Targeting the Three Stages of Retrieval from Secondary Memory in a Double-Blinded, Placebo-Controlled, Randomized Working Memory Training Study

Kathryn J. Ralph, Bradley S. Gibson, Dawn M. Gondoli, Pedro Sztybel, Joseph R. Pauszek, Robert W. Miller, and Emily Litzow

University of Notre Dame

Manuscript Published in the Journal of Cognitive Enhancement (2017) 1:455-477

Author note:

Kathryn J. Ralph, Department of Psychology, University of Notre Dame

Bradley S. Gibson, Department of Psychology, University of Notre Dame

Dawn M. Gondoli, Department of Psychology, University of Notre Dame

Pedro Sztybel, Department of Psychology, University of Notre Dame

Joseph R. Pauszek, Department of Psychology, University of Notre Dame

Robert W. Miller, Department of Psychology, University of Notre Dame

Emily Litzow, Department of Psychology, University of Notre Dame

Please send all correspondence to:

Kathryn J. Ralph

Department of Psychology

106 Haggar Hall

University of Notre Dame

Notre Dame, IN 46556

kralph@nd.edu

RUNNING HEAD: SECONDARY MEMORY TRAINING

Abstract

Working memory (WM) is the ability to temporarily store and retrieve a limited amount of information during complex cognitive activities, especially in the face of distraction. The dual-component model describes WM as including active maintenance in primary memory (PM) and cue-dependent search and retrieval from secondary memory (SM). Previously, researchers have found that WM training (WMT) fails to enhance SM capacity, a component that mediates the relationship between WM and fluid reasoning (gF). Thus, a double-blinded, placebo-controlled, randomized trial was conducted to elucidate whether retrieval from SM could be targeted using a two-component WMT regimen versus two control conditions: adaptive one-component WMT targeting solely PM capacity and non-adaptive one-component WMT. Participants were 174 adolescents, aged 10 to 13 years, who were assessed before, after, and six-moths following training. Retrieval from SM was measured using delayed free recall tasks, far transfer to gF was assessed with matrix reasoning and verbal inference tests, and far transfer to academic performance was assessed with reading and math tests. It was predicted that solely two-component WMT would enhance retrieval from SM and result in far transfer. ANCOVAs with pre-test scores as the covariate indicated that two-component participants increased total errors over controls. There were no significant differences between the groups on recall latency, total correct, or gF measures. The non-adaptive one-component group significantly improved on reading, although a drop in the other two groups drove the effect. Additional research is needed to elucidate whether theoretically-motivated WMT can positively impact higher-level cognition through SM retrieval mechanisms.

Keywords: working memory, working memory training, secondary memory

Conflict of Interest Disclosure

During the course of this study and the writing of this report, Kathryn J. Ralph, M.A. worked part time as a content specialist for the Clinical Assessment division of Pearson Education, which is the distributor of Cogmed Working Memory Training. All other authors have no conflict of interest to disclose.

Targeting the Three Stages of Retrieval from Secondary Memory in a Double-Blinded, Placebo-Controlled, Randomized Working Memory Training Study

Working memory (WM) is a multifaceted cognitive ability that allows for the temporary storage and retrieval of a limited amount of goal-relevant information during complex cognitive activities, especially in the face of distraction (Kane & Engle, 2002; Unsworth & Engle, 2007a). As a construct, WM has taken a core role in modern theories of cognition because it strongly relates to our ability to solve novel complex problems and adapt to new situations in daily life (i.e., fluid reasoning (gF)) (Carpenter, Just, & Shell, 1990; Catell, 1943; de Abreu, Conway, & Gathercole, 2010; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, Conway, 2005). Furthermore, WM capacity is predictive of a variety of higher order abilities including performance on verbal and mathematical aptitude tests in children (Cowan et al., 2005; Gathercole & Pickering, 2000) and adults (Cowan et al., 2005; Daneman & Carpenter, 1980; Turner & Engle, 1989).

Working Memory Training Background

Given the crucial role WM plays in everyday functioning, there has been increased interest in whether it can be enhanced by computerized WM training (WMT). Often such training paradigms include participants practicing on simple span, complex span, or n-back tasks that adapt in difficulty and therefore push participants to continually challenge their WM capacity. At the basis of this method is the contention that sustained and intense practice on WM tasks should lead to enhancement of neural substrates underlying WM (Klingberg, 2010) and by extension result in near transfer to unpracticed WM tasks and far transfer to more distal abilities such as gF and academic performance (Barnett & Ceci, 2002).

However, despite the promise of early studies, evidence of far transfer to gF and academic performance measures following WMT has been far from definitive. For instance, using the n-back training paradigm, researchers have demonstrated training-related improvements on single (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008) and composite gF measures (Jaeggi, Buschkuehl, Shah, & Jonides, 2014). However, these results have been tempered by numerous failed replication studies (Chein & Morrison, 2010; Chooi & Thompson, 2012; Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Redick et al., 2013; Seidler et al., 2010; Thompson et al., 2013). Additionally, it is not well understood whether tasks like the n-back are measuring the same construct as the well validated complex WM span task (Jaeggi, Buschkeuhl, Perrig, & Meier, 2010; Kane, Conway, Miura, & Colflesh, 2007) and therefore, whether improvements following n-back training truly represent the enhancement in WM that would theoretically lead to far transfer.

Similarly, the commercially available simple span WMT called Cogmed Working Memory Training™ (Cogmed; Pearson Education, Inc.) has sometimes (Caeyenberghs, Metzler-Baddeley, Foley, & Jones, 2016; Klingberg, Forssberg, & Westerberg, 2002; Klingberg et al., 2005) but not always (Brehmer, Westerberg, & Bäckman, 2012; Dahlin, 2011; Dunning, Holmes, & Gathercole, 2013; Holmes et al., 2010; van Dongen-Boomsma, Vollebregt, Buitelaar, & Slaats-Willemse, 2014) led to improvement on measures of gF or verbal or performance IQ. Such mixed results are also apparent on academic outcome measures following Cogmed, with participants showing no improvement on academic achievement and classroom performance measures compared to controls in numerous reports (Bigorra, Garolera, Guijarro, & Hervás, 2016, Chacko et al., 2014; Dunning et al., 2013; Gray et al., 2012; Partanen, Jansson, Lisspers, & Sundin, 2015; Roberts et al., 2016; van der Donk, Hiemstra-Beernink, Tjeenk-Kalff, van der Leij, & Lindauer, 2015; van Dongen-Boomsma et al., 2014). Although researchers have proposed other explanations for the lack of far transfer to academic performance following Cogmed (Bergman Nutley & Söderqvsit, 2017), it is worth noting that studies in which there has been non-significant far transfer to academic achievement are almost exclusively rigorously designed, randomized trials with active control groups.

These discrepant findings surrounding far transfer have led to numerous meta-analyses and reviews challenging whether WMT is a viable method for inducing cognitive enhancement (Au, Buschkuehl, Duncan, & Jaeggi, 2016; Au, Sheehan, Tsai, Duncan, Buschkuehl, & Jaeggi, 2015; Cortese et al., 2015; McCabe, Redick, & Engle, 2016; Melby-Lervåg & Hulme, 2013; Melby-Lervåg, Redick, & Hulme, 2016; Peijnenborgh, Hurks, Aldenkamp, Vles, & Hendriksen, 2016; Schwaighofer, Fischer, & Bühner, 2015; Shipstead, Redick, & Engle, 2012; Spencer-Smith & Klingberg, 2015). In response to these evaluations, some researchers have suggested that methodological inconsistencies across studies (e.g., varying sample characteristics, training paradigms, gaming elements, control groups, outcome measures) may explain the divergent results or hamper clear interpretations of the literature at the meta-level (Schwaighofer et al., 2015; Simons et al., 2016). However, another explanation put forth by our group is that existing paradigms may fail to target the appropriate theoretical mechanisms underlying WM (Gibson et al., 2013; Gibson et al., 2012).

