CrossMark

A simultaneous examination of two forms of working memory training: Evidence for near transfer only

Meredith Minear[1] · **Faith Brasher**[2] · **Claudia Brandt Guerrero**[2] · **Mandy Brasher**[2] · **Andrew Moore**[2] · **Joshua Sukeena**[2]

Abstract The efficacy of working-memory training is a topic of considerable debate, with some studies showing transfer to measures such as fluid intelligence while others have not. We report the results of a study designed to examine two forms of working-memory training, one using a spatial n-back and the other a verbal complex span. Thirty-one undergraduates completed 4 weeks of n-back training and 32 completed 4 weeks of verbal complex span training. We also included two active control groups. One group trained on a non-adaptive version of n-back and the other trained on a real-time strategy video game. All participants completed pre- and post-training measures of a large battery of transfer tasks used to create composite measures of short-term and working memory in both verbal and visuo-spatial domains as well as verbal reasoning and fluid intelligence. We only found clear evidence for near transfer from the spatial n-back training to new forms of n-back, and this was the case for both adaptive and non-adaptive n-back.

Keywords Working memory

Working memory is of great interest across psychology. Defined as the ability to simultaneously store and process information in the service of a goal (Baddeley, 1992; Miyake & Shah, 1999), working memory is frequently measured by performance on computerized tasks where one must remember a series of items while performing accurately on an interleaved secondary task such as reading a sentence or solving arithmetic problems (Daneman & Carpenter, 1980; Turner & Engle, 1989). What makes the construct so central is its ability to predict performance on a wide array of complex cognitive tasks such as reading and language comprehension (Daneman & Carpenter, 1980; Daneman & Merikle, 1996), reasoning (Kyllonen & Christal, 1990), and mathematical (LeFevre, DeStefano, Coleman, & Shanahan, 2005) performance as well as measures of fluid intelligence (Kane, Conway, Hambrick, & Engle, 2007). Additionally, poor working-memory performance is seen in disorders such as attention-deficit hyperactivity disorder (ADHD; Barkley, 1997), schizophrenia (Goldman-Rakic, 1994), and learning disabilities (Swanson, 2006). Therefore, a natural question is whether improvements in working-memory capacity will lead to improved performance on complex tasks.

The broader question of whether cognitive enhancement can be achieved through the training of basic cognitive tasks is an old one (Thorndike & Woodworth, 1901), but the 21st century has seen a growing number of studies suggesting that such training regimens can be effective, especially when working memory is the target. Klingberg, Forssberg, and Westerberg (2002) reported that the use of a computerized working-memory program in children with ADHD showed transferable improvements to working memory as well as a measure of fluid intelligence. One feature deemed critical for success was the adaptive nature of their program, i.e. the difficulty was constantly adjusted for each individual based on their current performance so they would always be challenged but not frustrated. Participants run on a non-adaptive version of the program did not show the same benefits (Klingberg, Forssberg, & Westerberg, 2002; Klingberg et al., 2005). More startling and more controversial was a report by Jaeggi, Buschkuehl, Jonides, and Perrig (2008) demonstrating the effects of working-memory training in college students.

✉ Meredith Minear
  meredithminear@gmail.com

1 Department of Psychology, University of Wyoming, Laramie, WY, USA

2 The College of Idaho, Caldwell, ID, USA

 Springer

Here working memory was trained using an adaptive dual n-back paradigm and transfer was seen to measures of fluid intelligence thus raising the possibility that working-memory capacity could be increased in an already high functioning sample and show generalizable benefits. Chein and Morrison (2010) also reported transferable improvements following a different form of working-memory training based on complex span tasks similar to those commonly used to measure working-memory capacity. They reported transfer to both working and short-term memory as well as to the Stroop task and a reading comprehension measure. However, they did not find transfer of training to fluid intelligence.

These data were met with both excitement (Sternberg, 2008) and considerable skepticism. A meta-analysis of working-memory training studies found little support for transfer to other cognitive abilities (Melby-Lervåg & Hulme, 2013), while Shipstead, Redick and Engle (2012) highlighted a number of methodological weaknesses with many of the initial studies including (1) inadequate measurement, i.e. using a single task to measure a construct such as fluid intelligence, (2) conflation of working memory with short-term memory, and (3) inadequate control groups. There have been a growing number of studies addressing these methodological concerns. However, the question of whether working-memory training is effective and the extent to which it shows transfer to measures of fluid intelligence and cognitive control is still unclear, with some studies reporting evidence of such transfer (Jaeggi, Buschkuehl, Shah, & Jonides, 2014; Jaušovec & Jaušovec, 2012; Rudebeck et al., 2012; Stephenson & Halpern, 2013;) while others have not (Chooi & Thompson, 2012; Redick et al., 2013; Schwarb, Nail, & Schumacher, 2015; Sprenger et al., 2013; Thompson et al., 2013). Therefore, important questions that still have no clear answer are what is being trained in working-memory training and what are the mechanisms through which transfer, especially far transfer, can be achieved?

Jaeggi and colleagues have proposed that working-memory training may be viewed as analogous to basic physical conditioning where working memory plays a role similar to the cardiovascular system (Jaeggi et al. 2011). Improving cardiovascular fitness benefits performance on a wide array of sports. Therefore, working-memory training may be viewed as improving the efficiency of key cognitive processes such as attentional control or inhibitory processing that are involved across multiple areas of cognition. Von Bastian and Oberauer (2014) have proposed two routes working-memory improvements could take. The first is that training may improve the capacity of working memory such that after training individuals can keep a larger amount of information active. If working-memory training increases capacity, then wide transfer across different tasks would be predicted (von Bastian & Oberauer, 2014). However, theoretically, capacity based changes should be domain specific, i.e. improvements in verbal capacity transfers to tasks using verbal working memory

and visuo-spatial improvements transfer to visuo-spatial tasks. The second route is training induces increased working-memory efficiency due to knowledge and skills learned during training. An example would be participants developing specific strategies such as chunking to improve on the training task. Transfer would then be more specific to tasks where participants could apply these strategies. However, a third possibility could simply be that some participants become more motivated or exert more effort on cognitive tasks after working-memory training especially if they hold a strong belief in the efficacy of the training. Such a motivational shift might manifest itself in improved performance across transfer tasks when transfer of training would not be predicted by mechanistic or strategic changes with training.

Another important question is: are the mechanisms responsible for transfer of training necessarily the same in different working-memory training paradigms? Working memory is involved in many different tasks, therefore it would be possible to develop different training regimes with one showing transfer due to changes in cognitive architectural parameters such as capacity while another leads to more skill-based improvements in working-memory efficacy such as active strategies developed by participants to maximize task performance (Kieras, Meyer, Mueller, & Seymour, 1999). To date, two tasks have been most frequently used for training working memory, the n-back (e.g., Jaeggi et al., 2008) and the complex span (e.g., Chein & Morrison, 2010). Briefly, in the n-back task, participants see or hear a series of stimuli and for each stimulus must judge whether it is the same stimulus seen n items back. In the dual n-back, participants perform a visual and auditory n-back simultaneously. In complex span tasks, participants must remember a series of stimuli (e.g., letters) with a processing task such as judging the veracity of a mathematical statement interleaved between each item. Both tasks are deemed working-memory tasks because they require the simultaneous processing and storage of information and both predict variance in fluid intelligence scores (Kane, Conway, Miura, & Colflesh, 2007). Both tasks would appear to require selective attention, attentional control (Engle, 2002), storage of items in primary memory and retrieval from secondary memory (Unsworth & Engle, 2007). However, a recent meta-analysis of studies measuring the relationship between the n-back and complex span found only a weak relationship between the two tasks and concluded they tap different components of working memory (Redick & Lindsey, 2013). Therefore, training based improvements may rely more on processes specific to the particular training task. For example, the n-back clearly involves a great deal of updating information while span tasks require shifting between two unrelated tasks. Both may require inhibiting no longer relevant items, but in the n-back this is required constantly within trials while for span tasks inhibition is needed between trials. Performance on a working-memory task may improve with training due to

any mixture of reasons such as (1) increased capacity of short-term store/primary memory, (2) improved executive function which may involve better attentional focus or the suppression of irrelevant information or both, or (3) other task-specific processes such as updating or shifting.

A third important consideration for researchers attempting to demonstrate the efficacy of any form of training is the choice of control group. Since a pre to post increase on a particular task could be due simply to practice effects, a comparison group is needed to demonstrate effects due to the training intervention. While some studies have employed a simple no-contact control group in which participants are simply tested on the transfer measures twice with no other interaction with experimenters, an active control group in which participants undergo some form of intervention is preferred. However, recently a number of concerns about active control groups have been raised including participant awareness that they are in the control group and the role of differential expectancies of both the participants in a particular intervention and the laboratory assistants testing them (Boot, Blakely, &Simons, 2011; Boot, Simons, Stothart, & Stutts, 2013). As Boot et al. (2011) have noted, placebo effects are a big concern for researchers studying cognitive training as participant expectancies and effort alone may improve performance.

Finally, a growing concern has been raised with the limitations of the standard analysis approach of null hypothesis significance testing (NHST) when evaluating the evidence of transfer of training (Dougherty, Hamovitz, & Tidwell, 2016; Sprenger et al., 2013). While a number of studies as noted above have reported no evidence for transfer effects from working-memory training, most have only reported their results using NHST, which does not allow an adequate assessment of the extent to which the data collected support the null hypothesis, i.e. that transfer was not present. However, the use of Bayesian analysis does allow for an estimate of the evidence in favor of the null hypothesis (Rouder, Morey, Speckman, & Province, 2012). Sprenger et al. (2013) included Bayesian analyses in their studies of working-memory training and reported evidence in favor of the null. Dougherty et al. (2016) conducted a meta-analysis of working-memory training studies using a Bayesian approach and found that studies that using a passive control group were more likely to favor the alternative hypothesis that training was effective, while studies employing active control groups were more likely to support the null. Therefore, in our study, we report our results using both traditional NHST and Bayesian analyses.

In the current study, we took a novel approach to investigating the possible sources of transferable improvement resulting from training working memory by both comparing the effects of two different forms of working-memory training, n-back and complex span, on the same set of transfer measures as well as utilizing two different active control groups. Both n-back training and complex span training have been proposed as having domain general transfer effects upon working memory although typically the training protocols include both verbal and spatial variants of the training task. In the case of the n-back, single n-back training in which the only spatial or only verbal stimuli are used has been reported as being as effective as the dual n-back (combining verbal and spatial) on measures of spatial reasoning (Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010). However, most studies of complex span training have incorporated both spatial and verbal variants into the training protocol making it difficult to ascertain whether transfer depends on separate improvements in verbal and spatial SHORT-TERM MEMORY capacity or the possible executive components involved in performing a complex span task. Therefore, we chose to isolate each form of working-memory training to one domain in order to better compare whether any possible transfer effects were isolated to the domain trained or generalized across verbal and visuo-spatial domains.
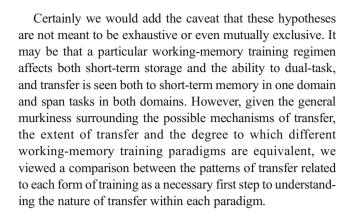
In addition to comparing two different forms of working-memory training, we included two active control groups. The first has been used in prior studies of working-memory training in which participants practice a non-adaptive version of the n-back task (Klingberg et al. 2002; 2005). We also had this group complete fewer training sessions (12 compared to 20) so that this group both lacked the adaptive component and the number of training sessions initially reported as important for transfer to occur (Jaeggi et al., 2008). However, as noted above, this form of control group can been criticized for failing to control expectancy effects as the less challenging and repetitive nature of the non-adaptive n-back task might be less motivating for participants as well as leading some participants to be become suspicious they were in a control condition. Therefore, we also included a second active control group in which participants learned and practiced a cognitively demanding real-time strategy (RTS) video game, Starcraft. Real-time strategy games would appear to place a demand on working memory as players must maintain information over time in the service of complex goals. However, previous research has found no effects in younger college aged adults of playing RTS games on working memory and fluid intelligence measures such as the 2-back task, Operation Span, and Raven's Matrices for either expert video-gamers compared to non-gamers or novices who were trained on RTS games (Boot et al. 2008). Additionally, a recent meta-analysis of the effects of playing video games on cognition reported little to no improvement of executive function including measures of intelligence and working memory as a result of playing video games (Powers, Brooks, Aldrich, Palladino, & Alfieri, 2013). Therefore, in this group we expected the participants to be more engaged and have greater expectations for transfer, but past research has not found evidence of transfer to many of the tasks used in our pre-test post-test battery.

