Do individuals with higher and lower levels of WMC benefit equally from working memory training?

Working Memory Training

WMC is the ability to maintain and use information in the face of distraction (Engle, 2002). This ability is heavily reliant on the ability to control attention (Engle & Kane, 2004). One of the most common methods of measuring WMC is with complex span tasks such as the operation span (Conway et al., 2005; Turner & Engle, 1989; Unsworth, Heitz, Schrock, & Engle, 2005). In this task, subjects must alternate between the presentation of to-be-remembered letters and solving simple math problems. Importantly, this measure indicates the ability to retain information (letters) while completing a secondary distractor task (math problems) that stops subjects from attending to and rehearsing the letters. The reason that many researchers have studied WMC and why many cognitive training programs have targeted WMC is because of its relationship to a multitude of important cognitive

This article was published Online First May 29, 2017.

Jeffrey L. Foster, School of Social Sciences and Psychology, Western Sydney University; Tyler L. Harrison, Kenny L. Hicks, and Christopher Draheim, School of Psychology, Georgia Institute of Technology; Thomas S. Redick, Department of Psychological Sciences, Purdue University; Randall W. Engle, School of Psychology, Georgia Institute of Technology.

Correspondence concerning this article should be addressed to Randall W. Engle, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA. E-mail: randall.enge@gatech.edu

abilities and activities such as Gf (Kane et al., 2004; Kyllonen & Christal, 1990), multitasking (Hambrick, Oswald, Darowski, Rench, & Brou, 2010), following directions (Engle, Carullo, & Collins, 1991), and programming a computer (Shute, 1991). The logic behind working memory training is that, if we can improve WMC, we should be able to improve abilities that are highly related to WMC. This logic is not sound, however, because correlation does not imply causation.

A critical issue in any type of cognitive training is the extent to which the training transfers to performance on other tasks. Barnett and Ceci (2002) presented a taxonomy on the various variables that describe the extent of transfer. For instance, whether the transfer tasks are performed at the same location as the training tasks, the time between training and transfer, and the surface similarities of between training and transfer tasks. Training researchers typically use a greatly simplified version of Barnett and Ceci's taxonomy and talk about transfer as being either near or far. For researchers in the working memory training literature, evidence for near transfer occurs when training improves performance on a task that is structurally similar to the training task. Far transfer occurs when training improves performance on a task that has no structural similarities to the training tasks and measures a cognitive ability different than the one trained.

There is still a debate as to whether working memory training works. The key issue is centered on whether working memory training produces far transfer, particularly to Gf (Au, Sheehan, Tasi, Duncan, Buschkuehl, & Jaeggi, 2014; Harrison et al., 2013). Some studies have suggested that working memory training leads to far transfer (e.g., Chein & Morrison, 2010; Diamond & Lee, 2011; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Klingberg, Forssberg, & Westerberg, 2002). However, there are many studies that show no evidence of far transfer (e.g., Chooi & Thompson, 2012; Harrison et al., 2013; Redick et al., 2013; von Bastian & Oberauer, 2013). A recent meta-analysis has made sense of the discrepancy in the literature (Melby-Lervåg, Redick & Hulme, under review). The studies that have a no-contact control group show far transfer and the studies which use an active control group (i.e., a control group that completes a training task that is equally demanding but unrelated to WMC) do not show far transfer. From this meta-analysis, it seems likely that Hawthorne-type expectancy effects may be a major factor in whether a study finds evidence of far transfer.

Who Benefits the Most From Training?

Numerous companies have developed and marketed software aimed at improving cognitive abilities by training WMC with some being sold for use in public education (see Shipstead, Hicks, & Engle, 2012a, 2012b). Although the question of the effectiveness of working memory training is still controversial (Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Morrison & Chein, 2011) one variable potentially important to whether WM training is effective has not been studied. It is possible that high- and low-ability individuals differ in the benefit of WM training.

The question dealt with in the present study is whether people of different ability levels will show differences in the effects of working memory training. As mentioned earlier, working memory training has been shown to be beneficial for individuals with

WMC deficits. In fact, Jaeggi et al. (2008) stated that "there was also a main effect of performance group . . . showing that subjects with initially lower Gf generally showed even larger gains in Gf" (p. 6830). Likewise, one commentary on the Jaeggi et al. paper claimed the differential improvement for people with low Gf was an important feature of their paper (Sternberg, 2008). In a meta-analysis, Au and colleagues (Au et al., 2015) found a marginally significant negative correlation between people's initial ability levels with an n-back training task, and the amount of improvement on transfer tasks: Leading the researchers to conclude that "those who start with more room to improve . . . may also gain the most" (Au et al., 2015, p. 375). Other researchers make similar claims that using interventions like working memory training will close the gap between high- and low-ability students (see Diamond & Lee, 2011).

However, some have found evidence that suggests those with initially higher abilities can gain the most from training. For example, Swanson (2014) found that training verbal and visuospatial strategies aided high WMC children more than low WMC children, and Fuchs et al. (2014) found that high WMC subjects benefitted the most from fluency-based training on solving math problems, whereas low WMC subjects benefitted the most from more basic conceptual training. Taken together, these differences in findings from a variety of research offer only competing hypotheses for who may be more likely to benefit from training on WMC tasks.

Method

Subjects

A total of 116 subjects completed all 23 sessions of the study.¹ These subjects included both university students ($n = 84$) and community members ($n = 32$) and were all aged 18–35. Subjects were recruited from one of two previous studies that measured their individual WMC using three complex span tasks. We recruited only subjects for whom a composite score (average of z-scores) across the three complex span tasks found them in the top third of scores (high spans; $N = 59$) or the bottom third of scores (low spans; $N = 57$). Low spans tended to be older ($M_{\text{Low}} = 24.56$, $M_{\text{High}} = 19.90$, $t(114) = 6.56$, $p < .01$), and less likely to be a university student than high spans ($N_{\text{Low}} = 27$, $N_{\text{High}} = 57$, $\chi^2(1) = 35.19$, $p < .01$). Subjects were compensated with a base-pay of \$300 for completing all 23 sessions. In addition,

¹ In addition to the 116 subjects that successfully completed all stages of the study, an additional 38 subjects (25%) completed the pretest assessment and at least one day of training, but failed to complete all 23 sessions of the study. Many of these subjects ($N = 14$) completed only one or two days of training, or dropped out immediately after the mid test ($N = 6$) but, on average, these subjects completed 5.7 training sessions ($SD = 4.20$). More of these subjects were low spans ($N = 23$) than high spans ($N = 15$), but overall dropout rates were similar between low spans and high spans who completed at least 1 training day, $\chi^2(1) = 1.49$, $p = .22$. Subjects who were recruited but did not complete any training sessions were never assigned a training condition and are not included in these numbers. The total number of high spans and low spans that completed all training and assessment tasks was 116 (complex span training: $n_{\text{high}} = 20$, $n_{\text{low}} = 20$; running span training: $n_{\text{high}} = 20$, $n_{\text{low}} = 19$; visual search training: $n_{\text{high}} = 19$, $n_{\text{low}} = 18$).

subjects earned bonus pay during the training sessions. They received \$4 for every new level achieved for the first time, and \$2 for not decreasing their level during a given session. We chose this bonus structure to ensure subjects were motivated to perform well during the training sessions. All bonuses were paid at the completion of the study, while the base pay was distributed across the three assessment sessions.

Materials and Procedure

The present study consisted of 23 sessions: an initial pretest assessment session, 10 training sessions, a midtest assessment, 10 additional training sessions, and a final posttest assessment session. During each of the 20 training sessions, subjects completed two training tasks in one of three conditions: complex span training, running span training, or visual search (active control) training.

Training Sessions

In each of the 20 training sessions subjects completed two related tasks. One task had a verbal component, and another had a spatial component. Each task was setup to have eight blocks of trials per training session with each block of trials corresponding to a level of difficulty. If the subject had higher than 85% accuracy on a given block of trials, they advanced to the next level. If they scored below 70% accuracy, they decreased to the previous level. If they scored between 70% and 85% they remained on the same level (see Harrison et al., 2013; Jaeggi et al., 2008; Redick et al., 2013). Subjects in the complex span training and running span training conditions always completed the verbal task first, and the spatial task second. Both tasks in the visual search training condition were visual, and subjects always completed the hands task first, and the letters task second.