Dual-component Model of WM

One modern WM theory, the dual-component model (Unsworth and Engle, 2007a), holds that WM is comprised of both primary memory (PM) and secondary memory (SM) components. PM refers to the active maintenance of a limited amount of goal-relevant information over the short term, specifically in the face of internal and external distraction or interference. Given that PM capacity is limited by both storage and attentional resources, some to-be-remembered items may not be maintained and must be recovered through strategic, cue-dependent search of long-term memory or SM (Mogle, Lovett, Stawski, & Sliwinksi, 2008; Unsworth, 2007, 2009, 2016; Unsworth & Engle, 2007a, 2007b; Unsworth & Spillers, 2010). Not only do both PM and SM capacities distinguish between high and low WM capacity individuals (Unsworth, 2007, 2009, 2016; Unsworth & Engle, 2007a) but also, recent work has shown that both PM and SM capacity account for 78% of the variance in gF (Unsworth, Fukuda, Awh, & Vogel, 2014). It thus stands to reason that for WMT to effectively lead to near and far transfer, training paradigms must target both components of WM.

Importantly, the dual-component model is an appropriate framework within to evaluate the efficacy of simple span WMT because the theory itself is rooted in analysis of span task performance. Indeed, Unsworth and Engle (2007b) used the dual-component model to explain the stronger correlation that is typically found between complex span tasks and higher-order cognition than between simple span tasks and higher-order cognition. In particular, they argued that simple span and complex span tasks both measure the same processes (rehearsal, maintenance, and updating in PM, as well as retrieval from SM), albeit to different extents.

For instance, using confirmatory factor analysis, Unsworth and Engle showed that simple span tasks such as forward word span loaded more highly on the PM factor than on the SM factor; whereas, complex span tasks such as operation span loaded more highly on the SM factor than on the PM factor. However, successful recall in simple span tasks may increasingly measure SM ability (as opposed to PM capacity) as list-length increases beyond the capacity of PM. Indeed, Unsworth and Engle (2006) showed that simple span tasks can predict fluid IQ scores just as well as complex span tasks when list-length exceeds the capacity of PM (greater than 4 items) because the simple span task now measures both the PM and SM components of WM capacity. Thus, according to this analysis, both simple span and complex span tasks appear capable of targeting the PM and SM components of WM capacity under certain conditions.

However, although simple span tasks are appropriate measures of PM and SM components, there is some evidence to suggest that the Cogmed simple span WMT is currently designed to target solely PM capacity (Gibson et al., 2011). In fact, experimental studies comparing training on spatial versus verbal Cogmed tasks and simple versus complex Cogmed tasks have failed to improve SM capacity (Gibson et al., 2013; Gibson et al., 2012). There thus remains a critical need to evaluate whether SM capacity can be enhanced by simple span WMT.

Recall Accuracy Threshold

One suggestion for targeting SM capacity has been to reduce the recall accuracy threshold (RAT) during training. The RAT describes the number of items that must be correctly recalled in order for a span length to increase. For instance, historically the standard algorithms for Cogmed (Klingberg et al., 2005), as well as other training paradigms (Chein & Morrison, 2010), have required that 100% of items in a given trial be remembered correctly for span lengths to increase (i.e., RAT = 100%) (Klingberg et al., 2005). Furthermore, regardless of whether the training is on simple, complex, or n-back tasks, 100% accurate recall leads the span length (or the number of items back in the sequence for n-back) to increase by one item (Harrison et al., 2013; Jaeggi et al., 2008). Indeed, even in more recent iterations of Cogmed, the 100% RAT has remained the standard, although now participants are penalized for the severity of their incorrect responses (i.e., the number of incorrect items recalled on each trial).

However, although the 100% RAT remains the standard for WMT paradigms, Gibson et al. (2013) proposed that the 100% RAT was perhaps too stringent. Indeed, it was hypothesized that requiring every sequence to be recalled with 100% accuracy might bar participants from reaching longer spans where information was more likely to overload PM and to be retrieved from SM. To test this hypothesis, Gibson et al. conducted a Cogmed training study with a reduced RAT, thereby allowing participants to make errors and to reach longer spans over the course of training. Specifically, the reduced RAT was set so no errors were allowed on spans of length two or three; one error was permitted on spans of length four, five, or six; two errors were allowed on spans of seven, eight, or nine, and so on. Based on this algorithm, the RAT for four items was 75% (3/4), for five items it was 80% (4/5), and for six items it was 83% (5/6) whereas for higher spans, the RAT ranged from 71% to 69%. Thus, although the overall reduced RAT was around 80%, the proportion correct required was span-contingent.

Applying the reduced RAT to two training conditions: standard Cogmed (i.e., simple span WMT) and an experimental complex span version of Cogmed, Gibson et al. (2013) found that all participants significantly improved their SM (and PM) capacity, regardless of exercise type. Therefore, despite complex spans being more distracting and more likely to interfere with the contents of PM, improvements in SM capacity could only be attributed to the reduced RAT. In addition, the findings of Gibson et al. (2013) were supported by another study that found improvements in SM capacity with a reduced RAT (87.5%), irrespective of whether participants trained on simple or complex spans (Harrison et al., 2013). Thus, early evidence indicated that WMT with a reduced RAT might be a viable method for targeting SM capacity.

Current Randomized Controlled Trial

Although the findings of Gibson et al. (2013) were encouraging, the study was limited by a small, adult sample, use of a passive control group, and a lack of far transfer measures. Further, the only other study evaluating WMT effects on SM capacity also focused solely on adults

(Harrison et al., 2013). Accordingly, the present study was designed to examine SM enhancement in a more rigorous and detailed fashion.

With respect to rigor, the present study compared the impact of three WMT conditions in a large heterogeneous sample of adolescents, using a double-blinded, placebo-controlled, randomized design. The three training conditions were designated as the “two-component,” the “adaptive one-component,” and the “non-adaptive one-component” conditions. Two-component WMT was designed to target both PM and SM components of WM, whereas adaptive one-component WMT was designed to target solely PM and non-adaptive one-component WMT was designed to weakly target PM.

The three training conditions were identical albeit with one critical change between them. For instance, the two-component condition was identical to the adaptive one-component condition with the sole exception being that a reduced RAT was implemented in the former condition whereas a 100% RAT was implemented in the latter condition. Likewise, the adaptive one-component condition was identical to the non-adaptive one-component condition with the sole exception being that span length was determined adaptively in the former condition whereas it was fixed at two items in the latter condition.

The adaptive and non-adaptive one-component conditions were designed to serve as control conditions for the two-component condition because neither was expected to target the SM component of WM. Use of the adaptive one-component condition as a control condition was especially important because it should have induced the same placebo-based benefits as the two-component condition (Boot, Simons, Stothart, & Stutts, 2013; Simons et al., 2016). Given that non-adaptive control conditions have been criticized for possibly not inducing placebo-based benefits (Boot et al., 2013; Simons et al., 2016), the non-adaptive one-component condition was included to assess the potential presence of these benefits in the adaptive one-component condition. Beyond inclusion of two active controls, further strengths of the current trial included: blinded participants, parents, assessors, and coaches, ratings of participant motivation, as well as far transfer gF and academic measures at post-training and long term follow-up.

With respect to detail, the present study provided a closer examination of SM enhancement by focusing on the specific mechanisms underlying retrieval from SM. Indeed, once information is lost from PM, cue-dependent/strategic search and retrieval from SM is determined by three stages: delimiting the search set, recovery, and error monitoring (Unsworth & Engle, 2007a, 2007b; Wixted & Rohrer, 1994). First, individuals must use internal and/or external retrieval cues to delimit the amount of information they search through in SM (i.e., the search set size). Both target items and irrelevant items are included in the search set. After the search set is delimited, participants sample item representations, which are assumed to be of equal strength, randomly with replacement. In the second phase of retrieval, the recovery phase, items that have been sampled are recovered into consciousness if they exceed some absolute threshold. Finally, a decision making/ monitoring process is used to determine whether the recovered item is a target and should be recalled or an irrelevant item that should be forgotten (Unsworth & Engle, 2007a).