We designed transfer to range from very near, i.e., the same task used in training with different content to far transfer such as reasoning. This was done with the goal of trying to isolate possible improvements in short-term memory, executive function, and working-memory capacity and their contributions to any observed transfer to measures of complex cognition such as reasoning, fluid intelligence, and mathematical performance. We collected multiple measures of each construct so that evidence of transfer would not rely upon a single task and when possible tried to choose tasks that had shown prior evidence of being sensitive to some form of training intervention. We included verbal and visuo-spatial variants of our short-term memory, working memory, and reasoning transfer measures. This allowed us to test the claim that the effects of working-memory training are domain general. In addition to reporting our results using the standard NHST, we also computed the Bayes Factor assessing the extent to which the evidence favored the alternative or the null hypothesis. Finally, we measured individual differences in personality as well as self-reported motivation, and beliefs in the efficacy of the training as these have been proposed as one source of variance in why some studies see transfer effects while others do not (Jaeggi et al., 2014). We tested whether these variables predicted either the amount of improvement on the training task or the degree of transfer observed.

Based on the hypothesized possible mechanisms of working-memory training described above, we developed the following set of predictions.

1) If a particular form of working-memory training has its effects primarily on short-term capacity, then transfer should be domain specific and include at least the measures of short-term memory and working memory within the domain (verbal or spatial) trained.
2) If a particular form of working-memory training improves processes specific to the structure of the training task, then we would predict task based transfer (e.g., specific to span tasks or to n-back tasks) across verbal and non-verbal domains.
3) If a form of training generally benefits executive processes such as attentional capacity or inhibitory function, then we predict wider transfer across working-memory tasks in both domains as well as improvements on our executive function transfer tasks.
4) If improvements over training are accomplished primarily through the development of specific strategies, then we predict a narrow pattern of transfer related to that strategy such as specific improvements on tasks using stimuli specific to a strategy (such as using chunking to remember letters or creating a shape to remember a series of locations).

Certainly we would add the caveat that these hypotheses are not meant to be exhaustive or even mutually exclusive. It may be that a particular working-memory training regimen affects both short-term storage and the ability to dual-task, and transfer is seen both to short-term memory in one domain and span tasks in both domains. However, given the general murkiness surrounding the possible mechanisms of transfer, the extent of transfer and the degree to which different working-memory training paradigms are equivalent, we viewed a comparison between the patterns of transfer related to each form of training as a necessary first step to understanding the nature of transfer within each paradigm.

Method

Participants

A total of 132 College of Idaho students were recruited by campus flyers advertising a "brain training" study. They were paid US$10 an hour with a US$50 completion bonus with an average total compensation of US$300. After agreeing to participate in the study, participants were initially randomly assigned to one of the four conditions. Participants who were assigned to the video game group were asked during pre-testing how often a week they played video games and if they had played real-time strategy (RTS) games before. If an individual indicated that he or she had played RTS games on a weekly basis, that participant was reassigned to one of the three other groups. This was done simply to ensure that we did not have participants who were already skilled players in our RTS group. Only four participants were reassigned based on these criteria (one to the spatial n-back, one to verbal complex span, and two to the non-adaptive n-back). In our final sample, 34 participants were assigned to the spatial n-back training with three participants dropping out. Thirty-five participants were assigned to the verbal complex span training with three dropouts, 33 were assigned to the video game training with seven dropouts, and 30 were assigned to the non-adaptive n-back group with three dropouts. We based our planned sample sizes on the medium effect size (Cohen's d = .65) reported by Jaeggi et al. (2008) using a similar sample size. The demographics for the final sample are shown in Table 1.

Materials

Training

For each of our working-memory training regimens, we preserved each type of training in its original form both in terms of the length and the number of sessions rather than trying to alter one to fit the other. This resulted in a difference of

Table 1 Demographic information

Group	N	Gender		Age	
		Male	Female	M	SD
Spatial N-back	31	8	23	19.9	1.6
Verbal Complex Span	32	12	20	19.7	1.4
Video Game	26	10	16	19.8	1.5
Non-adaptive N-back	27	12	15	19.6	1.3

approximately 10 min per session with the Verbal Span group having slightly longer sessions. Similarly for the real-time strategy game group we used a training procedure similar to that reported by Boot et al. (2008). In this we erred on the side of faithfully replicating the procedures that had or in the latter case had not resulted in transfer rather than altering these procedures to exactly match each other. For the non-adaptive n-back group, we chose to have participants complete fewer sessions than the working-memory training groups so that this group not only lacked an adaptive element of training, but also had less exposure to the task, i.e., a lower dosage of training.

Adaptive spatial n-back To train visuo-spatial working memory, we used the spatial n-back procedure described by Jaeggi et al. (2010). In this task, participants saw a series of blue squares appear in eight different possible locations based on a 3 × 3 grid excluding the center square. For each square, they were to determine whether the current location was the same as the location n-back by pressing the "y" key. If the location did not match the location n-back, no response was made. Each stimulus appeared for 500 ms with an inter-stimulus interval of 2,500 ms. If participants made less than three errors during a block, the n-back level was increased by one. If participants made more than five errors during a block, the level was decreased by one. Each training session consisted of 15 blocks with each block consisting of 20+n trials and lasted approximately 20 min. Participants were scheduled for five sessions each week for 4 weeks for a total of 20 sessions.

Verbal complex span To train verbal working memory, we used the verbal complex span task developed by Chein and Morrison (2010). In this task, participants were asked to re-member a series of letters presented one at a time for 1 s. However, between each letter, participants completed 4 s of a lexical decision task, therefore engaging in simultaneous storage of the previous letters and processing of un-related material. At recall, participants were shown a 4 × 4 array of letters and reported the sequence by clicking on the letters in the same order they were presented. There was feedback on both recall accuracy as well as performance on the lexical decision task (% correct). Each training session began with a 4-item list. If a participant remembered all the letters in the

correct sequence while maintaining a level of 75 % correct on the lexical decision task for two consecutive trials, then the number of items in the sequence was increased by one. If participants performed two incorrect trials in a row, the num-ber was decreased by one. Similar to Chen and Morrison (2010), participants performed 30 trials of the task each train-ing session which lasted between 30 and 40 min. Participants were scheduled for five sessions each week for 4 weeks for a total of 20 sessions.

Real-time strategy game training Participants in this group were trained on Blizzard's 1996 RTS game "Starcraft: Brood War". We adopted a training protocol similar to those used by Boot et al. (2008) and Basak et al. (2008). Training occurred across 4 weeks and consisted of tutorial missions and matches against a computer opponent. The participants trained for 1.5 h/session, three times a week. Each session provided time for 3–6 games. During the first week participants played as the "Terran" race, as the "Protoss" in the second, "Zerg" in the third, with the final week being dedicated to randomized matches of all three. Aside from the tutorial missions, partic-ipants played a different map each week. During the fourth week, players were randomly assigned to a different map for each game.

Non-adaptive spatial n-back This group was trained on the same n-back task as the adaptive group with all the same task parameters with the subtraction of the adaptive component that increased or decreased the n-back level based on partici-pant performance. Instead this group alternated between seven blocks of 2-back training and seven blocks of 3-back training with a single block of 4-back training at the end of each session. These participants completed 12 approximately 20-min sessions of 15 blocks each over a 4-week period.

Transfer tasks

N-back We used two variants of the n-back task to test for specific transfer in n-back performance.

> Verbal n-back: This task was a verbal analogue of the n-back training task. Instead of locations, participants were given a continuous stream of block letters presented in the center of the screen. Eight letters were used: C, G, Q, T, H, K, P, and W. Participants were instructed to press the "y" key if the letter was the same as the letter n-back, otherwise they were not to respond. There was an initial practice block of a 2-back task with 12 trials with three hits. This was followed by three blocks of 2-back judg-ments, three blocks of 3-back judgments, and three blocks of 4-back judgments. Each block consisted of 20 stimulus presentations in addition to the first 2, 3, or 4 stimuli needed to establish a pattern. In each block, there

were six target trials (hits) and 14 non-target trials. The dependent measure was the total number of hits minus false alarms.

Object n-back: This task was identical in construction to the verbal n-back except that instead of letters, the stimuli were eight random polygons.

Speed of processing We used simple choice reaction time tasks to measure speed of processing using verbal and visuo-spatial stimuli. In the verbal domain, the Animal and Word tasks were used. In the visuo-spatial, the Dot Judgment and Array Matching tasks were used (Chen et al. 2007). All consisted of 20 trials and used the "z" key for left responses and the "/" key for right side responses with an equal number of correct "left" and "right" responses. All tasks had a short practice block of five trials. Both accuracy and reaction time were measured for each of these tasks.

Animal task: Participants were shown two words on the left and right of the screen and asked to decide as quickly and accurately as possible which of the words was an animal.

Word task: Participants were shown two strings of letters and asked to decide as quickly and accurately as possible which letter string was an English word.

Dot judgment Participants saw three dots on the screen, a central white dot and a blue and a red dot and were asked to decide as quickly and accurately as possible which of the colored dots was the closest to the white dot.

Array matching Participant saw three arrays of shapes, one array at the top of the screen and two choice arrays on the bottom left and right. Participants were asked to decide which of the choice arrays would be the best match to the top array based on color, number, or shape. For example, if the sample array was a set of four blue circles and the choice arrays were a set of two red squares and a set of three blue triangles, the blue triangles would constitute the best match.

Short-term memory To measure short-term storage, we used two tasks to measure verbal storage, letter span, and digit span, and two tasks to measure visuo-spatial storage, arrow span, and circle span. All four tasks consisted of 20 trials with lengths of three items in a series up to 12 with two trials at each length and a short block of four practice trials with 2-, 3-, and 4-item trials. The score for each task was the number of trials correctly answered.

Digit span: Participants were shown a series of digits presented one at a time for 1,500 ms with an ISI of 500 ms and were asked to remember the sequence of digits in their presented order. At test, participants saw a number pad with the digits 1–9 as well as a series of boxes presented in a horizontal line at the bottom of the screen. The box that was colored green indicated the position of the digit in the number series that was to be recalled. Participants then used the mouse to click on the number that was presented in that position.

Letter span: This task was identical to the Digit Span except that instead of digits, the letters B, F, H, J, L, N, P, R, and X were used.

Arrow span: In this task, participants were asked to remember a series of locations pointed to by a sequence of arrows that appeared one at a time in a 5 × 4 grid for 1, 500 ms followed by a 500-ms ISI. At test, participants saw a set of black checks in the grid marking all but one of the locations to which the blue arrows pointed to and were asked to click on the missing location.

Circle span: In this task, participants saw a set of 25 circles randomly distributed on the screen. During each trial, a series of circles turned red one at time for 1,500 ms with a 500-ms ISI. At test, participants saw the same cloud with all the circles that turned red filled in with black except one and they were asked to click on the missing circle with the mouse.

Working memory We used three measures of verbal working-memory span and three measures of visuo-spatial working-memory span.

Reading span: In this task participants were asked to remember a sequence of two to sevwn letters. Between each letter presentation they were asked to read a sentence and judge whether the sentence made sense. At test, participants were shown a grid of letters and had to click on the letters that were presented in the order that they were presented. Participants completed 15 trials. We used the absolute score calculated as the sum of all correctly recalled sets as described in Unsworth, Heitz, Schrock, and Engle (2005).