Each training session lasted approximately 20 to 75 min.² During the training sessions, each subject completed two training tasks in one of three conditions: complex span training, running span training, or a visual search (active control) training. For all tasks, subjects were given full instructions on how to complete the task, including suggestions about strategies, during the first training sessions. The remaining 19 training sessions gave only a brief reminder of the task and potential strategies to use.

Subjects were provided with a range of possible strategies they could use to improve their scores. Importantly, these strategies were merely suggestions and subjects were explicitly told “The technique you use is entirely up to you.” For both the complex span and running span training conditions, suggestions were focused on visualization of material, such as “When you see the letter ‘T’ you could imagine a turtle” for the verbal tasks or “You could also try forming shapes out of the locations on the 4 × 4 grid” for the symmetry span task. These instructions would not aid in the visual-search training condition, and those instructions suggested not looking too hard, and allowing the search item to jump-out at you—a suggestion based on anecdotal data from subjects in a previous study. These strategies were included to ensure all subjects had a minimum level of knowledge about how to improve their scores, but were included merely as a suggestion so that subjects did not feel obligated to use any of the suggested strategies.

Complex span training. Subjects in the complex span training condition trained on two modified complex span tasks:

operation span and symmetry span. In the operation span, subjects were first presented with a simple math problem to solve, followed by a to-be-remembered letter (see Panel A of Figure 1). The number of math problems and to-be-remembered letters increased with each level of difficulty. After viewing all of the letters, subjects were asked to recall the letters in the order they were presented. In the symmetry span, subjects were shown shapes and responded with whether the shape was symmetrical along its vertical axis. After each symmetry judgment, subjects were shown a red square in a 4 × 4 grid of locations and were asked to recall the locations of each red square in the order they were presented. Each block of the training tasks was setup as levels with each level containing three trials made up of one trial each with the number of distractor-memory pairs equal to the level number (i.e., three pairs), one smaller (two pairs), and one higher (four pairs). Both the choice of task and this varying trial number was set to maintain consistency with prior research (see Harrison et al., 2013).

Running span training. The running span training condition used training on two running span tasks: running letters and running locations. These tasks used the same to-be-remembered stimuli as used in the complex span training. In the running letters task, subjects were told they would have to remember the last X number of letters with X increasing with each level of difficulty. They then saw letters, one at a time, briefly flashed on the screen. The number of letters seen varied with each trial, but was never fewer than subjects were told to recall. For example, a subject may need to recall the last four letters shown, but actually see seven letters. This task requires subjects to update their memory with new letters as they appear. Again, each block contained three trials, but unlike the complex span training each trial contained the same number of to-be-recalled items (i.e., three trials of recalling the last three items) to avoid updating the number of to-be-recalled items on every trial. The running locations task followed the same methods as the running letters task, but used the location of a red square in a 4 × 4 grid rather than letters.

Visual search training. The visual search training condition acted as our active-control group and consisted of two tasks: Visual search letters and visual search hands (see Panel B of Figure 1). We used a visual search task as our active control since visual search time is unrelated to WMC (Kane, Poole, Tuholski, & Engle, 2006). In the visual search letters task, subjects saw a grid of E’s and T’s facing in a variety of directions, along with a single F that faced either left or right. Subjects saw this grid of letters for 500 ms, followed by a 500-ms mask. Their task was to identify whether the F was facing left or right. The size of the grid increased with each level. The visual search hands task followed the same methods,

² In the training sessions, variations in the time to complete were due to both the training task being completed, and the difficulty level of the task. At the easiest levels, each training session took approximately 25 min to complete. As the difficulty of the tasks increased the time to complete each session separated between the three training conditions with the complex span training condition taking the longest, the running span training condition taking the second longest, and the visual search training condition being the shortest. The number of blocks across all 20 training sessions in all three conditions remained constant.

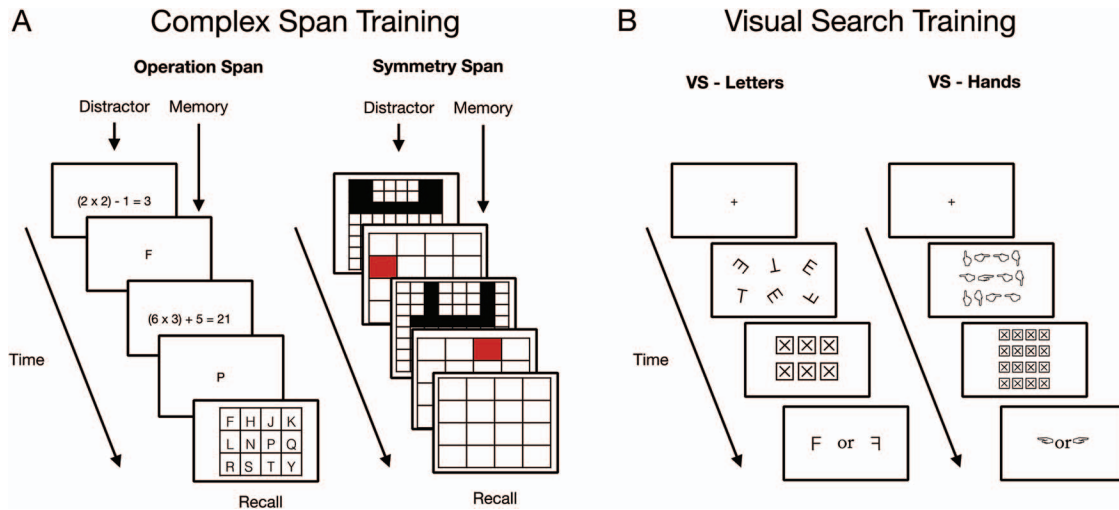


Figure 1. Two sample trials from each of the complex span training (A) and visual search training (B) tasks. See the online article for the color version of this figure.

but used hands pointing up, down, left, or right instead. Importantly, one hand appeared with a profile view (as opposed to a palm-view) of the hand, and the subject's task was to determine whether that profile hand was pointing left or right. For the visual search task, each block contained 24 trials. With each level increase one new row or column was added interchangeably to increase difficulty. For example, at the starting level, a 2×2 grid was used, the next level up used a 3×2 grid, and the next level after that used a 3×3 grid.

Assessment Sessions

Each assessment session lasted approximately 2–3 hours³ during which each subject completed 13 tasks measuring transfer effects due to training. The assessment tasks were selected to measure eight cognitive abilities: WMC, running span ability, attention control, primary memory (PM), secondary memory (SM), memory updating ability, Gf, and multitasking ability. The first two represent near transfer, whereas the last two represent far transfer and the remaining can be classified as moderate transfer (Harrison, Shipstead, & Engle, 2014).

WMC. To assess changes in WMC, two different complex span tasks were administered in the pretest, midtest, and posttest: a modified reading span task, and the rotation span task. Figure 2 demonstrates a sample item from each of these tasks. All of the complex span trials were randomized, so no counterbalancing took place between assessment sessions.

Modified reading span. In the standard reading span task, subjects first read a sentence (e.g., “Andy was stopped while he crossed the pizza”) and respond with whether the sentence makes sense. They then see a to-be-remembered letter, followed by another sentence, and another to-be-remembered letter (Daneman & Carpenter, 1980; Unsworth, Heitz, Schrock, & Engle, 2005). In the modified reading span task used here, the letters were replaced with line drawings of common objects (e.g., a cup) to keep the to-be-remembered items different from the to-be-remembered items used in the training tasks. The

number of sentence/image pairs ranged from three to 11 per trial, and subjects were asked to recall the images in the order they were presented. Scores were calculated by summing the total number of correctly recalled to-be-remembered items—also known as the partial scoring method (Conway et al., 2005; Turner & Engle, 1989).

Rotation span. In the rotation span task, subjects first saw a letter presented either normally, or mirrored, which was rotated on its vertical axis (Kane et al., 2004). Their task was to determine whether the rotated letter was presented normally or mirrored. They then saw an arrow pointing in one of eight directions, and of short or long length. The number of letter/arrow pairs ranged from 2 to 9 per trial, and subjects recalled the arrows in the order they were presented. Scores were calculated using the partial scoring method.