Measuring Retrieval from SM

Researchers have operationalized the three stages of retrieval from SM with three corresponding measures: search set size is indexed by recall latency, recovery is measured by total recall (total correct + total errors) or as mentioned below by total correct alone, and error monitoring is measured by intrusions (total errors). Using primarily delayed free recall (DFR) tasks (Unsworth, 2007, 2009), researchers have been able to reliably distinguish between high and low WM capacity individuals on these three measures of retrieval from SM. The current study thus focused exclusively on verbal and spatial DFR tasks because they allowed us to isolate the effect of WMT on the SM retrieval parameters.

In studies comparing the retrieval profile of low versus high WM capacity adults, it has been shown that high WM capacity individuals use better cues to delimit the search set, consequently leading to smaller search set sizes including a smaller proportion of intrusions. High WM capacity individuals therefore spend less time sampling items in smaller search sets resulting in shorter recall latencies. Furthermore, high WM capacity individuals tend to have better error monitoring at the decision making phase than lower WM capacity individuals resulting in fewer errors (Unsworth, 2007, 2009; Unsworth & Engle, 2007a).

Despite committing fewer errors, higher WM individuals also have higher total recall scores (which includes both correct and incorrect responses) than lower WM individuals, indicating that the difference in total recall is attributed to superior total correct scores (Unsworth, 2007). Thus, although total recall serves as a measure of the recovery stage of retrieval, it is also vital to measure the total number of items that were correctly recalled in order to disambiguate the nature of improvement in recovery. In addition, this total correct measure is especially important in estimating the correlations between the three stages of retrieval as the correlation between total recall and total errors would otherwise be artificially inflated due to the fact that both include the number of incorrect responses. For these reasons, total correct was used in place of total recall as a measure of recovery in the present study.

Predictions

Generally, we expected that WMT targeting SM capacity would result in participants resembling high WM capacity adults (Unsworth, 2007, 2009; Unsworth & Engle, 2007a). That is, two-component trainees would have shorter recall latencies, higher total correct, and fewer total errors. In order to inform our predictions regarding how the expected enhancement of SM might influence far transfer, bivariate correlations between pre-test scores on the three stages of retrieval from SM, gF, and academic measures were inspected for the current sample. Improvements in the stages of retrieval from SM that were significantly correlated with gF and academic measures were expected to be more important for far transfer than improvements in those stages that were not so correlated.

Method

Participants

Participants were a heterogeneous sample of 190 adolescents, aged 10 to 13 years, recruited from three middle schools (grades 6 – 8) in a school district serving 2500 students. An a priori power analysis indicated that 41 participants were needed per condition (N = 123) to have an 80% chance of detecting a medium effect size, but we recruited the additional adolescents to account for possible attrition. Letters describing the researcher’s interest in studying the impact of cognitive training on working memory, reasoning, and academic achievement were mailed to the caregivers of all adolescents in the district by their schools. The letter informed caregivers that adolescents would be assessed once before and twice after taking part in a 25-day regimen of WMT. Families were also informed that adolescents would be randomly assigned to one of three WMT regimens and that their group assignment would be revealed only after the completion of the study. Adolescents were informed that they would receive $40.00 for each assessment appointment and $4.00 for completing each day of the assigned 25-day training regimen ($100.00 total).

Two cohorts of adolescents participated in the study. Cohort 1 was assessed and trained between August 2015 and April 2016 and cohort 2 between August 2016 and April 2017. In order to be enrolled, adolescents needed to meet the following inclusion criteria: 1) access to a computer with stable internet connection, 2) no previous exposure to WMT interventions, 3) no parent-reported motor problems that might interfere with use of a computer mouse, and 4) no ongoing participation in another research study. Prior to scheduling the pre-test appointment, 16 participants withdrew from the study. The remaining 174 participants were scheduled for pre-test and randomly assigned to one of three WMT conditions (see Figure 1). Preliminary examination of training results after cohort 1 revealed that some participants in the two-component WMT group were not experiencing the experimental manipulation intended to train SM. Thus, in cohort 2, we oversampled slightly for the two-component and adaptive one-component WMT groups (see below for details). As a result, a total of 68 participants were assigned to the two-component WMT condition, 65 participants to the adaptive one-component WMT condition, and 41 to the non-adaptive one-component WMT condition.

Procedure

The first assessment appointment included a brief meeting with caregivers and adolescents to obtain consent and assent respectively, as well as to provide instructions for how to access the WMT. All participants received the same demonstration of the WMT regardless of their condition. Adolescents were informed that a researcher would call each week to monitor compliance and to troubleshoot technical issues. After the introduction meeting, adolescents were then assessed on a battery of neuropsychological and academic outcome measures. Additionally, parents completed a battery of questionnaires providing basic personal information about the student (e.g., age, grade, handedness, and race and ethnicity), a brief educational and mental health history including a list of current medications, as well as ratings of ADHD symptoms and executive functioning.

Following the pre-test, participants trained at home on the Internet during the school week and their training data was uploaded in real time to a secure database supported by Cogmed. All participants were instructed to train for 30 to 45 minutes, 5 days per week for 5 weeks (25 training days total). Each week the principal investigator reviewed participant training data to assess compliance status. Then, graduate students trained as Cogmed coaches, who were blind to participant condition, called participants once per week.

To keep coaches blind, they did not have access to participant training data. Also, they used scripts during calls that did not elicit feedback regarding the adaptivity level of the participant's training. For instance, coaches never asked about the difficulty of tasks or how many items were presented. In order to ensure that participants were themselves completing the training rather than other family members or friends, coaches also asked participants to reveal if any new exercises appeared during the week's training. Post-training assessments were never conducted by a researcher who had also served as that participant's coach, to avoid any possibility that assessment could be biased by the coach having had repeated contact with the participant and discussing their training.

Following training, participants returned for post-test on same battery of tasks as the pre-test appointment. Participants returned again for a six-month follow-up on the same assessment battery as pre- and post-test.

Parent Ratings of Executive Functions

Primary caregivers completed two rating-based measures of their child's executive functions prior to WMT. First, ADHD symptoms were measured using the DuPaul ADHD rating scale (DuPaul, Power, Anastopoulos, & Reid, 1998). The DuPaul scale requires parents to rate the frequency of 18 inattentive and hyperactive/impulsive symptoms from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; American Psychiatric Association, 2000). Second, executive functions were measured using the 86-item Behavior Rating Inventory of Executive Functioning (BRIEF; Gioia, Isquith, Guy, & Kentworthy, 2000). Of particular interest was the 10-item subscale of the BRIEF that measures WM problems. Both measures have demonstrated excellent reliability with, α = .92 and α = .80 to .98 for the DuPaul scale and BRIEF respectively. These measures were collected for descriptive purposes and were not intended to serve as covariates or outcomes in the main analyses.

Outcome Measures

Participants completed six neuropsychological measures: one verbal and one spatial task of each immediate free recall (IFR), DFR, and gF, as well as standardized measures of math and reading. As the current report is focused on training-related enhancement of SM capacity, data from the IFR tasks, which measured PM capacity, will not be discussed further. For all neuropsychological measures, parallel forms were administered. At pre-test, half of the participants in each cohort were given form A and the other half were given form B. The alternate form was administered at both the post-training and the six-month follow-up assessments. Participants also completed a measure of sight word reading at pre-test and were asked about their motivation during training at post-test.

SM Capacity Measures

Following prior research (Unsworth, 2007, 2009; Unsworth, Brewer, & Spillers, 2009), SM capacity was measured using two DFR tasks. Unlike IFR tasks, DFR tasks include a distractor task after the presentation of to-be-remembered items and are therefore understood to overload the contents of PM and require recall solely from SM. The verbal DFR task included six lists of 10 unique high frequency words presented in random order. All participants saw the same word lists presented on a standard CRT monitor in 20-point white font on a black background. Each word was presented in the middle of the computer screen for one second. After presentation of the entire word list, a fixation-cross appeared to signify the start of the delay period.