Operation span: In this task, participants were again asked to remember a series of two to seven letters. In between letter presentations, participants were asked to solve a simple mathematical problem and then asked to click on the letters seen in the order presented at the end of each trial. Participants completed 15 trials and the absolute score was used (Unsworth et al., 2005).

Letter-number span: In this task, participants were asked to remember a series of alternating letters and numbers with the instruction that they would then be asked to recall the sequence reorganized with the digits first in numerical order followed by the letters in alphabetical

order. Participants' memory for the sequence was tested by asking them to identify one member of the reordered sequence. Participants were cued to report the nth item in the correctly ordered sequence and then to click on the identity of that item using a visual keypad presented on the screen consisting of either the digits 1–9 or the letters B, F, H, J, L, N, P, R, or X. There were four practice trials consisting of sequence lengths of from two to four items. The experimental block consisted of 20 trials with two trials of each length between three and 12 items with the trial lengths presented in random order. The dependent measure was the total number of items correctly recalled. Symmetry span: In this task, participants were asked to remember a sequence of two to five locations presented one at a time. Between each location presentation, participants were shown a shape and asked to decide whether it was symmetrical or not. At the end of each trial, participants were asked to indicate the locations seen in order using a 16-location array. Participants completed 12 trials. The score was thus the total number of correct items in the correct position (Unsworth et al., 2005). Rotation span: Participants were asked to memorize a series of locations indicated by black Xs appearing one at a time for 1,500 ms on a 5 × 4 grid. Between each X, participants were shown two figures (letters chosen from the Thai alphabet), one upright and the other rotated by 60° or 120°. The participant had to judge whether the rotated figure was identical to or a mirror image of the upright figure while remembering the previous presented locations. At test, participants were asked to click on all the locations in which a black X appeared in the order that they were presented. Participants completed three blocks of four trials consisting of 2, 3, 4, and 5 items for a total of 12 trials. The dependent measure was the sum of the number of locations correctly recalled in its proper sequence. Alignment span: Participants were shown a series of red dots appearing on a 4 × 5 grid. Each red dot was presented with two white dots for 1,500 ms. Participants judged whether the accompanying white dots formed a straight line with the red dot while simultaneously remembering the locations of each of the previous red dots (Hale et al., 2011). At test, a grid containing all the red dots in the current sequence but one appeared and the participants clicked on the location of the missing dot. Participants completed a practice block of four trials and a block of 20 trials consisting of two trials of three to 12 items each. The dependent measure was the sum of trials correctly completed.

Executive function We used two measures of executive processing.

Attention network task: In this task, developed by Fan et al. (2002), participants performed a standard flanker task in which they indicated whether a central arrow is pointing to the right or left. The flanking three arrows on each side were either compatible or incompatible with the central arrow. In addition, the arrows appeared on either the top or bottom of the screen. Participants were either uncued as to when a trial will start, received a centered cue that alerts them to the beginning of the next trial, or an informative spatial cue appearing on either on the top or bottom of the screen, allowing them to orient to the location of the next stimulus. This resulted in three measures of attention, Alerting Attention (AA) the difference between average response time (RT) on trials where no cue is present compared to a center cue, Orienting Attention (OA) the difference between trials cued with a central cue and a spatial cue, and Executive Attention (EA) the difference between RT on trials with incongruent flankers compared to compatible flankers.

Simon task (Simon, Acosta, Mewaldt, & Speidel, 1976): In this task, participants saw an arrow pointing to the left or right side of the screen. Their job was to indicate which direction the arrow pointed and to press "l" key if the arrow pointed right and "a" if it pointed left. Participants first completed a block of 36 neutral trials where the arrow was presented in the center of the screen. The remaining two blocks consisted of 120 mixed spatially compatible and incompatible trials where the arrow appears on the left or right side of the screen. For compatible trials, the arrow appears on the same side of the screen that the arrow is pointing to, e.g., a right-pointing arrow appears on the right of the screen while for incompatible trials, the arrow appears on the opposite side of the screen compared to the direction the arrow is pointing. One of the mixed blocks consisted of 25 % incompatible trials and the other of 75 % incompatible trials. The order of these two blocks was randomized across participants.

Reasoning To measure possible transfer to reasoning abilities, we used two verbal asoning tasks from the ETS Kit of Factor-Referenced Tests (Ekstrom, French, & Harman, 1976), the Nonsense Syllogisms and Inference Tests, as well as two reasoning measures in the visuo-spatial domain, the Ravens Progressive Matrices (RPM; Raven, Raven, & Court, 1998) and the Cattell Culture Fair Tests (Cattell, 1961). For the ETS and Cattell measures, there were two forms already existing that were then counterbalanced across pre- and post-test administrations. For the RPM, the items were split into odd and even sets and also counterbalanced across testing times.

Inference test: Participants were given one or two statements similar to those seen in newspaper articles and then

a set of possible inferences based on those statements. Participants were instructed to choose the only inference that can be drawn from the given statements without assuming any additional information. There were ten items for which 6 min were given. The dependent measure was the number of correct items subtracting for incorrect items.

Nonsense syllogisms test: In this test, participants were given 15 sets of syllogisms and instructed to evaluate whether the conclusion followed logically from the premises. The dependent measure was the number of correct minus incorrect items completed within 4 min.

Raven's: We used a computerized version of the Raven's Standard Progressive Matrices test separated into two forms of 30 items. Participants saw a 3 × 3 matrix of shapes presented with the last shape missing and chose the item out of a set of six to eight choices that best completes the pattern both across the rows and columns. Participants were given 15 min to complete the task. The dependent measure was the total number of items answered correctly.

Cattell: This is a timed paper and pencil measure of non-verbal intelligence consisting of four tests, Series (13 items), Classifications (14 items), Matrices (13 items), and Conditions (ten items). There are parallel forms, A and B, which were counterbalanced at pre- and post-test across participants. The dependent measure was the total number of items correct on each test and then summed across tests.

Mathematical problem-solving measures To test the possibility of far transfer to tasks measured in real-world settings, we measured pre- and post-test performance of basic mathematical and arithmetic aptitudes also drawn from the ETS kit.

Mathematical aptitude test: This test consisted of 15 basic algebraic word problems in which participants had to solve for a single variable and identify the correct answer out of five choices. Participants were given 10 min. The dependent measure was the correct number of responses minus the number of incorrect responses.

Arithmetic aptitude test: In this test, participants were asked to solve 15 simple arithmetic word problems by choosing the correct alternative out of five possible responses. Ten minutes were given and the dependent measure was the number of correct subtracting incorrect items.

Survey measures In addition to surveying participants on their levels of motivation, strategies during training and beliefs about the efficacy of the training, we also collected a set

of individual difference measures that we felt may help us distinguish between more and less motivated participants in our samples.

Need for Cognition Questionnaire (Caccioppo & Petty, 1982): This 18-item questionnaire measures how motivated individuals are to engage in challenging activities.

The Short Grit Scale (Duckworth & Quinn, 2009): This eight-item questionnaire measures trait-based perseverance and passion for long-term goals.

The Brief Self-Control Scale (Tangey, Baumeister & Boone, 2004): This is a 13-item measure of self-reported self-control with items such as: "I am self-indulgent at times," and "I never allow myself to lose control."

The Big Five Inventory (BFI) (John, Donahue, & Kentle, 1991): This is a short 44-item inventory that measures the personality traits of Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness.

The Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983): This is a ten-item self-report questionnaire that measures a person's evaluation of the stressfulness of the situations in the past month of his or her life.

The Life Orientations Test (Revised) (Scheier, Carver & Bridges, 1994): This ten-item survey measures traits based optimism.

The Adult Hope Scale (Synder et al., 1991): This consists of 12 items designed to measure how hopeful an individual typically is.

The Barratt Impulsiveness Scale, Version 11 (Patton et al., 1995): The BIS-11 is one of the most frequently used self-report instruments for impulsivity and consists of 30 questions. It is designed to give an overall measure of general impulsiveness as well as three subscales of attentional, motor, and non-planning impulsiveness.

Participant motivation and strategies: After post-testing, participants completed a laboratory-based measure asking them a series of open-ended questions on the nature of their training experience focusing on: (1) what strategies did they use to improve their performance on the training task, (2) how motivated they were during training and how challenging did they find the tasks, and (3) whether they felt that the training had improved their cognition and did they notice any specific improvements on any of the post-test tasks. These questions are reproduced in the Appendix.

Procedure

All participants completed a week of pre-testing with approximately 60–90 min of testing each day. Pre-testing included an

initial measurement of all the transfer tasks as well as the personality measures. The following week participants began their training regimens. The week after training, participants had a week of post-testing in which we took a second measure of all the transfer tasks. Participants also completed our motivation and strategy questionnaire on the last day of post-testing.

Results

Training results

All four groups showed improved performance with practice. The dependent measure for the spatial n-back training was the highest n-back level achieved in each session. Figure 1a shows the average amount of improvement across the 20 sessions where the mean n-back level increased from 3.6 to 4.5, $F(19, 570) = 2.8$, $p < .001$, $\eta_p^2 = .09$. This is a smaller training improvement than reported by Jaeggi et al. (2010) where participants were achieving an average n-back level of close to 7-back by the end of a similar training regime. A closer examination of our data revealed considerable variability in the maximum n-back level achieved during training and how well that level was maintained across the 20 sessions. Of our 31 participants, three peaked at a maximum level of 4-back, six at 5-back, 12 at 6-back, seven at 7-back, one at 8-back, and two at 9-back. The average maximum level reached during training at any point was 6.1 with an average gain of 2.5 levels (from lowest n-back level to highest). However, we did see a great deal of fluctuation for most participants in n-back level achieved from session to session.

We performed a similar analysis for the verbal span group using the highest level (i.e., the longest string of letters correctly reported) achieved in a session as the dependent variable. We found the average number of letters retrieved went from 6.6 to 9.9 across 20 sessions of practice, $F(19, 589) = 13.8$, $p < .001$, $\eta_p^2 = .31$. These training data are shown in Fig. 1b.

The dependent measure for the video game group was developed using the participants' game scores. To keep track of each participant's progress, we recorded the scores following each game with the total score consisting of the player's performance with their units, structures, and resources. Performance within the units' sub-score was determined by how many units were created, how many were lost in battle, and how many enemy units the player's army killed. Performance within the structures' sub-score was calculated similarly to the units category (built, raised, lost), and the resources' sub-score was determined by totaling the amount of resources collected (the amount of resources a player spends is also tracked). This score ranged from 700 to 123,

580 points at Session 1 and from 34,896 to 305,840 points by Session 12. These data are shown in Fig. 1c. Again we saw significant improvements in game performance with training as well as considerable variability, $F(1, 25) = 40.4$, $p < .001$, $\eta_p^2 = .62$. However, only one individual showed no or negative improvement with a high score of 123,580 on the first day with notably worse performance with an average score of 46,000 across the remaining days of training. Although we initially screened based on self-reported RTS gaming experience, it is possible this individual misrepresented her previous experience or just scored exceptionally well the first day and then showed no improvement across the remaining sessions. Therefore, we dropped this participant from further analysis. All other participants demonstrated positive improvements in game score across training sessions.

Finally we also compared the accuracy data for the first versus the last block of the non-adaptive n-back control training to ensure that participants were performing well on the task throughout training. We found a significant improvement in accuracy from the first session ($M = .85$) to the last ($M = .92$), $F(1, 25) = 12.8$, $p < .001$, $\eta_p^2 = .33$.