Running span ability. To measure running span ability, two running span tasks were used: running span images and running span arrows. Running span tasks have been used widely in the executive function literature as a measure of WMC and, more recently, as an indicator of memory updating (Broadway & Engle, 2010; Dahlin et al., 2008). The same to-be-remembered items from Figure 2 were used for these tasks. In both tasks, subjects were informed—before each block of trials—precisely how many of the items they will need to remember. However, they were not told how many items would be displayed. To perform well at this task, subjects must update the items in memory by dropping off old items as new ones appear. In both tasks, the number of to-be-recalled items ranged from 3 to 11. Scores were calculated using the same partial scoring method used with the WMC tasks. All of the running span trials were randomized, so no counterbalancing took place between assessment sessions.

³ The variations in time during the assessments were due mostly to familiarity with the tasks; later assessments tended to be completed faster than earlier assessments.

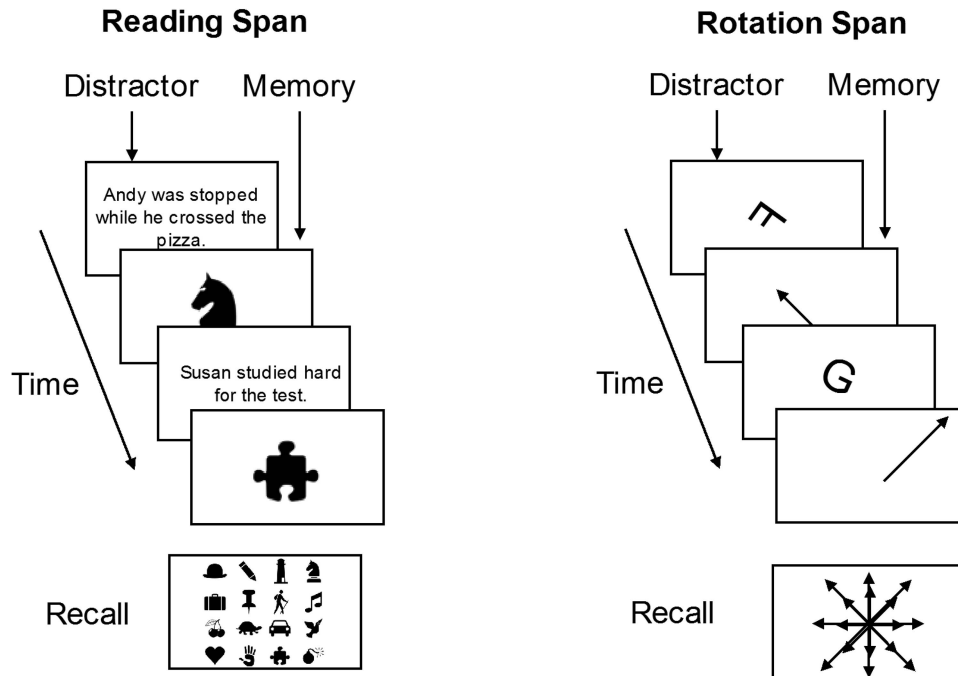


Figure 2. Two sample trials from the reading span (left) and rotation span (right) tasks.

Attention control. Two tasks were used to measure transfer effects in attention control—the antisaccade task and a visual arrays task that requires subjects to quickly focus on a specific type of item in a visual array. All of the attention control task trials were randomized, so no counterbalancing took place between assessment sessions. These two tasks are demonstrated in Figure 3.

Antisaccade. The antisaccade task used here followed the methods of Kane et al. (2001) and Unsworth et al. (2004). In this version of the antisaccade task subjects are first shown a fixation point in the center of the screen and are then presented with a brief star-shaped stimulus that flashes on either the left or right side of the screen (Kane et al., 2001; Unsworth et al., 2004). The subject has to then look quickly toward the opposite side of the screen to see a letter stimulus (B, P, or R). The letter presentation is very brief and masked afterward. The subject indicates which letter they saw with the keyboard. Scores are calculated by summing the total number of correct responses.

Controlled attention visual arrays. In this visual arrays task, subjects are first shown the name of a color (e.g., BLUE) very briefly (Vogel, McCollough, & Machizawa, 2005). Next, either 10 or 14 colored rectangles (five or seven of each color) appear rotated at various angles in 20° increments for 500ms. After a 500ms blank screen, one of the rectangles (in the color previously indicated) reappears. The subject's task is to determine whether the angle that the rectangle appears in is the same as previously shown, or whether it has been rotated. Scores are calculated by summing the total number of correct responses. Although the visual arrays task was originally developed as a measure of WMC (Luck & Vogel, 1997), recent evidence suggests that this specific visual arrays task is a good indicator of attention control (Shipstead, Lindsey, Marshall, & Engle, 2014).

PM and SM. To measure PM and SM, subjects completed two immediate free recall (IFR) tasks. Three versions of each task (IFR Words and IFR Pictures) were made with the words and pictures used for recall changed between each version. Each subject completed each version across the three assessment sessions, but the order in which each version was completed was counterbalanced across subjects.

Each version consisted of six trials containing 12 words or pictures. In each trial the 12 items were presented one at a time for 1,000 ms. After all 12 items were presented, subjects heard a beep indicating that they should immediately begin recalling the presented items (in the pictures task, subjects were instructed to recall a name for the picture—e.g., a horse). Following the procedure of Dalezman (1976), subjects were informed before beginning that they should first recall the items from the end of the list, before recalling earlier items. This instruction is used to ensure that subjects recall items from PM before they are lost. Tulving and

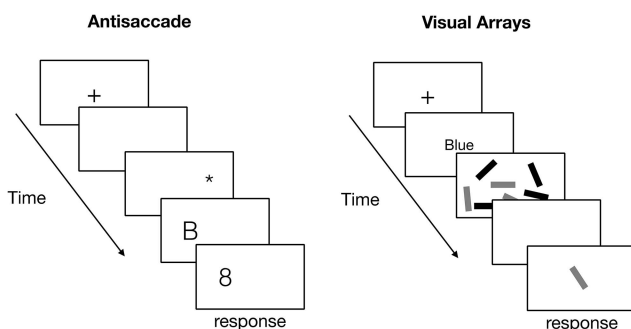


Figure 3. Two sample trials from the antisaccade task (left) and visual arrays task (right).

Colotla's (1970) procedure was used to calculate PM and SM scores for each trial.

Memory updating ability. Research on individual differences in cognitive ability has investigated the importance of memory updating ability (Morris & Jones, 1990; Shipstead et al., 2017)—that is, the ability to forget no-longer-important information in order to focus on newly important information: like forgetting your old phone number so that it doesn't interfere with remembering your new phone number. Importantly, some evidence suggests that memory updating ability may account for a large portion of the relationship between WMC and Gf (Shipstead et al., 2017). If this is true, then it would suggest that the ability to increase memory updating ability with training would be critical in showing a causal relationship between training on WMC tasks and increases in Gf.

Keep track. In the keep track task (Yntema & Trask, 1963), subjects are shown a list of words one-at-a-time and asked to remember the most recently presented instance of a word in a specific category or categories. For example, subjects may be asked to remember the most recent relative and country categories, and then see, one at a time: Aunt, Germany, Blue, Mile, Uncle. They were then presented with a list of possible words, and asked to recognize the most recent instance of those categories (Germany and Uncle). During the presentation, words fitting into one of 6 categories are shown, and subjects are asked to remember the most recent instance of anywhere from 1 to 5 categories. Each session used the same set of 6 categories of words, but which category set was used was counterbalanced between each assessment session. Scores are calculated by summing the number of correctly recalled words at the end of each trial.

Trail making. In the trail making task (modified from Ricker & Axelrod, 1994), subjects are given a starting letter-number pair (e.g., K-43). The words LETTER or NUMBER are then presented one at a time for 1 s each. The subjects' task is to increase the starting pair by one letter (L) or one number (44) each time they see LETTER or NUMBER respectively. Scores are calculated by summing the total number of correctly recalled pairs at the end of each trial. Trials in the trail making task were randomized, so no counterbalancing took place between assessment sessions.

Gf. We used two measures of Gf to measure far transfer effects to intelligence in training: Matrix reasoning and paper folding.