During the delay, participants saw a series of three digit numbers that they needed to reorder aloud from smallest to largest. For example, if participants saw 597 they needed to say "5", "7","9". A total of 8 three-digit numbers, one every two seconds, appeared during the delay. Following the delay, question marks appeared in the center of the screen signifying the recall period. Participants then had 30 seconds to recall aloud as many words as possible in any order and they needed to wait for the recall period, as signified by a brief tone, to end before moving to the next trial. Prior to the test trials, two practice trials using letter stimuli were presented. Participant responses were recorded via a microphone and digital recorder and scored as correct if they matched a word on the list, if they uttered a plural version of a list item that was singular, or if they said a past tense version of a list item that was in the present tense.

For the spatial DFR task, participants were shown six lists of 10 different spatial locations marked by white squares on a black background presented on a standard CRT monitor. Each of the 10 locations per list was cued by temporarily changing the color from white to red for one second. Squares appeared at any one of 60 (6 x 10) unique locations on the computer screen and each location was cued only once across the six different lists. To make the task more manageable, 30 of the possible 60 squares were randomly selected and appeared in three consecutive trials. After the first three trials, 30 new locations were randomly selected from the possible 60 without replacement and appeared in the last three trials.

After all of the locations were cued in a list, the computer screen went blank to signify the start of the delay period. During the delay, participants completed a puzzle by placing a set of Styrofoam shapes on a paper template with outlines for each puzzle piece. Because participants were focused on placing the puzzle pieces in their appropriate locations, the test administrator informed the participant when the delay period ended (i.e., when the spatial locations re-appeared for recall). Two practice trials that were identical to the test trials were presented before the six test trials. Participants were required to click as many of the cued locations as possible during the 30-second recall period in any order. Participant responses were recorded by the computer and include the location of the mouse clicks, the order of the mouse clicks, and the number of correct responses.

For each verbal and spatial DFR task, three measures were obtained following Unsworth (2007, 2009). First, a measure of search set size was estimated using recall latency (i.e., the average time during the recall period when a response was emitted, including both correct and incorrect items). Second, as a measure of recovery, total correct was the number of items correctly recalled summed over the six trials. Lastly, as measure of error monitoring, total errors was the number of previous list intrusions, extra-list intrusions, and repetitions summed over the six trials. Recall latency, total correct, and total errors scores were then each averaged over the verbal and spatial modalities.

Fluid Reasoning Measures

Participants were assessed with one verbal and one spatial gF test. Verbal reasoning was assessed with the Educational Testing Service (ETS) Inference test (Ekstrom, French, Harman,

& Dermen, 1976). The ETS Inference test required participants to select one of five conclusions that could be drawn from a given statement. Each test consisted of 10 statements and participants were given six minutes to provide their answers. Participants' scores were the total number of correct conclusions reported, with 10 as the maximum score.

Spatial reasoning was assessed with a Bochum Matrices-Advanced-Short Version (BOMAT; Hossiep, Turck, & Hasella, 1999). Participants were given 10 minutes to view as many of 29 matrices as possible, each with one empty square in the 5 x 3 matrix. The time limit was imposed to avoid ceiling effects (Moody, 2009). Participants were required to decide which of six options completed each matrix. The number of correct responses was considered a measure of spatial reasoning. Both gF measures were chosen because they have been used in previous studies that investigated the relation between WM capacity and fluid IQ (e.g., Kane et al., 2004), or the effects of WMT on fluid IQ (Chein & Morrison, 2010; Jaeggi et al., 2008; Klingberg et al., 2005). Each of these measures has demonstrated considerable reliability, with $\alpha = .76$ to .78 for the ETS Inference test (Ekstrom et al., 1976) and $\alpha = .58$ for the BOMAT (Hossiep et al., 1999; Jaeggi et al., 2010).

Academic Growth Measures

Participant reading and math performance was assessed with the Measures of Academic Progress (MAP; Northwest Evaluation Association, 2011a, 2011b). The MAP is a computerized, adaptive assessment that can be administered up to three times during the school year in a group setting and is intended to measure academic growth over time. The reading and math sections consist of 40 and 53 multiple choice questions respectively. For reading there is a Reading Total score and three sub-scores for Literature Comprehension, Non-Fiction Comprehension, and Vocabulary. For math, there is a Math Total score and five sub-scores for Number Sense, Computation, Algebra, Geometry, and Data Analysis. MAP scores fall on an equal interval scale known as the Rasch Unit (RIT) scale and these scores have demonstrated considerable reliability and validity (r = .77 to .90; Northwest Evaluation Association, 2011a).

Sight Word Reading Measure

The Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999) was administered to participants at the pre-test assessment. The TOWRE is comprised of sight word efficiency and phonemic decoding efficiency subtests which have each demonstrated considerable reliability (α = .93 to .96; Torgesen et al., 1999). In the current study, solely the sight word efficiency subtest, which required participants to read aloud as many words from a list as possible in 30 seconds, was administered examine the extent to which performance on the verbal WM tasks might be influenced by reading ability.

Participants who scored at or below the 30^{th} percentile were considered “at risk for reading difficulty” (Torgesen et al., 1999). Of the sample included in the final analysis (see below), 17 participants were identified as at risk. In order to determine the extent to which reading difficulty might influence the potential benefits of WMT, the final analysis was conducted with and without at risk readers and the results were unchanged (see also, Gibson, Gondoli, Ralph, & Sztybel, in press). Thus, at risk readers were included in the final analysis.

Motivation

After training, participants completed the Intrinsic Motivation Inventory (IMI; McAuley, Duncan, & Tammen, 1989; Ryan & Deci, 2000). The IMI is a 12-item scale with two subscales: Interest/Enjoyment (7-items) and Effort/Importance (5-items). Each item is rated from 1 (*not true at all*) to 7 (*very true*), with greater motivation indicated by higher scores. The IMI has demonstrated excellent reliability (α = .91).

WMT Interventions

Participants were randomly assigned to two-component, adaptive one-component, or non-adaptive one-component WMT. The two-component training was considered to be the "treatment", whereas adaptive and non-adaptive one-component trainings were active control groups. All three training conditions included the same bank of 12 verbal and spatial simple WM span tasks. Verbal exercises involved remembering the forward serial order of digits or letters and repeating them back using the computer mouse. Spatial exercises involved remembering the forward serial order of stationary or moving locations presented in two-or-three dimensional arrangements (see Appendix for exercise details). In order to equate the amount and type of feedback received across the groups, features of the standard Cogmed program such as verbal and textual feedback, a level meter showing how many items needed to be recalled, and the high score list showing the participant's highest level of performance on each exercise were removed.

Out of the 12 possible exercises, eight were labeled as "common" exercises and participants trained on four of these common exercises, two verbal and two spatial, each day. Although the eight common exercises rotated throughout the duration of training, participants in all three conditions were exposed to the same four common exercises each training day and thus, it was possible to make a direct comparison across conditions. The remaining four exercises, two visual and two verbal, were designated as "critical" exercises and were presented on every training day in all three conditions. Thus, participants trained on the same four common and the same four critical exercises each training day, with only common exercises rotating.

Following Gibson et al. (2013), two-component WMT was adaptive and had a reduced RAT on the critical exercises and a 100% RAT on the common exercises. All participants in two-component WMT began training by recalling two to-be-remembered items. Because spans of two and three items tend to be easy, 100% accuracy was required at those levels for the critical exercises but thereafter, participants in two-component WMT were able to make one error on spans of four, five, and six items; two errors on spans of seven, eight, and nine items; three errors on spans of 10, 11, and 12 items, and so on.

Just like two-component WMT, adaptive one-component WMT also started at a span of two items and adapted in difficulty. However, one-component WMT had a 100% RAT on both critical and common exercises. Every exercise in adaptive one-component and two-component training required participants to complete 15 trials, each with a unique arrangement of to-be-remembered items, for a total of 120 trials each training day. Thus, solely the RAT on the critical exercises (i.e., the number of correct items required for the exercise to increase in difficulty) was the parameter that differed between the two- and adaptive one-component training.