Personality measures and training improvements
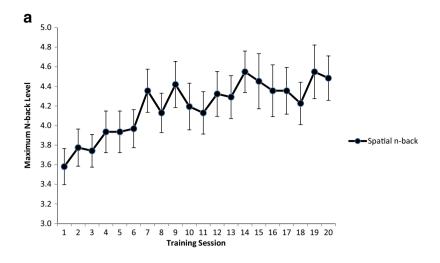
First, we tested whether there were any significant differences between our four groups on any of the personality measures. The means and standard deviations for each group are shown in Table 2. The only difference between any of the groups that approached significance was on the Adult Hope Scale, $F(3, 116) = 2.59$, $p = .056$, $\eta_p^2 = .07$.
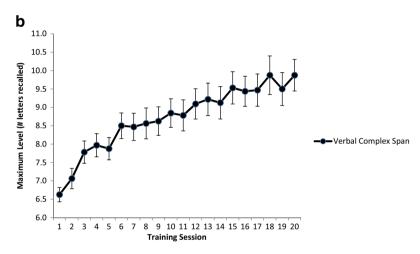
In order to test whether any of the personality measures would help predict working-memory training improvements, we took the difference in performance on the first session of training compared to the last and converted this into a z-score so that spatial n-back and verbal complex span improvements were on the same scale. We then tested for any correlations between personality and improvement and found that only self-reported stress, $r(63) = -.26$, $p = .04$ and neuroticism, $r(63) = -.25$, $p = .05$, showed any relationship with training improvements. The scatterplots of these relationships are shown in Fig. 2.

Transfer data

The descriptive statistics for all tasks broken down by group and testing time are presented in Table 3. To examine possible transfer, we were interested in creating composite gain scores similar to those used in previous studies (Redick et al., 2013) based on: (1) General N-back performance, (2) Speed of Processing, (3) Verbal Short-term Memory, (4) Visuo-spatial Short-term Memory, (5) Verbal Working Memory, (6) Visuo-spatial Working Memory, (7) Executive Function, (8) Verbal Reasoning, (9) Visuo-spatial Reasoning, and (7) Math
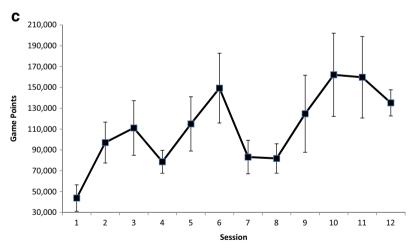
Fig. 1 (**a**) The training data for the spatial n-back group. (**b**) The training data for the verbal complex span. (**c**) The game- based improvement across sessions for the video game- training group based on game points per session

Performance as described above in the Methods section. This entailed first computing a standardized gain score for each measure by dividing the pre- to post-test gain for each participant by the standard deviation of the pre-test score for all participants, then averaging the gain scores for each task entered into the composite.

To ensure that the measures chosen for each construct were in fact related in our dataset, we computed the

Table 2 Survey responses by group
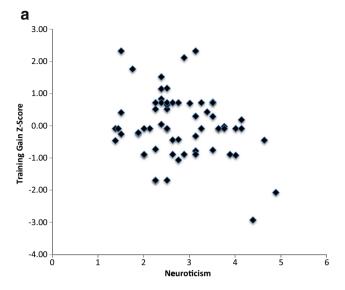
Survey Measures	Spatial N-back	Verbal Complex Span	Video Game	Non-adaptive N-back
Need for Cognition	64.40 (8.58)	64.13 (9.72)	64.40 (9.23)	61.96 (9.88)
Grit	3.51 (.51)	3.36 (.57)	3.43 (.59)	3.36 (.55)
Self-Control	41.48 (7.51)	39.91 (8.10)	41.47 (8.85)	40.85 (7.99)
Extraversion	3.38 (.83)	3.27 (.72)	3.27 (.84)	2.97 (.82)
Agreeable-ness	3.86 (.66)	3.80 (.65)	3.71 (.66)	3.65 (.70)
Neuroticism	2.71 (.76)	2.91 (.88)	2.82 (.63)	2.66 (.74)
Conscien-tiousness	3.66 (.60)	3.68 (1.16)	3.52 (.75)	3.58 (.62)
Openness	3.84 (.51)	3.58 (.58)	3.78 (.57)	3.62 (.55)
Perceived Stress	15.58 (5.76)	17.06 (6.64)	17.96 (5.71)	15.33 (6.59)
Optimism	17.10 (3.85)	15.03 (4.49)	16.31 (4.45)	15.81 (3.95)
Hope	49.93 (4.16)	46.59 (10.19)	51.14 (5.17)	50.37 (5.71)
Empathy	57.82 (9.39)	57.47 (13.91)	57.75 (6.94)	54.08 (7.20)
Impulsive-ness	33.10 (12.95)	35.05 (10.38)	34.73 (10.73)	33.59 (12.27)

correlations between tasks at pre-test across all four groups. This correlation matrix is shown in Table 4. All tasks were significantly correlated with at least one other measure in their proposed construct with two exceptions. The first was for the Visuo-spatial Short-Term Memory construct in which the Arrow Span and Circle Span scores were not significantly correlated at pre-test. This was likely due to the poor test-retest reliability shown by the Arrow Span task ($r = -.12$, $p = .26$ across groups) as compared to the Circle Span ($r = .45$, $p < .001$). One possibility is that some participants initially misunderstood that they were to indicate the squares the arrows were pointing to as opposed to the locations where the arrows appeared. However, as the program only recorded the accuracy of the response and not the actual location, we could not test this possibility directly. We can only infer this or some other initial difficulty with the task from the greater variability and larger range of the Arrow Span task at pre-test. Therefore, although Arrow and Circle Span were correlated at post-test, $r = .59$, $p < .001$, we decided against combining the two measures into a composite and report the results for each task separately. The second was the lack of any correlation between the AA, OA, or EA sub-scores on the ANT and the Simon task at either pre- or post-test.

Therefore, we did not combine these measures and examined the effect of group separately for each score.

The remaining final composites were then analyzed using ANOVA with group as a between-subjects factor. We also computed the Bayes factor (BF_{01}) comparing the null model to the alternative model using the Bayesian ANOVA provided in the JASP statistical package (JASP Team, 2016). BF_{01} quantifies the strength of evidence favoring one model over the other as represented by the relative probability of the observed data under the two models with a $BF_{01} > 1$ as evidence favoring the null, a $BF_{01} < 1$ favoring the alternative, and a $BF_{01} = 1$ as equivocal evidence with neither model more probable than the other. As noted by Sprenger et al. (2013), a $BF_{01} < .05$ (odds of 1/20) indicates positive support for the alternative hypothesis while a $BF_{01} > 3$ is taken as positive support for the null with values >10 indicating strong support for the null (Kass & Raftery, 1995). The results of the analyses on the composite variables are presented in Table 5 while the task level results for which we also included the Bayes factor comparing the null to the interaction model (Group × Session) are given in Table 6. Only two composites showed a significant effect of group. For the remaining composites, the Bayes factors were all >3 indicating
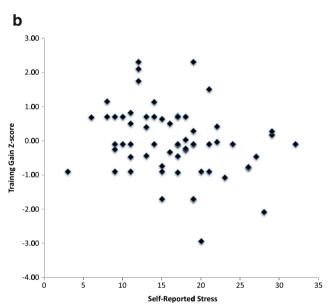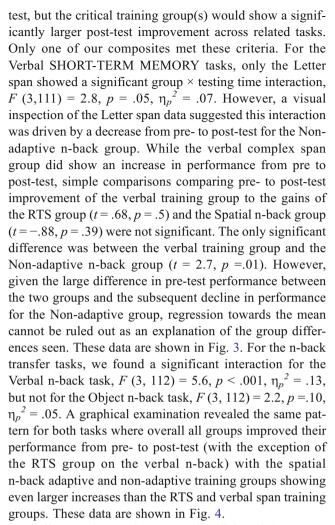
a

b

Fig. 2 The correlations between working memory training improvement in z-scores and (**a**) neuroticism and (**b**) self-reported stress

support for the null. The composites showing strong support for the alternative were Verbal short-term memory (VSTM) and the General N-back. We then examined each of the component tasks within each composite (Letter and Digit Span for the VSTM) (Verbal and Object n-back for the General N-back) using mixed factorial ANOVAs with group as the between-subjects factor and testing time (pre- vs. post-test) as a within-subjects factor. We also graphically examined the pattern of the group pre- to post-test changes in performance to determine whether any significant interaction provided more or less ideal evidence for positive transfer as described by Redick (2015) in which pre-test performance would be similar across groups and all groups would show improvement at post-

test, but the critical training group(s) would show a significantly larger post-test improvement across related tasks. Only one of our composites met these criteria. For the Verbal SHORT-TERM MEMORY tasks, only the Letter span showed a significant group × testing time interaction, $F(3,111) = 2.8$, $p = .05$, $\eta_p^2 = .07$. However, a visual inspection of the Letter span data suggested this interaction was driven by a decrease from pre- to post-test for the Non-adaptive n-back group. While the verbal complex span group did show an increase in performance from pre to post-test, simple comparisons comparing pre- to post-test improvement of the verbal training group to the gains of the RTS group ($t = .68$, $p = .5$) and the Spatial n-back group ($t = -.88$, $p = .39$) were not significant. The only significant difference was between the verbal training group and the Non-adaptive n-back group ($t = 2.7$, $p = .01$). However, given the large difference in pre-test performance between the two groups and the subsequent decline in performance for the Non-adaptive group, regression towards the mean cannot be ruled out as an explanation of the group differences seen. These data are shown in Fig. 3. For the n-back transfer tasks, we found a significant interaction for the Verbal n-back task, $F(3, 112) = 5.6$, $p < .001$, $\eta_p^2 = .13$, but not for the Object n-back task, $F(3, 112) = 2.2$, $p = .10$, $\eta_p^2 = .05$. A graphical examination revealed the same pattern for both tasks where overall all groups improved their performance from pre- to post-test (with the exception of the RTS group on the verbal n-back) with the spatial n-back adaptive and non-adaptive training groups showing even larger increases than the RTS and verbal span training groups. These data are shown in Fig. 4.

Similar to Colom et al. (2013) and Redick et al. (2013), we conducted a post hoc power analysis (G*Power; Faul, Lang, & Buchner, 2007) to ensure that our sample size which we had based on previous reports of working-memory transfer (e.g., Jaeggi et al., 2008) was large enough to detect significant Group × Testing time interactions in our transfer measures. Using the average correlation between our pre- and post-test measurements (average $r = .55$), our power to detect a medium (Cohen's d = .5) or a large effect (Cohen's d = .8) was over .99 and using our lowest correlation $r = .14$ (ANT), power was .94.

Survey data

Personality measures and transfer

We tested for any relationships between the personality measures gathered and the amount of transfer seen for the n-back training group on the verbal and visual n-back tasks as well as the improvements seen for the complex span group on the letter span task. However, unlike the training data, we found no personality measures that were related to transfer in our samples.