Matrices. In a typical matrix problem, subjects see a 3×3 grid of shapes with a blank space in the bottom right corner (Raven, Raven, & Court, 1998). The progression of shapes fit a logical pattern, and the subject's task is to determine what shape—from a list of 8 possible options—logically fits in the bottom right corner. Scores are calculated by summing the number of correct answers. Two sets of matrices were used for this study. In the first set of matrices, six items from the Raven's even set were used. The even set of Raven's consists of 18 items ordered in increasing difficulty. These 18 items were split into three subsets (e.g., Items 1, 4, 7, 10, 13, and 16 constituted subset 1) with one subset presented in each session with the order of each subset counterbalanced across subjects. The second set of matrices tasks consisted of three, 10-item, matrix sets created and normed in a separate study, with the presentation of each set counterbalanced across subjects (Hicks,

Foster, & Engle, 2016). Subjects were told that they would have 5 min and 6 min to complete each task respectively. Scores were calculated by summing the total number of correct answers across both tasks.

Paper folding. In the paper folding task, subjects are shown a line drawing of a piece of paper being folded one to three times (Ekstrom, French, Harman, & Dermen, 1976). A circle is then shown representing a hole being made through the entire thickness of the folded paper. The subjects' task is to choose—from a set of five possible answers—what the pattern of holes will look like after the paper is unfolded. Scores are calculated by summing the total number of correct answers. Similar to the matrices problems, two sets of paper folding tasks were used. The first set consisted of the 30 items from Ekstrom et al. These items were split into three subsets and counterbalanced across the three assessment sessions. The second set of tasks consisted of three 10-item folding sets created and normed in a previous study (Hicks et al., 2016). Scores were calculated by summing the total number of correct answers.

Multitasking. For our multitasking task, we used synthetic work.

Synthetic work. Figure 4 demonstrates the screen layout of the synthetic work multitasking task (Elsmore, 1994). Prior to completing synthetic work, each subject read detailed instructions on how the task is completed. They then spent 10 min on the synthetic work task. In synthetic work, subjects must perform four separate, simultaneous, tasks in the four quadrants of the screen shown in Figure 4: (a) a memory task where they are briefly shown a sequence of letters (they are later probed with a letter and they must decide whether the letter was or was not in the original sequence); (b) a math task where they solve simple math problems; (c) a fuel gauge task where they need to click on the gauge when it is low, but before it reaches 0, and (d) a beep task, where an auditory beep signals them to click a red ALERT button. The subject's task is to keep all four components of the synthetic work consistently updated. Failure to do so reduces the overall score. By default, synthetic work will output negative scores for subjects who failed to accurately perform on the task, and while the top end of positive scores reached 865, the low end of the negative scores could disproportionately drop to as low as $-2,760$. To correct for this skew, all scores were truncated to a minimum score of 0 to reflect a general failure at the task.

Results and Discussion

Our primary question was whether high spans or low spans would demonstrate the highest gains as a result of training. We asked this question in two ways: First, do we find group differences in the improvements across the trained tasks themselves?, and second, do we find group differences in the improvements across the assessment tasks? In addition to this primary question, we also wanted to know what near and far transfer benefits occur across the two different cognitive training domains, and whether there are differences in gains between verbal and visuospatial assessment tasks. We begin by analyzing the data from the training tasks in the three conditions (complex span training, running span training, and visual search training) before analyzing the data from the assessment tasks.

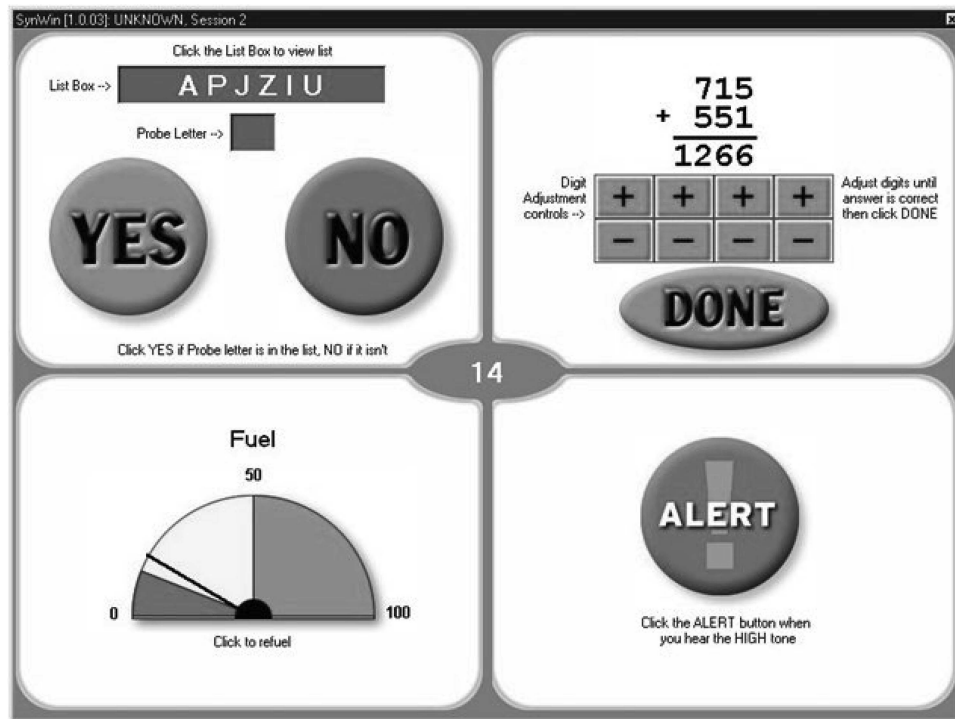


Figure 4. Screenshot from Elsmore's (1994) synthetic work task. Subjects complete four tasks simultaneously to receive points (the number located in the center). Clockwise from top left: list recall—subjects briefly see a list of letters, and are later probed with a letter and must respond with whether the letter was present in the list; addition—subjects solve simple math problems; visual monitoring—subject click the gauge before the needle reaches 0, with more points gained with the needle closest to 0; auditory monitoring—Subjects click the ALERT button when hearing a beep.

Training Tasks

We used the average level for each task as the subject's score for that day. These scores on each training task varied as a result of changes in difficulty between each task. That is, some tasks are more difficult than others, and the difficulty of a specific level number on one task does not necessarily reflect similar difficulty for the same level on another task. To adjust for this difference, we standardized scores for each task by converting scores to Z-scores based on the mean and standard deviation of all subjects completing a given task on the very first day of training. Therefore, the average z-score of high spans and low span on the first day of training will always be zero, with a standard deviation of 1, and will increase with increased performance relative to the standard deviation of the first day. Scores for the two training conditions (complex span and running span training) were calculated using the total number of to-be-remembered items in the task, while the visual-search training scores were calculated by the size of the search array: [(total number of rows) + (total number of columns)]. Although the scores have been standardized for each task, the nature of the tasks themselves still differ, and as such we conducted six 2 (span group) \times 20 (training session) multivariate analysis of variance, one for each training task. We present these data in Figure 5, along with a third line representing the trending difference between the two means—more specifically, it repre-

sents the size of the gap between high spans and low spans (the difference score on that training day). This trend line represents the three possibilities in the training results: A flat line would suggest similar gains in the training tasks between high spans and low spans, while a line with a negative slope would suggest a decreasing gap between the two groups (low spans showing larger improvements), and a line with a positive slope an increasing gap (high spans showing larger improvements).