Non-adaptive one-component WMT was identical in appearance to the other conditions except the number of to-be-remembered items was fixed at two and never increased in difficulty. Thus, participants needed to remember the forward serial order of two items on every trial on every exercise, both critical and common, for the entire duration of training. Note that non-adaptive one-component WMT had 20 trials total per exercise (160 trials per training day) in an attempt to address previous criticisms that non-adaptive training tends to be shorter than adaptive training (Chacko et al., 2014; van Dongen-Boomsa et al., 2014).

Results

Participants

The flow of participants through the study is shown in Figure 1. Out of the 174 participants randomly assigned to the three WMT groups, a total of 39 participants were excluded prior to the training performance analysis, with some meeting more than one exclusion criteria. Twenty-four participants dropped out of the study between pre-test and follow-up. Additionally, eight participants had scores +/- three standard deviations from the mean on the WM measures at pre-test. Of these eight, two were already excluded as dropouts and the additional six were excluded as outliers. Seventeen participants failed to complete at least 20 training days. Of these seventeen, eight were already excluded as dropouts or outliers and the additional nine were excluded for training less than 20 days. The remaining 135 participants were included in the subsequent statistical analysis (two-component WMT, n = 52; adaptive one-component WMT, n = 48; non-adaptive one-component WMT, n = 35).

Correlation Analysis

Bivariate correlations at pre-test between the three stages of SM retrieval (recall latency, total correct, total errors), gF, and academic growth measures were inspected for the entire original sample minus the eight outliers and the six participants that withdrew at pre-test (N = 160) (Table 1). It should be noted that only Reading Total and Math Total were included in Table 1, as sub-scores for reading and math were of the same magnitude and direction as their respective total scores.

Concerning the pattern of correlations between the three SM retrieval mechanisms, we expected pre-test correlations for our sample to mirror the adult literature. That is, we anticipated recall latency to negatively relate to total correct and positively relate to total errors, as the time it takes to sample items in the search set is shorter when the proportion of targets to non-targets is larger. In addition, we also expected total correct and total errors to strongly and negatively relate (Unsworth, 2007, 2009).

Consistent with these expectations, there was a significant, positive correlation between recall latency and total errors in the current study (see Table 1). However, the remaining two correlations deviated from these expectations: there was an unexpected small and non-significant, positive correlation between recall latency and total correct; and, the negative correlation between total correct and total errors was also smaller than anticipated (Unsworth, 2009).

Regarding correlations between the SM retrieval mechanisms and gF, the adult literature has shown that gF is negatively related to recall latency, positively related to total correct, and negatively related to total errors (Unsworth, 2009). In other words, the higher the gF, the shorter the time participants spend sampling the search set, the greater the number of target items recalled, and the lower the number of non-targets recalled. However, in the current study, only total correct significantly related to gF. There was no significant correlation between recall latency or total errors and gF. Interestingly, these correlations suggested that the recovery of items, rather than search set size or error monitoring, drove the relationship between SM capacity and gF in this sample.

Inspection of correlations between SM retrieval mechanisms and academic growth measures revealed that both reading and math were significantly positively related to total correct and significantly negatively related to total errors but unrelated to recall latency.
Thus, if two-component WMT successfully enhanced recovery from SM, and/or enhanced error monitoring, then, far transfer to reading and math performance might be possible. Taken together, the current sample did not follow the same pattern of relationships observed between SM capacity measures and gF in the adult literature (Unsworth, 2009). Instead, the significant positive correlation between total correct and gF suggests that improving recovery of items,

whether or not total errors or recall latency are decreased, should lead to far transfer following WMT. Similarly, the significant positive correlations between total correct and the academic measures, as well as the significant negative correlations between total errors and the academic measures suggests that improving recovery of items, as well as monitoring should lead to far transfer following WMT.

Training Performance Analysis

Prior to analysis of outcome measures, it needed to be verified that two-component trainees reached spans of four items or more on the critical exercises, where the span-contingent RAT manipulation came into effect. At spans below four items, participants in two-component training were actually experiencing the 100% RAT and would not be expected to differ from the adaptive one-component control group on critical exercises. Indeed, if the lower RAT was successful, then two-component trainees should have reached higher average span lengths than the adaptive one-component trainees on each of the four critical exercises. Regarding the common exercises, two-component trainees should have reached at least the same average span length as the adaptive one-component group, if not better, given their potentially more potent training regimen. Non-adaptive one-component trainees on the other hand were barred from increasing span length and thus, were not expected to reach comparable spans to the adaptive trainees.

To evaluate training performance for the two- and adaptive one-component WMT groups, we took the average span achieved for the two verbal critical exercises and the average span achieved for the two spatial critical exercises for each training day from 11 to 20, and we evaluated whether these average spans were greater or less than four items. Based on the separate verbal and spatial average spans, two main profiles emerged. Profile 1 characterized participants who spent 75% or more of training on spans of four or more items across the final 10 required days of training on each of the verbal and spatial critical exercises. Profile 2 characterized participants who on average spent less than 75% of the training on spans four or more items across the final 10 required days of training on each of the verbal and spatial critical exercises. Figure 2 presents the training curves averaged over modality for participants in two- and adaptive one-component groups as a function of these two profiles.

For each Profile 1 and Profile 2, repeated measures ANOVAs on average span achieved on each common and critical exercises were evaluated with training duration (days 1 to 20) as the within subjects factor and training condition (two-component vs. adaptive one-component WMT) as the between subjects factor (see Table 2). Mauchly's test indicated that the sphericity assumption had been violated, so degrees of freedom were corrected using Greenhouse-Geisser estimates.

Regarding Profile 1 participants, there was a significant main effect of training duration and condition on the critical exercises. Thus, the average span length achieved by Profile 1 participants on critical tasks significantly increased over the course of training and significantly differed between two- and adaptive one-component trainees. Although there was a significant interaction between training duration and condition, this interaction can be explained by participants in two- and adaptive one-component WMT starting training at 2-item spans and progressively differing over the course of training. As seen in Figure 2, participants in the two-component WMT group who trained at spans of four items or more for at least 75% of training outperformed adaptive one-component Profile 1 participants by about training day three. Thus, as intended, Profile 1 two-component participants were able to experience the lower RAT and reach higher spans than the adaptive one-component WMT group on the critical exercises.

Concerning the common exercises, there was a significant main effect of training duration. Profile 1 participants in two- and adaptive one-component WMT both significantly increased common exercise span length over the course of training. There was neither a significant main effect of condition nor interaction between training duration and condition on the common exercises. Thus, Profile 1 participants in the two- and adaptive one-component WMT groups did not achieve significantly different span lengths on the common exercises over the course of training, suggesting that any benefit associated with exposure to the lower RAT on critical exercises did not transfer to performance on the common exercises.

Regarding Profile 2 participants, there was a significant main effect of training duration on both critical and common exercises but no significant main effect of condition or interaction between training duration and condition for either exercise type. Profile 2 participants in both two-component and adaptive one-component WMT improved on the early training days and then decreased or stayed at the same average span level for the remainder of training on both critical and common exercises. Thus, contrary to the design of the two-component WMT, Profile 2 two-component trainees did not differ from Profile 2 adaptive one-component trainees on the critical exercises. Given that Profile 2 participants did not engage in training as intended, these participants were excluded from the final analysis. Although Profile 2 adaptive one-component trainees did indeed experience the 100% RAT, for consistency, all Profile 2 participants were excluded rather than some subset of the group.