Table 3 Descriptive data for the pre- and post-test measures of transfer broken down by group

	Pre-test				Post-test				Pre vs. Post-Test		
	Mean	SD	Min	Max	Mean	SD	Min	Max	p	r	η_p^2
Spatial n-back training											
Verbal N-back	16.6	9.8	−8.0	33.0	29.3	16.2	−16.0	52.0	.001	.57	.48
Object N-back	.40	16.1	−28.0	34.0	15.7	18.9	−26.0	46.0	.001	.42	.40
SOP: Animals	767.7	129.4	567.4	1,222.1	715.2	115.2	526.1	1,020.1	.003	.72	.25
SOP: Word	898.6	169.0	665.4	1,451.2	815.2	147.5	606.3	1,256.2	.002	.64	.28
SOP: Dot	860.7	279.8	549.8	1,796.7	740.9	185.2	460.6	1,247.5	.004	.65	.25
SOP: Array	1,387.1	350.2	823.3	2,211.6	1,221.9	271.5	662.6	1,794.3	.002	.65	.28
Digit Span	13.1	2.9	8.0	20.0	12.7	3.3	7.0	20.0	.52	.61	.01
Letter Span	10.9	2.9	4.0	18.0	11.1	3.9	3.0	19.0	.80	.50	.002
Circle Span	11.4	3.0	4.0	18.0	11.8	3.0	7.0	17.0	.32	.67	.03
Arrow Span	8.5	5.5	1.0	20.0	11.9	3.1	6.0	17.0	.43	-.42	.03
Reading Span	35.4	18.3	6.0	69.0	35.6	16.5	4.0	65.0	.98	.37	.001
Operation Span	36.8	18.4	7.0	71.0	43.0	20.9	3.0	68.0	.07	.48	.11
Letter-Number Span	7.0	3.3	1.0	16.0	7.4	3.3	2.0	14.0	.51	.61	.02
Symmetry Span	14.6	10.1	2.0	34.0	18.7	10.9	4.0	42.0	.02	.78	.20
Rotation Span	24.4	10.6	1.0	42.0	27.2	9.3	4.0	39.0	.03	.77	.15
Alignment Span	10.6	3.4	1.0	17.0	11.1	3.5	4.0	17.0	.35	.65	.03
ANT-AA	56.7	84.5	−17.1	482.4	41.1	28.7	−36.6	94.8	.27	.44	.04
ANT-OA	45.4	23.9	-.21	88.5	44.2	28.6	−21.1	93.8	.83	.40	.002
ANT-EA	109.5	48.1	4.3	200.3	81.7	31.0	20.7	160.5	.001	.78	.46
Simon Task (High Conflict)	169.6	91.6	25.2	350.4	102.4	54.9	1.7	236.3	.001	.62	.46
Simon Task (Low Conflict)	178.7	122.0	17.1	612.4	133.7	76.6	−2.8	284.0	.04	.50	.13
Inferences	5.1	3.8	−6.0	10.0	4.7	3.9	−4.0	10.0	.40	.66	.02
Nonsense Syllogisms	5.0	3.5	−3.0	12.0	4.1	5.1	−7.0	13.0	.39	.17	.03
Ravens SPM	24.0	4.1	16.0	34.0	24.1	4.5	18.0	34.0	.86	.76	.001
Cattell	28.8	4.3	20.0	38.0	31.9	3.6	25.0	38.0	.001	.31	.31
Mathematics Aptitude	.74	4.3	−7.0	9.0	1.4	3.9	−6.0	9.0	.37	.52	.03
Arithmetic Aptitude	4.4	3.9	−2.0	12.0	3.7	5.0	−8.0	12.0	.43	.42	.02
Verbal Complex Span Training											
Verbal N-back	17.4	13.0	−20.0	45.0	19.6	14.7	−15.0	43.0	.44	.36	.02
Object N-back	−1.7	18.9	−41.0	30.0	6.6	18.0	−37.0	38.0	.01	.61	.21
SOP: Animals	790.2	127.5	610.5	1217.1	732.1	104.5	578.7	941.4	.006	.56	.22
SOP: Word	909.6	170.3	618.3	564.1	784.7	156.9	564.1	1,061.9	.001	.53	.39
SOP: Dot	868.0	234.6	480.8	1,554.3	786.9	226.8	442.7	1,375.5	.08	.39	.10
SOP: Array	1,517.1	353.7	939.7	2,303.7	1,409.5	355.3	828.2	2,286.8	.04	.67	.13
Digit Span	12.3	2.7	8.0	17.0	12.4	3.1	4.0	19.0	.78	.43	.003
Letter Span	10.9	3.7	5.0	19.0	11.9	4.0	3.0	19.0	.19	.39	.06
Circle Span	13.4	2.9	7.0	20.0	12.9	2.8	6.0	19.0	.36	.43	.03
Arrow Span	10.1	5.0	1.0	19.0	11.6	4.8	1.0	18.0	.07	.04	.12
Reading Span	31.0	15.1	3.0	56.0	33.4	22.7	6.0	72.0	.55	.65	.01
Operation Span	37.4	19.2	6.0	70.0	46.9	17.8	20.0	75.0	.002	.65	.29
Letter-Number Span	7.1	2.2	3.0	12.0	7.0	2.6	1.0	12.0	.85	.33	.001
Symmetry Span	15.7	8.8	2.0	38.0	17.3	7.8	5.0	32.0	.35	.52	.03
Rotation Span	22.3	8.2	5.0	37.0	26.6	16.3	8.0	37.0	.003	.53	.25
Alignment Span	10.9	4.0	1.0	18.0	12.2	2.9	6.0	18.0	.06	.41	.11
ANT-AA	37.3	30.2	−29.4	102.6	26.5	96.3	−478.3	120.5	.55	.004	.01

Table 3 (continued)

	Pre-test				Post-test				Pre vs. Post-Test		
	Mean	SD	Min	Max	Mean	SD	Min	Max	p	r	η_p^2
ANT-OA	38.0	26.1	−54.5	75.6	40.3	25.6	−25.7	85.9	.63	.47	.008
ANT-EA	99.5	38.0	39.5	205.1	88.8	33.3	16.8	168.3	.11	.47	.08
Simon Task (High Conflict)	198.0	119.5	1.84	546.4	155.2	86.5	49.4	402.6	.01	.59	.18
Simon Task (Low Conflict)	200.9	106.2	46.6	464.7	150.2	97.4	27.4	478.9	.003	.54	.25
Inferences	4.4	3.6	−7.0	10.0	4.9	3.4	−4.0	10.0	.51	.29	.01
Nonsense Syllogisms	3.3	4.5	−7.0	15.0	3.5	5.4	−7.0	15.0	.83	.33	.001
Ravens SPM	22.2	4.1	14.0	28.0	23.0	4.4	15.0	33.0	.26	.60	.04
Cattell	29.4	4.1	23.0	38.0	30.8	3.9	22.0	37.0	.05	.53	.12
Mathematics Aptitude	1.8	3.1	−7.0	7.0	2.3	3.5	−8.0	10.0	.38	.56	.03
Arithmetic Aptitude	3.3	3.5	−4.0	10.0	3.9	4.2	−8.0	11.0	.33	.54	.03
Video Game Training											
Verbal N-back	15.3	12.4	−5.0	43.0	16.0	14.6	−13.0	39.0	.87	.70	.001
Object N-back	−4.5	14.3	−42.0	23	2.8	15.6	−35.0	29.0	.07	.15	.13
SOP: Animals	799.0	148.3	588.9	1,135.5	760.3	146.9	545.4	1,134.8	.03	.82	.17
SOP: Word	902.5	272.4	521.1	1,545.7	838.0	188.5	580.8	1,425.9	.16	.52	.08
SOP: Dot	1,382.5	430.6	760.5	2,520.3	758.8	155.7	507.9	1,096.2	.001	-.09	.65
SOP: Array	1,124.3	487.3	683.6	3,000.6	1,353.2	379.8	766.7	2,503.9	.007	.67	.27
Digit Span	10.4	3.7	2.0	17.0	11.1	2.3	6.0	18.0	.57	.15	.01
Letter Span	10.2	2.0	6.0	18.0	10.5	3.5	6.0	18.0	.52	.59	.02
Circle Span	12.1	2.5	8.0	17.0	13.1	.57	9.0	18.0	.38	.17	.03
Arrow Span	8.7	4.3	2.0	16.0	11.0	3.8	4.0	17.0	.05	-.17	.16
Reading Span	23.1	18.9	3.0	61.0	27.7	19.3	6.0	69.0	.04	.76	.19
Operation Span	34.4	19.8	4.0	68.0	34.8	18.1	3.0	71.0	.97	.70	.001
Letter-Number Span	6.2	2.5	1.0	12.0	6.8	2.5	3.0	11.0	.30	.13	.04
Symmetry Span	13.6	10.1	2.0	37.0	16.4	9.2	2.0	33.0	.07	.64	.13
Rotation Span	25.6	11.5	5.0	41.0	22.5	9.2	6.0	38.0	.11	.62	.10
Alignment Span	10.8	3.6	1.0	15.0	11.9	3.5	3.0	17.0	.08	.63	.12
ANT-AA	46.4	25.2	−23.9	89.5	49.1	28.3	−19.2	129.4	.64	.41	.009
ANT-OA	31.6	33.3	−37.5	103.4	58.2	104.6	−8.9	545.9	.22	.13	.06
ANT-EA	107.4	51.9	31.8	292.0	97.9	34.7	48.5	164.3	.24	.65	.06
Simon Task (High Conflict)	212.1	131.5	45.7	642.9	167.7	127.4	35.6	629.4	.12	.43	.09
Simon Task (Low Conflict)	239.6	150.9	96.7	692.2	182.4	109.0	45.3	504.1	.004	.79	.29
Inferences	4.0	4.5	−4.0	10.0	3.8	4.3	−5.0	10	.82	.35	.002
Nonsense Syllogisms	3.8	3.8	−3.0	13.0	2.7	3.9	−3.0	11.0	.23	.34	.06
Ravens SPM	20.5	4.4	12.0	29.0	22.8	4.6	13.0	29.0	.006	.54	.26
Cattell	27.1	5.1	16.0	37.0	29.9	4.2	19.0	37.0	.005	.53	.28
Mathematics Aptitude	1.4	3.8	−7.0	7.0	1.7	4.3	−6.0	9.0	.71	.47	.006
Arithmetic Aptitude	3.5	3.9	−5.0	10.0	2.8	4.1	−8.0	9.0	.22	.72	.06
Non-Adaptive N-back Training											
Verbal N-back	16.8	14.3	−14.0	43.0	27.7	14.2	−7.0	48.0	.001	.57	.45
Object N-back	2.6	16.0	−37.0	37.0	19.3	15.4	−5.0	49.0	.001	.60	.60
SOP: Animals	810.4	173.3	592.2	1,479.3	712.9	119.8	527.7	1,038.2	.001	.60	.34
SOP: Word	853.5	144.6	634.6	1,201.4	767.9	140.1	532.1	1,069.7	.001	.73	.41
SOP: Dot	787.1	55.7	463.7	1,484.6	719.1	251.2	429.6	1,470.9	.04	.76	.15
SOP: Array	1,430.7	77.2	764.7	2,424.9	1228.1	334.9	621.9	2,085.0	.001	.77	.42

Table 3 (continued)

	Pre-test				Post-test				Pre vs. Post-Test		
	Mean	SD	Min	Max	Mean	SD	Min	Max	p	r	η_p^2
Digit Span	11.7	3.0	3.0	18.0	13.0	4.2	5.0	20.0	.003	.83	.31
Letter Span	13.0	4.2	5.0	20.0	11.4	4.6	4.0	20.0	.01	.73	.22
Circle Span	12.5	3.2	6.0	17.0	13.4	3.3	7.0	19.0	.02	.67	.20
Arrow Span	10.7	5.4	5.0	17.0	12.2	4.3	5.0	19.0	.81	.06	.003
Reading Span	30.1	16.9	3.0	64.0	33.9	19.9	7.0	68.0	.65	.44	.008
Operation Span	36.0	21.1	3.0	68.0	42.0	22.9	3.0	75.0	.04	.78	.15
Letter-Number Span	8.7	3.3	2.0	14.0	8.4	2.9	3.0	13.0	.65	.37	.009
Symmetry Span	16.8	12.3	2.0	38.0	19.9	10.4	2.0	33.0	.06	.83	.14
Rotation Span	24.6	8.2	9.0	42.0	28.9	8.8	10.0	42.0	.001	.78	.38
Alignment Span	11.4	4.0	1.0	17.0	12.8	3.1	2.0	16.0	.08	.43	.11
ANT-AA	46.5	25.4	−35.5	92.0	39.7	28.3	−9.6	90.5	.31	.18	.04
ANT-OA	39.2	18.5	5.2	86.0	46.5	28.2	-.87	120.5	.10	.56	.10
ANT-EA	74.3	105.5	−427.5	187.0	87.2	27.0	45.3	191.2	.60	-.51	.01
Simon Task (High Conflict)	136.9	99.8	−57.6	454.8	128.3	86.4	9.7	450.8	.56	.74	.01
Simon Task (Low Conflict)	164.9	115.9	−21.3	603.5	136.6	72.6	29.8	392.3	.11	.66	.10
Inferences	4.0	3.7	−6.0	10.0	4.9	3.0	−1.0	10.0	.20	.56	.06
Nonsense Syllogisms	3.6	4.9	−7.0	13.0	4.3	3.3	−1.0	9.0	.31	.62	.04
Ravens SPM	22.9	5.6	13.0	32.0	24.0	5.1	15.0	33.0	.12	.85	.09
Cattell	29.3	5.0	20.0	38.0	30.9	4.6	24.0	41.0	.06	.60	.13
Mathematics Aptitude	2.3	4.5	−9.0	9.0	2.7	4.4	−7.0	11.0	.55	.57	.01
Arithmetic Aptitude	4.5	4.0	−4.0	12.0	5.7	3.9	−2.0	13.0	.12	.50	.09

Strategies reported for working-memory training tasks

For verbal complex span, all of our participants reported some sort of chunking strategy for remembering the letters with the most popular being the assignment of a word to each letter and creating sentences with those words. Twenty-four of our participants explicitly reported this strategy while the rest of our participants reported either chunking the letters into people's initials or acronyms. Only one participant reported an explicitly visual strategy in which words were assigned to the letters and then combined into a visual image.