The bottom two graphs of Figure 5 show the relative gains of high spans (solid line) and low spans (dashed line) for the two visual search—active control—tasks. The bottom graphs demonstrate several key points about subjects in the active-control condition for both the VS-hands and VS-letters tasks. First, a main effect of span demonstrates that high spans still showed slightly higher scores overall than low spans for the VS-hands task and faired marginally better on the VS-letters task, $F(1, 35) = 16.13, p < .01, \eta_p^2 = .31$, and $F(1, 35) = 3.41, p = .07, \eta_p^2 = .09$, respectively. Second, a main effect of training day shows that both high spans and low spans demonstrated large improvements in the visual search tasks across the 20 training sessions, $F(19, 17) = 67.49, p < .01, \eta_p^2 = .99$, and $F(19, 17) = 20.84, p < .01, \eta_p^2 = .96$, respectively. Importantly though, the flat gray lines representing the differences between the means demonstrate that there was no difference in the rate of improvement for either high spans or low spans, as represented by no

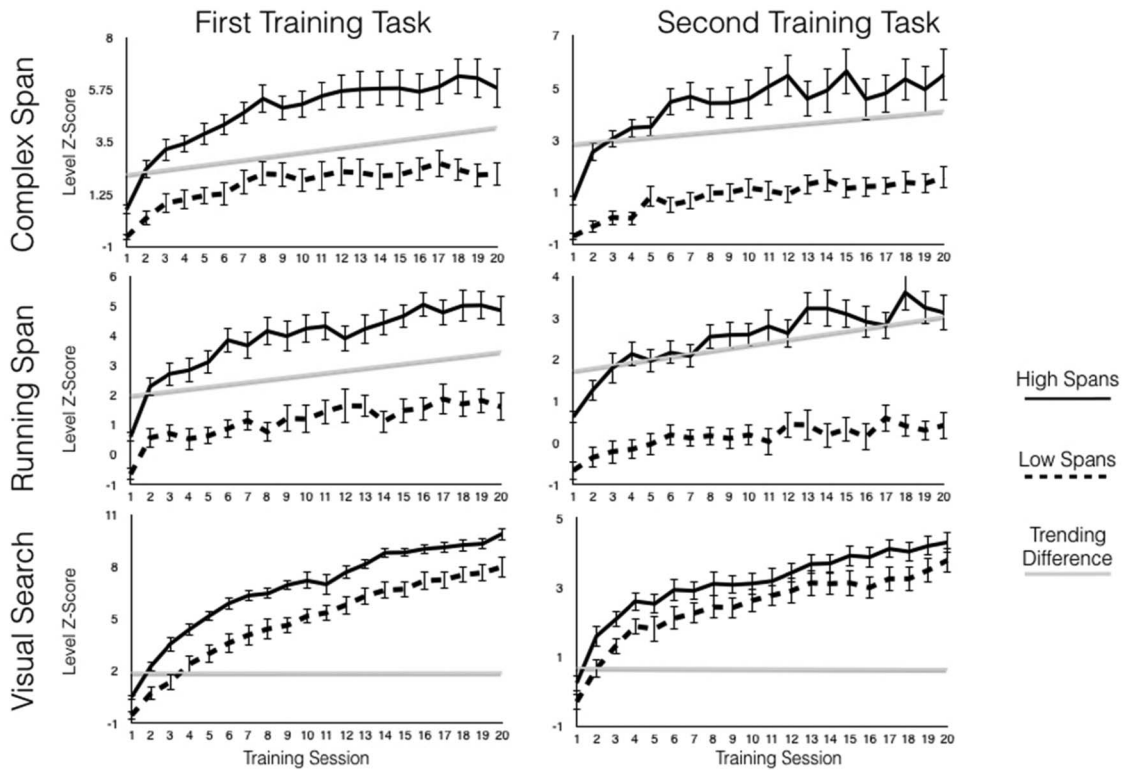


Figure 5. Results of the 20 training sessions by training condition, task number, and span. Scores were calculated by taking the mean and *SD* of all scores on the first training day, and then calculating the number of *SDs* each person scored above the mean on first day of training (compared to the whole group). The solid gray line represents the trending difference between the two means—more specifically, it represents the size of the gap between high spans and low spans.

statistically Significant Span \times Training Day interaction, $F(19, 17) = 0.69$, $p = .78$, $\eta_p^2 = .44$, and $F(19, 17) = 0.53$, $p = .91$, $\eta_p^2 = .37$, respectively.⁴

These findings are important for two reasons. First, the steady performance growth across the 20 training sessions for all three conditions suggests that our payment bonus structure was effective in motivating both high spans and low spans in doing their best and improving over time. Second, the lack of any increasing or decreasing gap in the visual-search ability between high spans and low spans suggests that there was little if any difference in the motivation to improve between the two span groups. That is, low spans were no more likely to be inattentive or uninterested during the training tasks, which suggests a similar level of determination for both span groups.

The middle row of Figure 5 shows the relative gains of high spans and low spans for the two running span training tasks. A main effect of span shows that, as expected, high spans outperformed low spans across the board with a relatively large difference between the two groups on both the letter running span and the location running span, $F(1, 37) = 28.77$, $p < .01$, $\eta_p^2 = .73$, and $F(1, 37) = 41.89$, $p < .01$, $\eta_p^2 = .53$, respectively. In addition, both groups of subjects increased their performance on the trained tasks over the 20 training sessions, as demonstrated by a main effect of training day, $F(19, 19) = 16.57$, $p < .01$, $\eta_p^2 = .94$, and $F(19, 19) = 5.99$, $p < .01$, $\eta_p^2 = .86$, respectively. Importantly

though, the positive slope of the line representing the differences between the two groups—along with the Span \times Training Day interaction—tells us that high spans demonstrated larger training gains than low spans on the letter running span, $F(19, 19) = 2.71$, $p = .02$, $\eta_p^2 = .73$, although this same effect did not materialize for the spatial running span, $F(19, 19) = 0.91$, $p = .58$, $\eta_p^2 = .73$. Although the spatial running span showed no statistically significant increase in the difference between the two means, the positive slope of the trend lines suggests that this null result may be due more to the increasing variance in each group over time, than the lack of an effect. Overall, these findings tell us that training on these tasks widens the gap between high-ability subjects and low-ability subjects on the trained task: a finding that suggests high spans benefit more from working memory training than low spans.

⁴ The finding that high spans generally outperformed low spans on the visual search task when measuring accuracy may seem inconsistent with the notion that visual search time is unrelated to working memory capacity (Kane et al., 2006). However, unlike Kane and colleagues, we measured accuracy in a rapid-presentation task instead of search time, which likely led to this main effect of working memory capacity. Importantly though, the lack of an interaction between span group and time would suggest the mechanism driving this small difference between spans was not related to how well high spans and low spans could train on the tasks.

Turning to the complex span training condition on the top line of Figure 5, we find a pattern of results nearly identical to those of the running span training condition. Once again, a main effect of span demonstrates that high spans outperformed low spans across the board for both the operation span and symmetry span training tasks, $F(1, 38) = 20.40, p < .01, \eta_p^2 = .43$, and $F(1, 38) = 31.04, p < .01, \eta_p^2 = .45$, respectively. In addition, the main effect of training day tells us that subjects improved on the training task over time, $F(19, 20) = 10.79, p < .01, \eta_p^2 = .91$, and $F(19, 20) = 5.56, p < .01, \eta_p^2 = .84$, respectively. The Span \times Training Day interaction failed to reach statistical significance for the operation span task, $F(19, 20) = 1.54, p = .17, \eta_p^2 = .09$, but the same interaction did reach statistical significance for the symmetry span task, $F(19, 20) = 3.50, p < .01, \eta_p^2 = .77$. Again, although the interaction term for one task failed to reach statistical significance, an inspection of the trend lines in both figures of Panel C demonstrates again that training on a given task widens the gap between high-ability subjects and low-ability subjects.

Overall, these findings demonstrate that while all people can be motivated to perform well on training tasks, it is those who already have high WMC that show the largest gains from using the training tasks. A finding that suggests any transfer effects found with the working memory training should be larger for high spans than for low spans.

These findings also shed light on the number of training sessions that are beneficial. That is, they can demonstrate the optimal number of training sessions before improvements begin to drop off to a rate that is statistically nonsignificant when compared to performance on the final day of training. For the sake of being liberal with these tests, we next collapsed across the two training tasks and the two span groups to conduct a 3 (training condition) \times 20 (training day) analysis of variance (ANOVA) and used uncorrected pairwise comparisons on the 20th day of training to the remaining 19 days. What we found is that subjects continued to improve to the very last day of training in the visual search task (Day 20 compared to 19; $M_{diff} = .33, p < .01$). However, in the running span training condition subjects did not significantly improve on their performance on the 13th training day (20 compared to 13; $M_{diff} = .13, p = .25$), and in the complex span training condition subjects did not significantly improve on their performance from the 11th day on (20 compared to 11; $M_{diff} = .35, p = .18$). It is important to note that these findings do not tell whether training after these days had any benefit on the transfer tasks, but it does suggest that the benefits of training past these time points should be suboptimal compared to earlier benefits, and that for the most part, any assessment effects we find should appear on the midtest after 10 training sessions, with little if any additional gains from the midtest to the posttest.