Final Analysis

The final analysis included a total of 101 participants with 33 two-component trainees, 33 adaptive one-component trainees, and 35 non-adaptive one-component trainees. Importantly, despite attempts to account for attrition, the sample size per group approximated, but did not quite achieve the 41 participants that were targeted in a priori power analysis. The first analysis examined possible pre-test differences between the training groups. Inspection of sample characteristics revealed that none of the WMT groups differed significantly from one another on continuous or categorical demographic variables (see Table 3). It is important to note that diagnostic statistics detailed in Table 3 were parent-reported, requiring either a yes or no answer, and therefore have not been corroborated with clinical interviews.

Consistent with the lack of differences on demographic variables, the three WMT groups also did not differ significantly on pre-test parent-rated WM abilities or ADHD symptoms, pre-test sight word reading, the number of training days competed, or self-reported motivation during training (see Table 4). Pre-test, post-test, and follow-up means and standard deviations on the outcome measures for the three WMT groups are reported in Table 5. Importantly, three groups did not differ significantly on SM capacity, gF, or academic growth measures at pre-test (Table 6).

Next, mixed ANCOVAs were conducted on each of the outcome measures with time (post-test and follow-up) as the within subjects factor, condition (two-component vs. adaptive one-component vs. non-adaptive one-component WMT) as the between-subjects factor, and corresponding pre-test scores as covariates. As can be seen in Table 7, there were no significant differences between the three training groups on recall latency (p = .80) or total correct (p = .94). There was however a significant main effect of condition on total errors (p = .020).

Exploratory Least Significant Difference post-hoc analysis revealed that the two-component WMT group had significantly more errors over the adaptive one-component (p =. 015) and non-adaptive one-component WMT groups (p = .014). The control groups did not differ from one another on total errors. Thus, the results of the ANCOVAs suggested that two-component trainees became more error prone (i.e., increased total errors) but did not reduce search set size (i.e., improve recall latency) or increase recovery (i.e., improve total correct) significantly better than the control WMT conditions.

Based on the lack of improvement in total correct, the only SM retrieval mechanism significantly related to both gF and academic performance measures, far transfer effects were not expected. Indeed, there were no significant differences between the WMT groups on gF (p = .89) or the majority of the academic performance measures (see Table 7). There were however significant interactions between time and condition on the Reading Total score (p = .020) and Literature Comprehension score (p = .001).

Review of simple main effects associated with these significant interactions showed that participants in non-adaptive one-component WMT had significantly greater Reading Total scores than both two-component (p = .044) and adaptive one-component (p = .010) groups at post-test but not follow-up. Furthermore, the non-adaptive one-component group also had significantly greater Literature Comprehension scores compared to the two-component (p = .036) and adaptive one-component WMT groups (p = .047) at post-test but not follow-up. Inspection of means revealed however that the significant difference between the non-adaptive one-component group and the two other groups was driven by an in increase in both reading scores for the non-adaptive group coupled with a decrease in both reading scores for both adaptive training groups from pre- to post-test.

Discussion

The current double-blinded, placebo-controlled, randomized trial was inspired by the need to develop theoretically motivated WMT interventions that target both PM and SM components of WM. Until now, our group has shown that the commercially available simple span training Cogmed targets solely PM capacity (Gibson et al., 2011; Gibson et al., 2012) but, may target SM capacity with a reduced RAT (Gibson et al., 2013). Thus, we investigated the impact of a novel two-component WMT version of Cogmed aimed at enhancing SM capacity, gF, and academic growth.

Participants were randomly assigned to either two-component WMT with a reduced RAT, adaptive one-component WMT with 100% RAT, or non-adaptive one-component WMT with a fixed span length. Importantly the span-contingent reduced RAT took effect at span lengths of four or more items but not at spans of two or three items. By examining two-component trainee performance, we confirmed that participants who trained at spans of four or more items 75% or more of the time across the last 10 required training days did train as expected. In contrast, we also confirmed that the two-component participants who trained at spans of four or more items less than 75% of the time across the last 10 required days of training did not train as expected. Thus, we excluded these low-span participants in the two-component training condition. For consistency, we also excluded the corresponding low-span participants in the adaptive one-component training condition.

Near Transfer

Comparing the appropriate two-component, adaptive one-component, and non-adaptive one-component WMT groups revealed that two-component trainees significantly differed from controls on error monitoring, but not on the size of the search set or the recovery of correct items. Moreover, the main influence of the reduced RAT led to worse error monitoring. These results suggested that two-component WMT with the reduced RAT did indeed target an SM retrieval mechanism, although not to the benefit of the two-component group.

The increase in total errors, although unexpected, does correspond with design of the reduced RAT, which allowed participants to make more errors and to reach longer span lengths where retrieval from SM would be required. It is thus possible that two-component trainees learned to use a lower threshold for reporting errors without changing the contents of the search set or the recovery from the search set. Thus, it seems that other modifications to WMT algorithms will be needed to target search set size and recovery parameters of retrieval from SM.

Furthermore, enhancement of SM in the two-component WMT group did not follow the pattern typically seen in high WM capacity adults (Unsworth, 2009). Indeed, previous research has indicated that high WM capacity individuals by virtue of their using better cues, have smaller search set sizes containing more targets and fewer intrusions than low WM capacity individuals. Typically high WM capacity participants have shorter recall latencies, and they recall more correct items. Accordingly, search set size and recovery have been identified as the more crucial parameters for distinguishing between high and low WM capacity (Unsworth, 2009). However, the pattern of correlations observed at pre-test in the present study between the three stages of retrieval were discrepant from the pattern typically observed in the adult literature (Unsworth, 2009). This finding suggests that adolescent retrieval mechanisms may not yet be fully formed. Thus, further validation of the SM retrieval mechanisms in adolescent populations may be required before it is known whether these mechanisms are appropriate targets for child-focused WMT.

Far Transfer

Regarding far transfer, our inspection of pre-test correlations indicated that total correct and total errors were the only parameters significantly related to higher-order cognitive abilities. Based on the correlation between total correct and gF, one possibility was that better recovery was of greater importance than the size of the search set and error monitoring. However, two-component trainees increased total errors rather than total correct. Thus, the retrieval mechanism by which far transfer to gF might have occurred namely, near transfer to total correct, was not improved by two-component training above that of the control WMT programs.

Inspection of pre-test correlations also indicated that enhanced total correct and/or decreased total errors would transfer to improved reading and math scores. However, as total correct was not enhanced and total errors did not decrease for any groups, the pattern of results defied expectation. Indeed, only the non-adaptive one-component WMT group significantly improved over both two- and adaptive one-component WMT groups on Reading Total and Literature Comprehension from pre- to post-test. However, as discussed in prior WMT reviews (Melby-Lervåg et al., 2016; Redick, 2015), the increased scores for non-adaptive participants coupled with decreased scores in both the adaptive training groups cautions interpretation of this effect beyond measurement error.

Overall, the lack of significant far transfer in the two-component WMT group was not surprising given the lack of SM capacity enhancement. These results suggest that current WMT paradigms still do not adequately target the SM component and therefore caution against the conclusion that WMT does not facilitate transfer to higher-order cognition (Melby-Lervåg et al., 2016; McCabe et al., 2016).

Study Strengths

Despite the lack of significant findings, a major strength of this investigation was not only the expansion upon prior work evaluating WMT in the dual component model of WM framework (Gibson et al., 2012; Gibson et al., 2013) but also, in the care taken to address a variety of criticisms in the WMT literature (Melby-Lervåg et al., 2016; Simons et al., 2016).

Heeding the call for use of appropriate controls in training studies, we included a placebo control condition (i.e., adaptive one-component WMT) that differed from the two-component WMT group solely on the reduced RAT on critical exercises. Moreover, a non-adaptive one-component WMT control was included in the event that our expectations were defied and the adaptive one-component WMT group improved on SM capacity measures over the two-component WMT group. It should be noted that except for the non-adaptive one-component group improving Reading Total and Literature Comprehension scores over the adaptive one-component group, these groups were otherwise not significantly different on outcome measures. It is therefore possible that both conditions conferred placebo-based benefits in this trial.