For the n-back training group, there was a greater diversity of strategies reported. The most common strategy was to use the locations where the stimuli appeared to draw shapes or patterns in the mind's eye. Twelve of our participants reported using this method either alone or in combination with counting the number of squares that had appeared as well as tapping or tracing out a corresponding grid on their leg or on the table. Eleven participants reported counting either as the sole strategy or in combination with tapping or drawing shapes. Three participants reported assigning a noise or musical note to each location and trying to keep track of the sequence that way. Finally, seven participants did not report a clear strategy or simply wrote they had no strategy. Amongst our best performers there was no one strategy that predominated.

Motivation and beliefs about training

We also collected open-ended responses to the following questions: (1) how motivated the participant was, (2) how challenging they found the training intervention, and (3) the extent to which they felt it had improved their cognitive function. Participant responses were coded onto a 5-point scale with 1 indicating lack of motivation, challenge, or improvement and 5 indicating a great deal of motivation, challenge, or improvement. Two individuals independently rated participants' answers blind to group membership with an inter-rater reliability of .90. On items where there was disagreement, the average of the two ratings was used as the final value. Overall participants reported being fairly motivated with an average score of 3.8 out of 5. We found group differences both in reported motivation, $F(3,114) = 2.96$, $p = .04$, $\eta_p^2 = .08$ and how challenging the training was perceived, $F(3,114) = 16.9$, $p < .001$, $\eta_p^2 = .32$ with the non-adaptive n-back group reporting being less motivated and challenged. However, there was no significant difference in the degree of improvement participants reported in their cognitive

Table 4 Pre-test correlations between all tasks collapsed across all participants

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	1.0																										
2	.32**	1.0																									
3	-.33*	-.08	1.0																								
4	-.19*	-.19*	.57**	1.0																							
5	-.12	-.11	.26**	.29**	1.0																						
6	-.22*	-.03	.40**	.38**	-.08	1.0																					
7	.22*	.39**	-.01	-.12	-.13	-.05	1.0																				
8	.16	.26**	-.11	-.17	-.24**	-.05	.52**	1.0																			
9	.37**	.21*	-.23*	-.15	-.07	-.12	.42*	.29*	1.0																		
10	.01	-.05	-.05	-.11	-.16	-.09	.01	.10	-.05	1.0																	
11	.36**	.43**	-.11	-.22*	-.21*	-.06	.41**	.38**	.12	.20	1.0																
12	.41**	.36**	-.14	-.22*	-.13	-.19*	.29**	.30**	.13	.004	.61**	1.0															
13	.01	.17	.13	-.05	-.10	-.03	.29**	.17	.18	.03	.14	.20*	1														
14	.30**	.55**	-.10	-.20*	-.17	-.03	.38**	.31**	.24*	.04	.53**	.53**	.14	1													
15	.39**	.42**	-.17	-.20*	.01	-.31**	.36**	.31**	.32**	-.11	.43*	.47**	.15	.14	1												
16	.38**	.29**	-.19*	-.26**	-.18	-.14	.28**	.11	.24*	-.19	.24*	.24*	.12	.32**	.38**	1											
17	.01	-.08	.02	.02	.002	.01	-.06	-.02	-.01	-.08	-.09	.05	.01	-.11	.06	-.10	1										
18	-.002	-.06	-.03	.21*	-.10	.16	.07	-.12	-.11	.05	-.17	-.18	.11	-.14	-.20*	-.10	-.01	1									
19	-.22*	-.11	-.02	-.01	.15	-.09	-.05	-.11	-.07	.08	-.11	-.13	-.09	-.22*	-.10	-.29**	-.07	-.01	1								
20	-.04	-.01	-.08	-.04	.16	.01	-.08	-.10	-.001	-.01	-.01	-.03	-.07	-.05	-.08	.04	-.13	-.06	-.04	1							
21	.03	-.16	-.09	-.003	.08	.04	-.003	.07	.02	-.11	-.08	-.01	-.09	-.08	-.12	.03	-.07	-.06	-.01	.63*	1						
22	.13	.07	-.20*	-.19*	.02	-.24**	.13	.18*	.04	.14	.33**	.19*	.16	.04	.14	-.07	-.12	-.06	.13	-.06	-.08	1					
23	-.09	.04	.06	.03	.02	.08	-.06	.09	-.14	.03	.21*	-.04	.02	.02	-.02	-.04	-.02	-.11	.17	-.05	-.13	.25**	1				
24	.31*	.27*	-.10	-.07	.03	-.02	.38**	.34**	.34**	-.05	.31*	.25**	.26**	.32**	.25**	.15	.02	.002	-.06	-.04	-.03	.30**	.17	1			
25	.44*	.32*	-.19*	-.15	-.11	-.24**	.23*	.22*	.19*	-.08	.38**	.41**	.25**	.42**	.48**	.28**	.002	.03	-.13	.03	.004	.25**	.01	.38**	1		
26	.31*	.22*	-.12	-.20*	-.01	-.16	.01	.21*	.09	-.08	.34**	.33**	.20*	.31**	.30**	.09	.18	-.12	.05	-.02	-.07	.25**	.25**	.32**	.47**	1	
27	.26*	.14	-.12	-.24**	-.07	-.20*	.11	.33**	.09	.21*	.31**	.37**	.29**	.26**	.30**	.13	-.09	-.15	-.08	-.05	-.16	.30**	.23**	.26**	.28**	.42**	1

1 = Verbal N-back, 2 = Object N-back, 3 = SOP animals, 4 = SOP words, 5 = SOP dot, 6 = SOP array, 7 = Digit span, 8 = Letter span, 9 = Circle span, 10 = Arrow span, 11 = Reading span, 12 = Operation span, 13 = Letter-number span, 14 = Symmetry span, 15 = Rotation span, 16 = Alignment span, 17 = ANT Alerting Attention (AA), 18 = ANT Orienting Attention (OA), 19 = ANT Executive Attention (EA), 20 = Simon task High compatibility, 21 = Simon task Low compatibility, 22 = Inferences, 23 = Nonsense, 24 = Ravens, 25 = Cattell, 26 = Mathematics aptitude, 27 = Arithmetic aptitude * p <.05, ** p <.01

Table 5 ANOVA results for transfer analysis at the level of composite variables

Composite	F	p	η_p^2	BF_{01}
General N-back	5.98	.001	.138	.03
Speed of Processing	.25	.86	.01	16.2
Short-term Memory Verbal	21.31	.0001	.36	<.0001
Working-memory Verbal	.05	.99	.001	20.2
Working-memory Visuo-spatial	1.69	.17	.043	3.3
Reasoning Verbal	1.26	.29	.03	5.3
Reasoning Non-verbal	1.57	.20	.04	3.7
Math Performance	.97	.41	.03	7.3

functioning, $F(3,114) = 1.74$, $p = .16$, $\eta_p^2 = .05$, and a majority of participants in each group endorsed recommending the training they participated in to others. These data are reported in Table 7.

Discussion

In this experiment we attempted to compare the results of two forms of working-memory training on the same set of transfer measures and using the same two active control groups. By doing so, we had hoped to elucidate the possible mechanisms by which training on two different working-memory tasks can show transfer to other working-memory measures as well as possibly improve performance on higher-level cognitive tasks such as fluid intelligence and reasoning. However, our data yielded strong evidence only for transfer of training in the case of the spatial n-back to other variants of the n-back task while at the composite level, we found positive support for no transfer of training to measures of speed of processing, working memory, reasoning, and mathematical performance using the Bayes factor. We will discuss our results for each type of working-memory training and how these results relate to previous

Table 6 ANOVA results for individual transfer tasks

| Task	Group			Session			Group x Session			
	F	p	η_p^2	F	p	η_p^2	F	p	η_p^2	BF_{01}
Letter Span	1.5	.221	.04	.002	.97	.0001	2.8	.05	.07	.75
Digit Span	3.1	.03	.07	2.3	.13	.02	1.3	.27	.03	3.3
Circle Span	2.1	.1	.05	3.1	.08	.03	1.7	.16	.04	2.0
Arrow Span	.79	.5	.02	27.6	.001	.2	1.3	.28	.04	7.4
Letter-Number Span	3.4	.02	.09	.29	.59	.003	.55	.65	.02	11.2
Reading Span	1.8	.16	.05	2.7	.10	.03	.36	.78	.01	10.3
Operation Span	.85	.47	.02	13.7	.0001	.11	1.6	.20	.04	3.2
Align Span	.78	.51	.02	10.9	.001	.09	.41	.75	.01	13.8
Symmetry Span	.591	.622	.02	15.1	.0001	.12	.37	.77	.01	11.5
Rotation Span	.63	.59	.02	9.3	.003	.08	6.0	.001	.14	.03
Verbal N-back	2.3	.08	.06	26.7	.0001	.13	5.6	.001	.13	.01
Object N-back	3.8	.01	.09	54.0	.0001	.33	2.2	.10	.06	1.6
Raven's	1.7	.17	.04	10.2	.002	.08	1.8	.15	.05	2.2
Cattell	1.3	.27	.04	30.1	.001	.21	1.1	.34	.03	5.9
ANT AA	1.2	.33	.03	1.3	.25	.01	.34	.79	.009	15.1
ANT OA	.21	.89	.006	2.9	.08	.03	1.4	.25	.04	4.1
ANT EA	1.5	.22	.04	1.9	.18	.02	1.7	.18	.04	2.7
Simon LC	4.5	.005	.11	1.5	.21	.01	.57	.64	.02	11.8
Simon HC	7.8	.001	.18	2.9	.09	.03	.10	.96	.003	3.4
Animal Task	.47	.71	.01	36.8	.001	.25	1.4	.23	.04	4.6
Word Task	.68	.57	.02	34.8	.001	.24	.63	.60	.02	10.5
Dot Task	12.2	.001	.25	66.6	.001	.37	22.4	.001	.38	<.0001
Matching Task	2.4	.07	.06	5.4	.02	.05	11.0	.001	.23	<.0001
Arithmetic	1.5	.21	.04	.09	.77	.001	1.7	.17	.04	3.0
Math	.87	.46	.02	1.8	.18	.02	.04	.98	.001	20.6
Inferences	.49	.69	.01	1.3	.16	.001	.69	.56	.02	9.9
Nonsense	.85	.47	.02	.27	.61	.002	.86	.47	.02	8.5