Assessment Tasks

Across all subjects, the scores for 105 assessment tasks were not included due to either experimenter error, or computer crashes. These 105 tasks represent 2.01% of the 5,220 assessment tasks that were completed in this study. Missing data were imputed using multiple regression.

For each of the 15 tasks, we liberally used a 2 (span) \times 3 (training condition) \times 3 (time) ANOVA to test for any signs of transfer effects among our training conditions. We chose this more liberal analysis to highlight the lack of evidence for transfer effects.

Table 1
Analysis of Variance Results From All Transfer Tasks

Effects	df	Critical <i>F</i> value ($p < .05$)	Running span images	Running span arrows	Reading span	Rotation span	Anti- saccade	Visual arrays	Primary memory pictures	Primary memory words	Secondary memory pictures	Secondary memory words	Keep track	Trail making	Matrices	Paper folding	Synthetic work
Main effects																	
Time	2,220	3.04	7.89	15.75	15.68	1.04	1.46	1.43	7.70	5.07	1.77	2.94	.38	1.77	.87	9.77	33.18
Condition	2,110	3.08	.47	1.08	3.32	10.83	.10	1.35	.17	.83	1.95	1.15	.07	.73	1.34	.08	.60
Span	1,110	3.93	105.95	93.5	113.36	150.24	41.37	78.06	24.10	38.68	53.21	36.46	72.16	43.16	80.38	71.40	93.83
2-way interactions																	
Time \times Condition	4,220	2.41	.22	1.14	1.82	7.43	.71	.39	.85	1.09	2.22	2.19	.35	.42	1.35	.22	.34
Time \times Span	2,220	3.04	2.01	1.06	.72	.57	1.01	1.96	.40	1.92	1.14	.69	1.13	1.01	.37	.93	.44
Condition \times Span	2,110	3.08	.68	.27	.71	.17	.07	.17	.03	.63	.32	.15	.94	.46	2.06	1.35	.30
3-way interaction																	
Time \times Condition \times Span	4,220	2.41	.34	1.18	.32	1.11	.92	.33	.64	1.03	.62	.24	.16	.49	.37	.69	.51

Note. Findings in bold are statistically significant with an alpha of .05.

Table 2

Means (SDs) of All Transfer Tasks by Test Session, Training Condition, and Span

Task	Complex span training						Running span training		
	High span			Low span			High span		
	Pre	Mid	Post	Pre	Mid	Post	Pre	Mid	Post
Running span images	37.50 (12.41)	40.37 (12.65)	44.02 (14.50)	20.25 (11.66)	21.75 (13.65)	21.20 (13.90)	42.45 (8.71)	42.95 (13.43)	45.25 (11.54)
Running span arrows	27.15 (8.95)	34.20 (11.55)	35.00 (16.21)	15.00 (8.80)	18.60 (10.24)	18.70 (8.83)	29.45 (10.21)	32.10 (11.19)	32.70 (12.38)
Reading span	48.70 (11.30)	52.85 (13.29)	45.00 (18.35)	28.55 (16.24)	30.60 (17.75)	23.75 (16.17)	48.60 (14.33)	46.64 (15.81)	43.10 (17.87)
Rotation span	38.95 (6.86)	48.35 (11.41)	50.15 (13.29)	20.34 (11.59)	23.80 (13.09)	25.95 (16.06)	38.80 (12.56)	35.80 (16.41)	34.65 (17.37)
Anti-saccade	.75 (.13)	.78 (.19)	.79 (.14)	.59 (.16)	.64 (.17)	.59 (.17)	.77 (.14)	.75 (.15)	.76 (.17)
Visual arrays	.74 (.11)	.73 (.10)	.70 (.11)	.57 (.07)	.59 (.08)	.57 (.08)	.69 (.11)	.70 (.12)	.69 (.13)
Primary memory pictures	3.34 (.79)	3.37 (.65)	3.36 (.65)	2.63 (1.07)	2.76 (.59)	2.95 (.69)	3.21 (.73)	3.55 (.67)	3.43 (.81)
Primary memory words	3.41 (.78)	3.51 (.81)	3.54 (.65)	2.80 (.77)	2.94 (.64)	2.86 (.84)	3.15 (.81)	3.60 (.63)	3.65 (.66)
Secondary memory pictures	2.32 (.80)	2.58 (.89)	2.68 (.91)	1.65 (.76)	1.52 (.74)	1.60 (.67)	2.38 (.86)	2.55 (.93)	2.38 (.99)
Secondary memory words	2.42 (1.36)	2.57 (1.19)	2.63 (1.15)	1.60 (.81)	1.74 (1.07)	1.87 (.71)	2.43 (1.02)	2.57 (.99)	2.22 (1.08)
Keep track	41.55 (8.85)	41.10 (10.53)	40.90 (8.42)	27.90 (12.13)	26.75 (13.91)	28.30 (12.14)	42.90 (7.01)	43.30 (7.93)	41.10 (9.12)
Trail making	31.35 (7.72)	32.10 (7.99)	29.80 (10.70)	21.05 (12.63)	18.39 (13.16)	20.70 (13.17)	30.05 (8.26)	31.47 (7.88)	28.70 (9.69)
Matrices	10.70 (2.45)	11.22 (2.51)	10.70 (2.54)	7.58 (3.58)	7.98 (3.38)	6.80 (3.64)	11.30 (2.38)	10.75 (2.29)	11.10 (2.20)
Paper folding	11.05 (3.39)	11.90 (3.73)	12.20 (3.53)	7.90 (4.43)	7.89 (4.71)	8.10 (4.42)	11.90 (2.81)	12.74 (1.85)	13.20 (1.67)
Synthetic work	385.40 (181.12)	431.80 (170.31)	519.60 (138.13)	132.15 (173.18)	205.46 (200.56)	240.56 (193.93)	346.45 (196.02)	448.05 (176.24)	522.85 (287.63)

Note. Means presented in bold are statistically different than baseline using *a*1-tailed paired samples *t*-test. means that are underlined remain statistically significant after a correction for multiple comparisons using *p* < .01 with a two-tailed *t*-test.

Table 1 outlines the results of the 15 separate, $2 \times 3 \times 3$ ANOVAs. Effects in bold are statistically significant. Although we will not go into detail for each of the assessment tasks, three important findings stand out from this table. First, and unsurprisingly, high spans outperformed low spans across every task: as represented by the main effects of span. Second, subjects did improve on several tasks across the three time-points (main effects of time), but some tasks showed no overall improvements with practice. This inconsistency of practice effects might be driven by the fact that these subjects had completed similar tasks in our lab in the past; some practice effects may have occurred prior to completing this study. A more likely explanation, however, is that some tasks were simply more prone to practice effects than others. Finally, the lack of any interaction effects—with the exception of the Time \times Condition interaction on the rotation span task—demonstrates very little support for the ability to improve executive functioning through working memory training.

To further elucidate these findings, Table 2 outlines the means and standard deviations of the assessment tasks by training condition, span level, and time for each of the 15 assessment tasks. To again highlight the lack of any far transfer effects, we've split these means by individual cells, and conducted one-tailed paired-samples *t* tests. Means that are highlighted in bold italics represent each mid and posttest score that is significantly higher than the baseline—pretest—scores. In addition, means that are underlined are statistically different from baseline using a less-liberal multiple-comparison adjustment of *p* values < .01 with a two-tailed *t* test.

The first four lines represent near transfer effects: That is, they represent measures identical in structure to the trained tasks, while utilizing unique stimuli. The majority of statistically significant improvements appear with these near assessment tasks: with nearly all effects specific to the running span assessment tasks. Although we do see increased scores over time for the running span tasks, the improvements in the complex span and running span training conditions are nearly mirrored in the active-control training condition as well: A finding that suggests these improvements are reflective of practice effects, rather than any improve-

ment in general ability. However, the improved scores on the rotation span task for high spans in the complex span training condition demonstrates the driving force behind the Time \times Condition interaction discussed earlier. More to the point, it suggests that some near transfer benefits may have occurred, but only for high spans, and only in the complex span training condition. However, it should be noted again these analyses utilized the most liberal of tests to highlight the lack of transfer effects in the data.

The remaining lines of Table 2 show a consistent lack of improvement across time on the mid and far assessment tasks. However, one of the Gf tasks (paper folding) does show some improvement across time: but the improvement across time is again mirrored in the active-control condition, which suggests these improvements are reflective of practice effects.