Additionally, other moderators of training (see Schwaighofer et al., 2015) were also equivalent across the groups. Indeed, all participants received the same instructions for training at a scripted introduction meeting, all trained on the exact same exercises each day, and all trained five days per calendar week. Furthermore, feedback elements of the commercially available version of Cogmed such as the level meter, high score list, and verbal and text feedback were removed from the three training conditions.

To ensure comparability between the conditions, we included a measure of motivation and found participants in all three groups did not differ significantly in their levels of interest, enjoyment, effort, and importance with regards to training. These results are consistent with other studies finding no motivational differences between adaptive and non-adaptive trainees (Bergman Nutley et al., 2011; Steeger, Gondoli, Gibson, & Morrissey, 2016). Thus, improvements following training cannot be attributed to differences in exposure to the training program, features of the training program, or motivational differences in trainees.

Another aspect of Cogmed (in particular) that has raised questions is the practice of coaching participants. In this study, not only were participants, parents, and assessors blind to condition but also, coaches were blind to participant group assignment as well. Scripted coach calls to ensure compliance and training fidelity were conducted each week for all three conditions and no coaches conducting calls had access to participant training data. Furthermore, coaches were asked to guess about each participant's group assignment following training and incorrectly identified the participant's condition the majority (63%) of the time.

Although it was important to include a control for adaptive one-component WMT, inclusion of the non-adaptive one-component group was also a potential limitation of the current trial. First, the non-adaptive one-component control limited our ability to provide appropriate instructions for training to the adaptive groups. That is, instructions for training could not reveal that the goal of training was to reach higher spans. Second, in situ feedback about training performance that might have encouraged participants to reach higher spans, such as the standard Cogmed level meter, were completely removed from our versions of Cogmed. Third, we conjecture that participants, particularly with two-component Profile 2 trainees, may have reached certain span lengths, experienced their difficulty, and decided to remain at a level of comfort rather than embrace the challenge of training. Indeed this behavior may have been related to their own beliefs or mindset about what they were capable of achieving during training (Appelgren, Bengtsson, & Söderqvist, 2016). In the absence of the non-adaptive control, coaches could have made use of the training data to encourage these participants to push their capacity rather than be restricted to generic statements regarding compliance.

Conclusion

The current trial explored whether WMT with a reduced RAT could improve SM capacity. However, despite prior findings that WMT with lower RATs may target SM (Gibson et al., 2013; Harrison et al., 2013), two-component WMT in the current study did not lead to significant near or far transfer. These results suggest that other manipulations, beyond or in addition to the reduced RAT, may be needed to adequately target SM retrieval mechanisms. Furthermore, because extant WMT paradigms are not enhancing the theoretical components of WM, it remains premature to draw conclusions about the potential of WMT to improve higher order abilities. Additional research will be needed to explore how SM mechanisms operate over the course of development and how those insights might be applied to improve WMT paradigms.

Author Notes

The research reported in this article was supported by Grant# R305A150305 awarded by the Institute of Education Sciences in the Department of Education. The authors would also like to thank Daniel Schor for his help with data collection and data coding tasks.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: Author

Appelgren, A., Bengtsson, S.L., & Söderqvist, S. (2016). Incremental view on intelligence and high intrinsic motivation increase working memory training compliance. *Applied Cognitive Psychology*, *30*, 289 - 293.

Au, J., Buschkuehl, M., Duncan, G. J., & Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychonomic Bulletin & Review*, *23*(1), 331 - 337.

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*(2), 366 - 377.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612 - 637.

Bergman Nutley, S., & Söderqvist, S. (2017). How is working memory training likely to influence academic performance? Current evidence and methodological considerations. Frontiers in Psychology, *8*(69).

Bergman Nutley, S., Söderqvist, S., Bryde, S., Thorell, L. B., Humphreys, K., & Klingberg, T. (2011). Gains in fluid intelligence after training non-verbal reasoning in 4-year-old children: A controlled, randomized study. *Developmental Science*, *14*(3), 591 - 601.

Bigorra, A., Garolera, M., Guijarro, S., Hervás, A. (2016). Long term far-transfer effects of working memory training in children with ADHD: A randomized controlled trial. *European Child & Adolescent Psychiatry*, *25*(8), 853 - 867.

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 on Psychological Science*, *8*(4), 445 - 454.

Brehmer, Y., Westerberg, H., & Bäckman, L. (2012). Working-memory training in younger and older adults: Training gains, transfer, and maintenance. *Frontiers in Human Neuroscience*, *6*(63), 1-7.

Caeyenberghs, K., Metzler-Baddeley, C., Foley, S., & Jones, D. K. (2016). Dynamics of the human structural connectome underlying working memory training. *Journal of Neuroscience*, *36*(14), 4056-4066.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, *97*(3), 404 - 431.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153 - 193.

Chacko, A., Bedard, A.C., Marks, D.J., Feirsen, N., Uderman, J.Z., Chimiklis, A., …Ramon, M. (2014). A randomized clinical trial of Cogmed Working Memory Training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, *55*(3), 247- 255.

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*(2), 193 - 199.

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531-542.

Cogmed Working Memory Training. (2010). Upper Saddle River, NJ: Pearson Education Inc.

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W., ... & Zuddas, A. (2015). Cognitive training for attention-deficit/hyperactivity disorder: meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child & Adolescent Psychiatry, 54*(3), 164 - 174.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42 - 100.

Dahlin, K.I.E. (2011). Effects of working memory training on reading in children with special needs. *Reading and Writing*, *24,* 479-491.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450 - 466.

de Abreu, P.M.J.E., Conway, A. R., & Gathercole, S. E. (2010). Working memory and fluid intelligence in young children. *Intelligence*, *38*(6), 552 - 561.

Dunning, D.L., Holmes, J. & Gathercole, S.E. (2013). Does working memory training lead to generalized improvements in children with low working memory? A randomized controlled trial. *Developmental Science, 16*(6), 915 - 925.

DuPaul, G.J., Power, T.J., Anastopoulous, A.D., & Reid, R. (1998). *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. New York: Guilford.

Ekstrom R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). *Manual for Kit of Factor-referenced Cognitive Tests*. Princeton: Educational Testing Service.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309 - 331.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177 - 194.

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*, *17*(6), 546 - 563.

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*(5), 726 – 737.

Gibson, B. S., Gondoli, D. M., Ralph, K. J., & Sztybel, P. (in press). Application of the dual-component model of working memory to ADHD: Greater secondary memory deficit despite confounded cognitive differences. *Child Neuropsychology*. Retrieved from http://dx.doi.org/10.1080/09297049.2016.1225706

Gibson, B. S., Kronenberger, W. G., Gondoli, D. M., Johnson, A. C., Morrissey, R. A., & Steeger, C. M. (2012). Component analysis of simple span vs. complex span adaptive working memory exercises: A randomized, controlled trial. *Journal of Applied Research in Memory and Cognition*, *1*(3), 179 - 184.

Gioia, G.A., Isquith, P.K., Guy, S.C., & Kentworthy, L. (2000). Test review Behavior Rating Inventory of Executive Function. *Child Neuropsychology*, *6*(3), 238 - 238.

Gray, S.A., Chaban, P., Martinussen, R., Goldberg, R., Gotlieb, H., Kronitz, R., Hockenberry, M., & Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD; a randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *53*(12), 1277 – 1284.

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*(12), 2409 - 2419.

Holmes, J., Gathercole, S.E., Place, M., Dunning, D.L., Hilton, K.A., & Elliot, J.G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, *24*(6), 827-836.

Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest. BOMAT - advanced - short version*. Gottingen: Hogrefe.

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*, *105*(19), 6829 - 6833.

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*, *108*(25), 10081-10086.

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*(4), 394-412.

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & cognition*, *42*(3), 464-480.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615 - 622.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*(4), 637 - 671.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66 - 71.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189 - 217.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317 - 324.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., ... & Westerberg, H. (2005). Computerized training of working memory in children with ADHD-a randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(2), 177 - 186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781 - 791.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48 - 58.