a

b
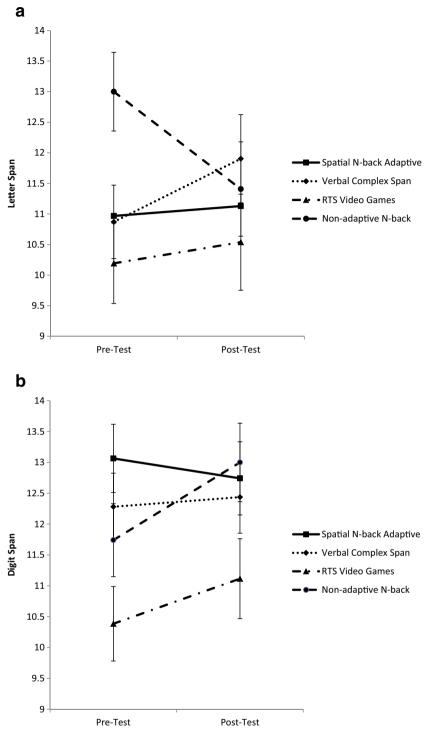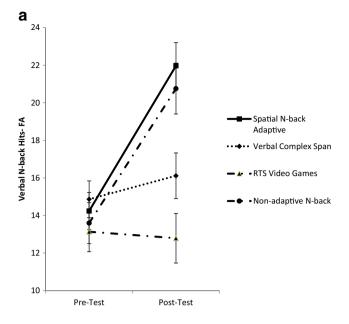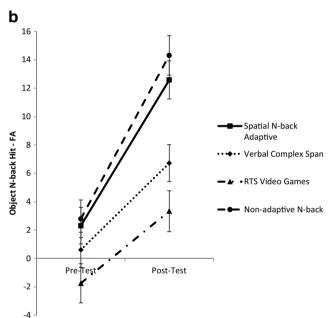
Fig. 3 The improvements for each of the training groups from pre-test to post-test for the Verbal short-term memory tasks (**a**) Letter Span and (**b**) Digit Span. *RTS* real-time strategy

work before moving to the larger picture of working-memory training and transfer.

We begin with the verbal complex span training as we were surprised to see no evidence for transfer to either verbal short-term memory or span tasks in spite of the robust training gains in which our participants showed

improvements equivalent to or greater than previous studies (36 % increase compared to the 25 % increase reported in Chein & Morrison (2010)). A number of laboratories have demonstrated transfer from complex span training to other span-based measures (Chein & Morrison, 2010; Harrison et al., 2013; Richmond, Wolk, Chein, & Olson,

Fig. 4 The improvements for each of the training groups from pre-test to post-test for the (**a**) Verbal n-back transfer task and (**b**) the Object n-back task. *RTS* real-time strategy

Ceci, 2002; Schmidt & Bjork, 1992) as most other studies using this form of working-memory training have used at least two variants of the tasks. This appears to be unlike n-back training in which equivalent effects have been reported for using a single or dual n-back paradigm (Jaeggi et al., 2010). Another possibility is that this form of working-memory training is more effective in fostering transfer when using spatial materials, although Gibson et al. (2011) reported equal transfer for groups of ADHD participants training on verbal or spatial versions of CogMed with each training program consisting of multiple training tasks. Gibson and colleagues (2013) also have reported that lowering the accuracy threshold for the items to be recalled from 100 % (as in our study) to 80 % can increase the amount of transfer seen. Therefore, it appears that unlike the n-back task, training using a lone complex span task is not effective in producing transfer.

Within our verbal span group, there was clear evidence of an almost universal strategy during training. All but one of our participants described using some form of chunking to help them remember the series of letters. However, while our verbal span training group did show a pre- to post-test increase on the Letter span task, the only group comparison that was significant was to a group showing a pre- to post-test decrease in performance, and there was no evidence of transfer to any other transfer task using letters such as the letter-number span or the letter n-back. It may be that the gains seen during the verbal complex span training were due at least in part to participants becoming more efficient at the lexical decision task which allowed them to better attend to the letters and apply chunking strategies in that specific context which then did not generalize to other span tasks even those using letters as stimuli. Therefore, it appears that single domain training using a span task is unlikely to yield any transfer, and the broader transfer seen in other studies may rely on training with multiple complex span tasks which allows participants to attend to the shared task structure and learn to shift between task demands more efficiently.

In contrast to the complex span training, our spatial n-back group did show clear evidence for transfer of training to letter and object variants of the n-back task compared to our complex span and video game control group. These results are similar to Sprenger et al. (2013), who reported only finding evidence for transfer to tasks similar in task structure or stimuli

2014; von Bastian & Oberauer, 2013). However, it may be that such transfer requires variability in training (Barnett &

Table 7 Means and standard deviations for post-test survey on participant perceptions of training and transfer

Group	Motivation	Challenge	Improvement	Recommend
Spatial N-back	4.0 (0.9)	3.9 (0.8)	3.7 (1.0)	83 %
Verbal Complex Span	3.8 (0.8)	3.3 (0.8)	3.7 (1.0)	93 %
Video Game	4.1 (0.8)	4.4 (0.7)	3.3 (1.0)	92 %
Non-adaptive N-back	3.4 (1.2)	2.9 (1.0)	3.9 (1.2)	77 %

to the training tasks. Their participants who had the n-back task as part of their training showed improvement on the n-back, but not to other working-memory measures. Although a number of our participants reported using strategies such as mentally tracing shapes to help keep track of the sequence of locations, we observed no measurable transfer to either of the visuo-spatial short-term memory tasks which involved keeping track of locations or to the three visuo-spatial working-memory tasks (Symmetry, Rotation, or Alignment spans). Other research has also reported no evidence of transfer of training to SHORT-TERM MEMORY or working-memory measures (Lilienthal, Tamez, Shelton, Myerson, & Hale, 2013; Schwab et al., 2015; Sprenger et al., 2013). Therefore, training on the n-back would not seem to increase the capacity of working memory and the transfer we observed did not appear to be related to the self-reported strategies used during training. We also saw a similar improvement on the letter and object n-back transfer tasks for the non-adaptive n-back group, which raises two possibilities: (1) participants in both n-back groups improved on a cognitive process common to n-back tasks, or (2) that the similarity in task structures led to an increased confidence or expectation that they could perform well on the task. In terms of cognitive processing, participants may have improved at familiarity based responding which has been shown to be a key component to n-back performance (Kane, et al., 2007). This was supported by our data, indicating that the primary locus of improvement was in the reduction of the number of false alarms as opposed to an increase in the number of hits. Unfortunately, we did not systematically manipulate the type of lures in our n-back paradigms to test whether our n-back groups would be less susceptible to near lures. Other research groups have also suggested that the ability to inhibit irrelevant information in working memory may be the locus of n-back training (Jaeggi et al., 2014; Schwarb et al., 2015). In terms of our study, improvements in inhibitory processing could fit our n-back data though we did not find any evidence for transferable improvement to the Simon task or the ANT. Sprenger et al. (2013) also did not find transfer to Stroop or antisaccade and only reported evidence for transfer to a flanker task when the training regimen included a flanker-like task. Therefore, while transfer of training was observed, we cannot rule out task-related expectancies, i.e., participants expect to improve on tasks similar to those used at training as the source.

While we were able to observe transfer to other n-back tasks, we were unable to replicate the transfer of training reported by some to other measures of fluid intelligence and reasoning (Jaeggi et al., 2014; Jaušovec & Jaušovec, 2012; Rudebeck et al., 2012; Stephenson & Halpern, 2013). In this our results mirror those of other researchers who have also been unable to demonstrate transfer to reasoning and fluid ability tasks from working-memory training (Chooi & Thompson, 2012; Colom et al., 2013; Harrison et al., 2013;

Heinzel, et al., 2014; Redick et al., 2013; Schwarb, Nail, & Schumacher, 2015; Sprenger, et al., 2013; Thompson, et al., 2013; von Bastian & Eschen, 2015). One possible reason for the lack of far transfer to the reasoning measures in our study is that the training improvements seen in our spatial n-back group were smaller than earlier reports (e.g., Jaeggi et al., 2010). Correlations between the amount of training improvement and the amount of transfer observed have been viewed by some researchers as evidence that transfer is moderated by training gains (Jaeggi et al., 2011). However, such responder analyses (i.e., individuals showing the most gains during training have larger transfer) have recently been criticized as logically insufficient and unusable as either support for or against the effectiveness of any particular training program (Tidwell et al., 2014). In addition, two recent meta-analyses of working-memory training have not found the rate of improvement over training to be a significant predictor of transfer (Au et al., 2015; Karbach & Verhaeghen, 2014).

A second possibility is participants' motivation for participating. Jaeggi et al. (2014) proposed that one explanation for why some studies find transfer to fluid intelligence and others do not may depend on how intrinsically motivated the participants are and that paying participants may undermine intrinsic motivation. We did not rely upon unpaid volunteers, but we did measure our participants' reported levels of motivation and whether they believed the training had improved their cognitive performance. Overall they reported feeling motivated and that they had benefited cognitively from the training. However, we did not measure how motivated our participants were at the beginning of the study, their reasons for participating in the study, or whether the degree of challenge or interest they had in their training task changed across the training.

Personality differences have also been raised as a factor that affects transfer of training (Jaeggi et al., 2014). It may be that individuals who volunteer and persist in a multi-week training intervention with little to no financial incentive may be systematically different in either personality or cognitive abilities than those who participate for compensation. Jaeggi et al. (2014) found that volunteers self-reported more cognitive failures and described these participants as having "some self-perceived deficit that many have influenced their interest in improving their memory and cognitive performance in the first place" (Jaeggi et al., 2014). While we did measure a number of personality and individual differences that we thought might contribute to training improvements and possible transfer, similar to previous studies, the few correlations seen between training and personality variables were small and underpowered. Studer-Luethi et al. (2012) reported weak correlations between neuroticism and training performance ($r = -.24$) and conscientiousness and training performance ($r = .28$). We did see a similar relationship between neuroticism and training ($r = -.25$) as well as between self-reported

stress and training ($r = -.26$). However, we did not see any relationships between training and conscientiousness or related variables such as need for cognition, grit, or self-control nor evidence of a relationship between personality and the near transfer observed.

We will conclude by discussing two challenges for our study and future studies of working-memory training: What constitutes an appropriate control group and the problem of small effect sizes. We embarked on a fairly large training study confident based on initial reports that we would observe substantial transfer of training. We used an active control group similar to those used in previous studies intended to control for the basic task structure and materials of one of our working-memory training groups while omitting the adaptive component of training that was believed to be important for transfer (Klingberg et al., 2005). Instead our non-adaptive n-back group showed the same evidence of improvement on the transfer n-back tasks as our n-back training group. This is consistent with very recent research suggesting that neither the length nor number of training sessions (Au et al., 2015) nor an adaptive element to training is necessary for transfer (Karbach & Verhaeghen, 2014; Vartanian et al., 2013; von Bastian & Eschen, 2015). However, this does leave open the question of what is being trained and how best to control for placebo effects if factors such as training duration do not affect transfer. We also used a RTS video game group as a second active control group to control for possible differences in engagement and expectancies for improvement between our first control and our training groups. While our first control group can be criticized as not a "control" group, but rather a training group based on the similar transfer seen, our second control group could also be criticized as possibly being more of a training group even though prior studies do not support that interpretation as Boot et al. found no transferable improvement after 21 h of training (although see Basak et al., 2008 for evidence of transfer in older adults). Glass, Maddox, and Love (2013) reported improvements in tasks measuring cognitive flexibility in a group trained on Starcraft, but only after 40 h of carefully managed play emphasizing maintained awareness of resources and rapid switching. A second group played Starcraft for the same 40 h, but without these specific demands, and showed no improvement. Additionally, video game training would not control for task-specific expectancies for improvement such as those possibly seen in our non-adaptive n-back control group. Given that the choice of control group has been shown to affect the observation of transfer across studies (Au et al., 2015; Melby-Lervåg, & Hulme, 2013), the question of what constitutes an appropriate active control merits greater attention in the field.