Looking only at the underlined items in Table 2—which represents a more conservative two-tailed test with a *p* < .01 threshold, most of these differences fail to reach statistical significance. The likely practice effects for the running span arrows task remains for both complex span training and visual search training conditions, along with an increase on the PM (words) task for high spans in the running span training condition. However, this later increase is not reflected in the ANOVA findings in Table 1 and is likely Type I error. In addition, the increased performance on the rotation span task for high spans in the complex span training condition also remains, and given the consistency of the increase at midtest and posttest, the fact that the task matches the training condition, and the reflected findings in the ANOVA for that task, this finding appears far more credible.⁵

In short, using even the most liberal of statistical tests, no evidence was found to support the notion that training on working memory tasks can transfer to improvement with other cognitive abilities. However, what little evidence we found for near transfer

⁵ A close examination of Table 2 will also demonstrate a reliable decrease in Reading Span scores over time across all conditions. Debriefing discussions with some subjects would suggest this was caused by fatigue as the reading span task was the last task of the assessment and was the longest task to complete.

Running span training			Visual Search Training					
Low span			High Span			Low Span		
Pre	Mid	Post	Pre	Mid	Post	Pre	Mid	Post
18.58 (14.11)	18.84 (11.06)	19.92 (11.80)	35.37 (11.81)	40.53 (13.00)	39.95 (16.75)	17.89 (11.30)	20.11 (11.79)	20.17 (11.73)
12.53 (8.06)	15.68 (9.36)	13.42 (7.15)	27.30 (8.34)	32.00 (8.98)	29.84 (11.77)	12.44 (7.94)	14.00 (8.88)	16.06 (6.92)
22.79 (14.84)	17.74 (11.47)	15.00 (12.43)	48.10 (14.37)	47.58 (16.04)	35.11 (19.19)	24.78 (15.64)	20.67 (11.56)	15.44 (9.24)
12.70 (9.32)	11.26 (8.23)	10.21 (9.82)	40.45 (8.08)	39.47 (12.81)	38.42 (14.25)	17.30 (10.28)	14.61 (6.90)	17.72 (8.98)
.59 (.15)	.61 (.14)	.58 (.16)	.75 (.15)	.78 (.12)	.77 (.11)	.61 (.13)	.62 (.14)	.63 (.15)
.54 (.07)	.55 (.09)	.56 (.07)	.71 (.10)	.73 (.10)	.70 (.12)	.58 (.09)	.59 (.08)	.59 (.08)
2.68 (.65)	2.99 (.81)	3.00 (.70)	3.25 (.65)	3.46 (.64)	3.46 (.53)	2.68 (.80)	3.03 (.69)	2.81 (.76)
2.54 (.80)	2.42 (.79)	2.82 (.90)	3.17 (.74)	3.35 (.71)	3.43 (.41)	2.73 (.70)	2.67 (.73)	2.78 (.59)
1.58 (.88)	1.42 (.81)	1.32 (.78)	2.31 (.89)	2.29 (.91)	1.83 (.82)	1.48 (.78)	1.48 (.89)	1.18 (.54)
1.33 (.68)	1.65 (1.02)	1.25 (.86)	2.54 (1.01)	2.38 (1.01)	2.09 (1.01)	1.42 (.62)	1.68 (.88)	1.24 (.68)
24.84 (11.40)	25.00 (10.96)	24.79 (10.61)	39.79 (8.53)	41.26 (8.41)	39.21 (9.80)	28.72 (9.16)	27.50 (9.95)	27.94 (9.74)
14.37 (11.27)	16.26 (12.87)	15.67 (13.08)	30.39 (8.21)	31.00 (9.14)	29.58 (11.36)	19.06 (11.30)	21.33 (9.07)	20.17 (8.23)
5.47 (2.55)	5.31 (3.46)	5.74 (2.77)	10.55 (2.62)	11.06 (2.39)	10.84 (2.95)	7.17 (3.32)	7.50 (3.25)	6.78 (3.64)
5.26 (3.74)	6.53 (3.65)	6.18 (4.16)	11.94 (2.62)	12.89 (1.66)	12.95 (2.84)	6.11 (4.06)	6.78 (4.68)	7.22 (3.96)
84.89 (118.76)	149.19 (158.15)	194.74 (185.44)	411.26 (161.88)	508.82 (188.21)	523.02 (196.64)	114.00 (130.70)	185.50 (200.80)	231.21 (202.97)

benefits seem to be driven by the improved scores of high spans: A finding that is consistent with the results of the scores on the training tasks.

General Discussion

Across all three training conditions, subjects showed improvements throughout the 20 training sessions. Subjects in all conditions were able to show remarkable gains in ability level across the tasks. But regardless of all conditions showing large improvements, the increasing gap between high spans and low spans in the two training conditions remains the predominantly visible effect—particularly when contrasted with the flat trend lines found in the active-control condition. It suggests that a person's ability to gain something from working memory training depends on the prior level of ability. More to the point, it suggests that the benefits of working memory training will incomparably benefit high spans and further increase the gap between high and low spans.

These findings do fit well with the research suggesting that some near transfer effects for similar tasks may occur, but that there is an overall lack of evidence for far transfer effects (Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016), and with research that suggests high spans would benefit more than low spans from a training regimen (Fuchs et al., 2014; Swanson, 2014).

Although we have answered the question of who benefits most from working memory training, one important question still remains: What is the mechanism behind near transfer? The question of transfer of learning has a long history in psychology. Thorndike and Woodworth (1901) showed transfer of learning on simple tasks (e.g., crossing out all the “e”s in a paragraph) but only when the assessment task used the same elements as the training task (e.g., crossing out “e”s made subjects better able at crossing out “e”s on a different paragraph but not all the “i”s). The finding of near transfer without far transfer is a consistent finding in the history of psychology across multiple domains. This pattern of results is found in discrimination learning with nonhuman primates (e.g., Harlow & Warren, 1952), strategy training for subjects with memory deficiencies (Belmont & Butterfield, 1971), and now with

working memory training (e.g., Redick et al., 2013). In addition, with this previous literature there is evidence to suggest that subjects with higher cognitive ability show more transfer than subjects with lower cognitive ability (Campione, & Brown, 1978; Cariglia-Bull & Pressley, 1990).

Our present findings are consistent with two hypotheses. First, it may be the case that near transfer effects are not a result of an improvement in WMC but rather that our subjects learned stimuli-specific strategies to increase their performance on the training tasks. In other words, high spans may be better able to generate and utilize new strategies during training. This hypothesis would explain both the disparate increases on the training tasks between high and low spans, and why no evidence could be found for far transfer effects: That is, the new strategies were task-specific and could only apply to similar, near transfer, assessment tasks. Second, near transfer effects may be a product of improvements in specific cognitive behaviors, namely, posterror adjustment (e.g., Rabbitt, 1966) and changes in speed/accuracy trade-off. For example, subjects could learn from training on the complex span tasks (i.e., operation span and symmetry span) that they need to spend more time performing the processing task on the assessment complex span tasks (e.g., making rotated letter judgments in the rotation span) after they have recently made errors. Subjects could also show near transfer effects because they identify that the memory component of the complex span tasks requires more cognitive effort than solving the processing the task and allot more time to remembering the stimuli.

We were able to address an additional question with this study concerning the measurement of WMC. Previous research has shown that the operation span task has good test–retest reliability (Klein & Fiss, 1999) but it is still unknown whether the complex span tasks would still be psychometrically valid measures of WMC after 20 training sessions. We correlated the WMC composite that we used to assign subjects to our high span or low span groups with subjects' performance on the complex span training tasks for Sessions 1 and 20. If the complex span training tasks were not valid measures of WMC after 20 training sessions we should

see that the correlation between our WMC composite and the complex span training tasks decreases from the first session to the 20th session. However, if the complex span training tasks still validly measured WMC, we should find similar correlations between Sessions 1 and 20. That is precisely what we found: the correlations for the operation span training task and our WMC composite went from $r = .56$ on Session 1 to $r = .57$ on Session 20 and the correlations for the symmetry span training task and the WMC composite went from $r = .70$ on Session 1 to $r = .49$ on Session 20 (these two correlations are not statistically different, $p > .05$). From these analyses, we can see that the complex span tasks are still psychometrically valid tasks even after 20 sessions of training.