A second major concern is that recent estimates of effect sizes especially for far transfer to reasoning and fluid intelligence appear to be much smaller than reported in the earliest working-memory studies. Therefore, perhaps our study did not have sufficient power to detect a training-related change in fluid intelligence. This seems possible given the results of recent meta-analyses (Au et al., 2015; Karbach & Verhaeghen, 2014; Melby-Lervåg & Hulme, 2013) reporting small effect sizes for far transfer. Based on these estimates, our current study would only have had a power of .81 to detect a group difference in Ravens ($r = . 8$) using a small effect size ($f = .10$) and we would have needed a sample size of 176 participants to achieve power of .95. Karbach and Verhaeghen (2014) have proposed that samples sizes of over 300 participants were necessary to achieve 80 % power and concluded that most studies to date have been underpowered. If this is correct, then power is a problem not only for the current study, but for most research conducted on working-memory training and transfer and the conclusions drawn from them.

Working-memory training and transfer remains a contentious area of research with some arguing that the evidence for far transfer is clear (Au et al., 2015; Jaeggi et al., 2014), while others remain skeptical (Haier, 2014; Redick, Melby-Lervåg, Hulme, & Grovak, 2014; Shipstead, Redick, & Engle, 2012). Certainly the results from our study and many others would suggest that the mechanisms underlying transfer are still not well understood. In our study, it would not appear that working-memory capacity was increased directly nor was the efficiency of working memory improved through the development and practice of different strategies. Instead our clearest evidence of transfer was from training on a spatial n-back, both adaptive and non-adaptive, to non-spatial n-back tasks. The use of Bayesian analyses was helpful in evaluating the extent to which our dataset favored the null hypothesis. In the cases of speed of processing and verbal working memory, the odds were greater than 10:1 in favor of the null model while more modest odds of approximately 3:1 in favor of the null were seen for visuo-spatial reasoning and working memory.

Conclusion

Our understanding of working-memory training continues to evolve at a rapid pace although key questions about the origins and breadth of transfer still remain. Given the large effect sizes reported in the earlier literature, we expected to observe far transfer in our study. However, it may be that such transfer is most easily found in certain subsets of the population and the effect size is much smaller than originally reported. Both merit serious consideration in future working-memory research and have important implications for the utility of working memory training in institutional settings such as schools (e.g., Rode, Robson, Purviance, Geary, & Mayr, 2014).

Appendix

1. Please describe in detail what strategies you used to improve your performance on the training task?
2. Please describe how motivated you were to improve on the training task?
3. Did you find the training task to be challenging? Why or why not?
4. Do you think the training task improved your cognitive ability? If so, were there any tasks in the post-test that you think you did much better at because of the training?
5. Would you recommend this brain training program to others? Why or why not?

References

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review, 22,* 366–377.

Baddeley, A. (1992). Working memory. *Science, 255,* 556–559.

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65–94. doi:10.1037/0033-2909.121.1.65

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.

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can training in a real-time strategy video game attenuate cognitive decline in older adults? Psychology and Aging, 23, 765–777. doi:10.1037/a0013494

Boot, W. R., Kramer, A. F., Simons, D. J., Fabiani, M., & Gratton, G. (2008). The effects of video game playing on attention, memory, and executive control. *Acta Pyschologica, 129,* 387–398.

Boot, W. R., Blakely, D. P., & Simons, D. J. (2011). Do action video games improve perception and cognition? *Frontiers in Psychology, 2,* 226. doi:10.3389/fpsyg.2011.00226

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives in Psychological Science, 8,* 445–454. doi:10.1177/1745691613491271

Caccioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Cattell, R. B. (1961). *The Culture Free Intelligence Test, Scale 3.* Champaign, IL: Institute for Personality and Ability Testing.

Chen, J., Hale, S., & Myerson, J. (2007). Predicting the size of individual and group differences on speeded cognitive tasks. *Psychonomic Bulletin & Review, 14,* 534–541. doi:10.3758/bf03194103

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.

Chooi, W.-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence, 40,* 531–542.

Colom, R., Román, F. J., Abad, F. J., Chun Shi, P., Privado, J., Froufe, M., Escorial, S., Martínez, K., Burgaleta, M., Quiroga, M. A., Karama, S., Haier, R. J., Thompson, P. M., & Jaeggi, S. M. (2013). Adaptive n-back training does not improve fluid intelligence at the construct level: Gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41,* 712–727.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24,* 385–396.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review, 3,* 422–433.

Dougherty, M. R., Hamovitz, T., & Tidwell, J. W. (2016). Reevaluating the effectiveness of the n-back training on transfer through the Bayesian lens: Support for the null. *Psychonomic Bulletin and Review, 23,* 306–316.

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment, 91,* 166–174.

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.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11,* 19–23.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience, 14,* 340–347.

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

Gibson, B. S., Gondoli, D. M., Johnson, A. C., Steeger, C. M., Dobrzenski, B. A., & Morrissey, R. A. (2011). Component analysis of verbal versus spatial working memory training in adolescents with ADHD: A randomized, controlled trial. *Child Neuropsychology: A Journal on Normal and Abnormal Development in Childhood and Adolescence, 17*(6), 546–563. doi:10.1080/09297049.2010.551186

Gibson, B. S., Gondoli, D. M., Kronenberger, W. G., Johnson, A. C., Steeger, C. M., & Morrissey, R. A. (2013). Exploration of an adaptive training regimen that can target the secondary memory component of working memory capacity. *Memory & Cognition, 41,* 726–737.

Glass, B. D., Maddox, W. T., & Love, B. C. (2013). Real-time strategy game training: Emergence of a cognitive flexibility trait. *PLOS One, 8*(8), e70350.

Goldman-Rakic, P. S. (1994). Working memory dysfunction in schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences, 6*(4), 348–357.

Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in Systems Neuroscience, 8,* 34. doi:10.3389/fnsys.2014.00034

Hale, S., Rose, N. S., Myerson, J., Strube, M. J., Sommers, M., Tye-Murray, N., & Spehar, B. (2011). The structure of working memory abilities across the adult life span. *Psychology & Aging, 26,* 92–110. doi:10.1037/a0021483

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.

Heinzel, S., Schulte, S., Onken, J., Duong, Q., Riemner, T.G., Heinz, A.···Rapp, M.A. (2014). Working memory training improvements and gains in non-trained cognitive tasks in young and older adults. *Aging, Neuropsychology and Cognition, 21,* 146–173.

Jaeggi, S. M., Buschkuehl, 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.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18,* 392–412.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 10081–10086.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition, 42,* 464–480. doi:10.3758/s13421-013-0364-z

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning – implications for training and transfer. *Intelligence, 38,* 625–635.

JASP Team (2016). JASP (Version 0.7.5.5)[Computer software].

Jaušovec, N., & Jaušovec, K. (2012). Working memory training: Improving intelligence Changing brain activity. *Brain and Cognition, 79,* 96–106. doi:10.1016/j.bandc.2012.02.007

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The big five inventory—versions 4a and 54*. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York, NY: Oxford University Press.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology: Learning, Memory & Cognition, 33,* 615–622.

Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological Science, 25,* 2027–2057.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association, 90,* 773–795.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Control* (pp. 184–223). New York: Cambridge University Press.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24,* 781–791.

Klingberg, T., Fernell, E., Olesen, P. J., 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 and Adolescent Psychiatry, 44,* 177–186.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14,* 389–433.

LeFevre, J. A., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 361–378). New York: Psychology Press.

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual *n*-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review, 20,* 135–141. doi:10.3758/s13423-012-0335-6

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49,* 270–291.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Patton, J. H., & Stanford, M. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51,* 768–774.

Powers, K. L., Brooks, P. J., Aldrich, N. J., Palladino, M. A., & Alfieri, L. (2013). Effects of video-game play on information processing: A meta-analytic investigation. *Psychonomic Bulletin & Review, 20,* 1055–1079. doi:10.3758/s13423-013-0418-z

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. New York, NY: Psychological Corporation.

Redick, T. S. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence, 50,* 14–20.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: A meta-analysis. *Psychonomic Bulletin & Review, 20,* 1102–1113.

Redick, T. S., Melby-Lervåg, M., Hulme, C., & Grovak, A. (2014). *PowerPoint presentation at the 55th Annual Meeting of the Psychonomic Society. Long: Beach, CA.* **Does working memory training improve intelligence? A meta-analytic update**.

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. doi: 10.1037/a0029082.

Richmond, L. L., Wolk, D., Chein, J., & Olson, I. R. (2014). Transcranial direct current stimulation enhances verbal working memory training performance over time and near transfer outcomes. *Journal of Cognitive Neuroscience, 26*(11), 2445–2454. doi:10.1162/jocn_a_00657

Rode, C., Robson, R., Purviance, A., Geary, D. C., & Mayr, U. (2014). Is working memory training effective? A study in a school setting. *PLOS One, 9*(8), e104796.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374.

Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. H. (2012). A potential spatial working memory training task to improve both episodic memory and fluid intelligence. *PLOS One, 7*(11), e50431. doi:10.1371/journal.pone.0050431

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063–1078.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3,* 207–217.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138,* 628–654. doi:10.1037/a0027473

Simon, J. R., Acosta, E., Mewaldt, S. P., & Speidel, C. R. (1976). The effect of an irrelevant directional cue on choice reaction time: Duration of the phenomenon and its relation to stages of processing. *Perception & Psychophysics, 19,* 16–22.

Sprenger, A.M., Atkins, S.M., Bolger, D.J., Harbison, J.I., Novick, J.M., Chrabaszcz, J.S.···Dougherty, M.R., (2013). Training working memory: Limits of transfer. *Intelligence, 41,* 638–663.

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41,* 341–357.

Sternberg, R. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences of the United States of America, 105,* 6791–6792.

Studer-Luethi, B., Jaeggi, S. M., Buschkuehl, M., & Perrig, W. J. (2012). Influence of neurotisicm and conscientiousness on working memory training outcome. *Personality and Individual Differences, 53,* 44–49. doi:10.1016/j.paid.2012.02.012

Swanson, H.L. (2006). Working memory and dynamic testing in children with learning disabilities. In S.J. Pickering (Ed.) *Working memory and Education.* (pp. 125–156). Elsevier Press.

Schwarb, H., Nail, J., & Schumacher, E. H. (2015). Working memory training improves visual short-term memory capacity. *Psychological Research*. doi:10.1007/s00426-015-0648-y

Snyder, C. R., Harris, C., Anderson, J. R., Holleran, S. A., Irving, L. M., Sigmon, S.···Harney, P. (1991). The will and the ways: development and validation of an individual-differences measure of hope. *Journal of Personality and Social Psychology, 60,* 570–585.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72,* 271–324.

Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., Chang, P., Pollard, K., Lala, N., Alvarez, G. A., & Gabrieli, J. D. E. (2013). Failure of working memory training to enhance cognition or intelligence. *PLoS ONE, 8*(5), e63614. doi:10.1371/journal.pone.0063614

Tidwell, J. W., Dougherty, M. R., Chrabaszcz, J. R., Thomas, R. P., & Mendoza, J. L. (2014). What counts as evidence for working memory training? Problems with correlated gains and dichotomization. *Psychonomic Bulletin & Review, 21,* 620–628.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8,* 247–261.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language, 28,* 127–154.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132.

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.

Vartanian, O., Jobidon, M. -E., Bouak, F., Nakashima, A., Smith, I., Lam, Q., et al. (2013). Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task. *Neuroscience, 236,* 186–194. http://dx.doi.org/10.1016/j.neuroscience.2012.12.060.

von Bastian, C. C., & Eschen, A. (2015). Does working memory training have to be adaptive? *Psychological Research*. doi:10.1007/s00426-015-0655-z

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.

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychological Research, 78,* 803–820.