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.
- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, 22, 366–378.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.
- Belmont, J. M., & Butterfield, E. C. (1971). Learning strategies as determinants of memory deficiencies. *Cognitive Psychology*, 2, 411–420. [http://dx.doi.org/10.1016/0010-0285\(71\)90024-7](http://dx.doi.org/10.1016/0010-0285(71)90024-7)
- Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, 42, 563–570. <http://dx.doi.org/10.3758/BRM.42.2.563>
- Campione, J. C., & Brown, A. L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence*, 2, 279–304. [http://dx.doi.org/10.1016/0160-2896\(78\)90020-X](http://dx.doi.org/10.1016/0160-2896(78)90020-X)
- Cariglia-Bull, T., & Pressley, M. (1990). Short-term memory differences between children predict imagery effects when sentences are read. *Journal of Experimental Child Psychology*, 49, 384–398. [http://dx.doi.org/10.1016/0022-0965\(90\)90066-H](http://dx.doi.org/10.1016/0022-0965(90)90066-H)
- Chein, 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. <http://dx.doi.org/10.3758/PBR.17.2.193>
- Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, 40, 531–542.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12, 769–786. <http://dx.doi.org/10.3758/BF03196772>
- Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, 23, 720–730. <http://dx.doi.org/10.1037/a0014296>
- Dalezman, J. J. (1976). Effects of output order on immediate, delayed, and final recall performance. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 597–608. <http://dx.doi.org/10.1037/0278-7393.2.5.597>
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450–466. [http://dx.doi.org/10.1016/S0022-5371\(80\)90312-6](http://dx.doi.org/10.1016/S0022-5371(80)90312-6)
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959–964. <http://dx.doi.org/10.1126/science.1204529>
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational testing service.
- Elsmore, T. F. (1994). SYNWORK: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments & Computers*, 26, 421–426. <http://dx.doi.org/10.3758/BF03204659>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19–23. <http://dx.doi.org/10.1111/1467-8721.00160>
- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *The Journal of Educational Research*, 84, 253–262. <http://dx.doi.org/10.1080/00220671.1991.10886025>
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, 44, 145–199. [http://dx.doi.org/10.1016/S0079-7421\(03\)44005-X](http://dx.doi.org/10.1016/S0079-7421(03)44005-X)
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128, 309–331. <http://dx.doi.org/10.1037/0096-3445.128.3.309>
- Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, 43, 226–236. <http://dx.doi.org/10.3758/s13421-014-0461-7>
- Fuchs, L. S., Geary, D. C., Fuchs, D., Compton, D. L., & Hamlett, C. L. (2014). Sources of individual differences in emerging competence with numeration understanding versus multidigit calculation skill. *Journal of Educational Psychology*, 106, 482–498. <http://dx.doi.org/10.1037/a0034444>
- Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, 24, 1149–1167.
- Harlow, H. F., & Warren, J. M. (1952). Formation and transfer of discrimination learning sets. *Journal of Comparative and Physiological Psychology*, 45, 482–489. <http://dx.doi.org/10.1037/h0057267>
- Harrison, T. L., Shipstead, Z., & Engle, R. W. (2014). Taxonomy of transfer to cognitive abilities: The case of working memory training. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger, III (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honour of Larry L. Jacoby* (pp. 307–321). New York, NY: Psychology Press.
- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, 24, 2409–2419. <http://dx.doi.org/10.1177/0956797613492984>
- Hicks, K. L., Foster, J. L., & Engle, R. W. (2016). Measuring Working Memory Capacity on the Web with the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition*, 5, 478–489.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6829–6833. <http://dx.doi.org/10.1073/pnas.0801268105>
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2012). Cogmed and working memory training—current challenges and the search for underlying mechanisms. *Journal of Applied Research in Memory & Cognition*, 1, 211–213. <http://dx.doi.org/10.1016/j.jarmac.2012.07.002>
- Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183. <http://dx.doi.org/10.1037/0096-3445.130.2.169>
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (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. <http://dx.doi.org/10.1037/0096-3445.133.2.189>
- Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of “executive attention”. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 749–777. <http://dx.doi.org/10.1037/0278-7393.32.4.749>
- Klein, K., & Fiss, W. H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments & Computers*, 31, 429–432. <http://dx.doi.org/10.3758/BF03200722>
- Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1076/jcen.24.6.781.8395>
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence*, 14, 389–433. [http://dx.doi.org/10.1016/S0160-2896\(05\)80012-1](http://dx.doi.org/10.1016/S0160-2896(05)80012-1)
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281. <http://dx.doi.org/10.1038/36846>
- Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49, 270–291. <http://dx.doi.org/10.1037/a0028228>
- Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer”: Evidence from a meta-analytic review. *Perspectives on Psychological Science*, 11, 512–534. <http://dx.doi.org/10.1177/1745691616635612>
- Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, 81, 111–121. <http://dx.doi.org/10.1111/j.2044-8295.1990.tb02349.x>
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18, 46–60. <http://dx.doi.org/10.3758/s13423-010-0034-0>
- Rabbitt, P. M. A. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, 71, 264–272. <http://dx.doi.org/10.1037/h0022853>
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales: Section 4. The advanced progressive matrices*. San Antonio, TX: Harcourt Assessment.
- 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. <http://dx.doi.org/10.1037/a0029082>
- Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26, 813–822.
- Ricker, J. H., & Axelrod, B. N. (1994). Analysis of an oral paradigm for the Trail Making Test. *Assessment*, 1, 47–51. <http://dx.doi.org/10.1177/1073191194001001007>
- Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, 132, 566–594. <http://dx.doi.org/10.1037/0096-3445.132.4.566>
- Shipstead, Z., Harrison, T. L., Trani, A. N., Redick, T. S., Sloan, P., Bunting, M. F., . . . Engle, R. W. (2017). *Individual Differences in cognitive control: Unitary working memory or diverse executive functions*. Manuscript submitted for publication.
- Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012a). Cogmed working memory training: Does the evidence support the claims? *Journal of Applied Research in Memory & Cognition*, 1, 185–193. <http://dx.doi.org/10.1016/j.jarmac.2012.06.003>
- Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012b). Working memory training remains a work in progress. *Journal of Applied Research in Memory & Cognition*, 1, 217–219. <http://dx.doi.org/10.1016/j.jarmac.2012.07.009>
- Shipstead, Z., Lindsey, D. R., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, 72, 116–141.
- Shipstead, Z., Redick, T., & Engle, R. (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. <http://dx.doi.org/10.1037/a0027473>
- Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research*, 7, 1–24. <http://dx.doi.org/10.2190/VQJD-TIYD-5WVB-RYPJ>
- Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, 105, 6791–6792.
- Swanson, H. L. (2014). Does cognitive strategy training on word problems compensate for working memory capacity in children with math difficulties? *Journal of Educational Psychology*, 106, 831–848. <http://dx.doi.org/10.1037/a0035838>
- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions: II. The estimation of magnitudes. *Psychological Review*, 8, 384–395. <http://dx.doi.org/10.1037/h0071280>
- Tulving, E., & Colotla, V. A. (1970). Free recall of trilingual lists. *Cognitive Psychology*, 1, 86–98. [http://dx.doi.org/10.1016/0010-0285\(70\)90006-X](http://dx.doi.org/10.1016/0010-0285(70)90006-X)
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127–154. [http://dx.doi.org/10.1016/0749-596X\(89\)90040-5](http://dx.doi.org/10.1016/0749-596X(89)90040-5)
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498–505. <http://dx.doi.org/10.3758/BF03192720>
- Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1302–1321. <http://dx.doi.org/10.1037/0278-7393.30.6.1302>
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, 438, 500–503. <http://dx.doi.org/10.1038/nature04171>
- von Bastian, C. C., & Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *Journal of Memory and Language*, 69, 36–58.
- Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Östenson, M. L., Bartfai, A., & Klingberg, T. (2007). Computerized working memory training after stroke—a pilot study. *Brain Injury*, 21, 21–29.
- Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 2, 65–74. [http://dx.doi.org/10.1016/S0022-5371\(63\)80069-9](http://dx.doi.org/10.1016/S0022-5371(63)80069-9)

Received November 3, 2015

Revision received March 9, 2017

Accepted April 8, 2017 ■