McCabe, J. A., Redick, T. S., & Engle, R. W. (2016). Brain-training pessimism, but applied-memory optimism. *Psychological Science in the Public Interest*, *17*(3), 187 - 191.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270 - 291.

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer” evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512 - 534.

Mogle, J. A., Lovett, B. J., Stawski, R. S., & Sliwinski, M. J. (2008). What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. *Psychological Science*, *19*(11), 1071 - 1077.

Moody, D. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*, 327 - 328.

Northwest Evaluation Association. (2011a). *Technical manual: For use with Measures of Academic Progress® (MAP®) and MAP for Primary Grades*. Portland, OR: Author.

Northwest Evaluation Association. (2011b). *RIT scale norms: For use with Measures of Academic Progress® (MAP®) and MAP for Primary Grades*. Portland, OR: Author.

Partanen, P., Jansson, B., Lisspers, J., & Sundin, Ö. (2015). Metacognitive strategy training adds to the effects of working memory training in children with special educational needs. *International Journal of Psychological Studies*, *7*(3), 130 – 140.

Peijnenborgh, J. C., Hurks, P. M., Aldenkamp, A. P., Vles, J. S., & Hendriksen, J. G. (2016). Efficacy of working memory training in children and adolescents with learning disabilities: A review study and meta-analysis. *Neuropsychological Rehabilitation*, *26*(5-6), 645 - 672.

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

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*(2), 359 - 379.

Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P.J., Gathercole, S., Gold, L., … &Wake, M. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *JAMA Pediatrics*, *170*(5):e154568.

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138 - 166.

Seidler, R. D., Bernard, J. A., Buschkuehl, M., Jaeggi, S., Jonides, J., & Humfleet, J. (2010). Cognitive training as an intervention to improve driving ability in the older adult (Technical Report No. M- CASTL 2010-01). Ann Arbor: University of Michigan.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628 - 634.

Simons, D.J., Boot, W.R., Charness, N., Gathercole, S.E., Chabris, C.F., Hambrick, D.Z., & Stine-Morrow, E.A.L. (2016). Do “Brain-Training” programs work? *Psychological Science in the Public Interest*, *17*(3), 103 - 186.

Spencer-Smith, M., & Klingberg, T. (2015). Benefits of a working memory training program for inattention in daily life: A systematic review and meta-analysis. *PLoS One*, *10*(3), e0119522.

Steeger, C. M., Gondoli, D. M., Gibson, B. S., & Morrissey, R. A. (2016). Combined cognitive and parent training interventions for adolescents with ADHD and their mothers: A randomized controlled trial. *Child Neuropsychology*, *22*(4), 394 - 419.

Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., ... & Gabrieli, J. D. (2013). Failure of working memory training to enhance cognition or intelligence. *PloS one*, *8*(5), e63614.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-ed.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127 - 154.

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 1020 - 1034.

Unsworth, N. (2009). Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. *Memory & Cognition*, *37*(6), 837 - 849.

Unsworth, N. (2016). The many facets of individual differences in working memory capacity. *Psychology of Learning and Motivation*, *65*, 1- 46.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Inter- and intra-individual variation in immediate free recall: An examination of serial position functions and recall initiation strategies. *Memory*, *19*(1), 67 - 82.

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*(1), 68 - 80.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104 - 132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038 - 1066.

Unsworth, N., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology*, *71*, 1 - 26.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392 - 406.

van der Donk, M., Hiemstra-Beernink, A., Tjeenk-Kalff, A., van der Leij, A., & Lindauer, R. (2015). Cognitive training for children with ADHD: A randomized controlled trial of cogmed working memory training and 'paying attention in class'. *Frontiers in Psychology*, *6*, 1081.

van Dongen-Boomsma, M., Vollebregt, M.A., Buitelaar, J.K., & Slaats-Willemse, D. (2014).Working memory training in young children with ADHD: A randomized placebo-controlled trial. *Journal of Child Psychology and Psychiatry*, *55*(8), 886-896.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin and Review*, *1*(1), 89 - 106.

Table 1

Bivariate correlations between pre-test SM capacity, gF, and academic growth scores

	1	2	3	4	5	6
1. Recall Latency	-					
2. Total Correct	.08	-				
3. Total Errors	.41**	-.12	-			
4. gF	-.06	.28**	-.09	-		
5. Reading Total	-.09	.38**	-.33**	.48**	-	
6. Math Total	-.08	.39**	-.35**	.49**	.76**	-

Note. ** $p < .01$; MAP reading sub-scores: Literature Comprehension, Non-fiction Comprehension, and Vocabulary, as well as MAP math sub-scores: Number Sense, Computation, Algebra, Geometry, and Data Analysis have been omitted for simplicity. The magnitude and direction of both the reading and math sub-scores approximate their respective total scores.

Table 2

Repeated Measures ANOVA comparing participants in two-component and adaptive one component WMT groups in each training profile on common and critical training tasks

		Repeated Measures ANOVA			Greenhouse-Geisser		
		F	*p*	η_p^2	χ^2	*p*	ε
Profile 1 (≥75%)							
Critical							
	Duration	$F(6.75, 432) = 102$	.001	.62	$\chi^2(189) = 768$	.001	.36
	Condition	$F(1, 64) = 5154$	.001	.99			
	Duration x Condition	$F(6.75, 432) = 2.62$	.013	.039			
Common							
	Duration	$F(9.50, 608) = 106$	.001	.63	$\chi^2(189) = 520$	.001	.50
	Condition	$F(1, 64) = .064$	.80	.001			
	Duration x Condition	$F(9.50, 608) = .82$	.61	.013			
Profile 2 (< 75%)							
Critical							
	Duration	$F(3.68, 118) = 10.3$	.001	.24	$\chi^2(189) = 574$	.001	.19
	Condition	$F(1, 32) = .16$	.69	.005			
	Duration x Condition	$F(3.68, 118) = .80$	.52	.024			
Common							
	Duration	$F(8.54, 184) = 10.6$	.001	.25	$\chi^2(189) = 456$	.001	.30
	Condition	$F(1, 32) = .72$	.40	.022			
	Duration x Condition	$F(8.54, 184) = 1.08$	.38	.033			

Note. Profile 1 = participants who on average spent 75% or more of training (days 1 to 20) at spans of four or more (two-component WMT, n = 33; adaptive one-component WMT, n = 33); Profile 2 = participants who on average spent less than 75% of training (days 1 to 20) at spans of four or more (two-component WMT, n = 19; adaptive one-component participants, n = 15).

Table 3

Participant descriptive statistics for two-component, adaptive one-component, and non-adaptive one-component WMT groups included in the final analysis

	2C (n = 33)	**$1C_A$ (n = 33)**	**$1C_{NA}$ (n = 35)**	**ANOVA**		
	M (SE)	M (SE)	M (SE)	*F*	*p*	η_p^2
Demographic						
Age	12.5 (.14)	12.5 (.14)	12.2 (.13)	$F(2, 98) = 1.33$	.27	.026
Grade	7.46 (.14)	7.46 (.14)	7.20 (.13)	$F(2, 98) = 1.21$	.30	.024
SES (USD)	91,578 (11, 084)	123,615 (11, 261)	120,101 (10, 753)	F(2, 94) = 2.51	.086	.051
				Chi-squared Test		
				χ^2	*p*	
Gender (M)	45.5%	36.4%	48.6%	$\chi^2(2, N = 101) = 1.10$	.58	
Handedness (R)	81.8%	87.9%	91.4%	$\chi^2(2, N = 101) = 1.42$	.49	
Race/ Ethnicity				$\chi^2(8, N = 101) = 13.2$	.10	
African-American	0.0%	6.1%	0.0%			
Asian-American	21.2%	9.1%	11.4%			
Caucasian	72.7%	66.7%	85.7%			
Hispanic	0.0%	6.1%	0.0%			
Other	6.1%	12.1%	2.9%			
Educational*						
LD diagnosis	6.1%	6.1%	2.9%	$\chi^2(2, N = 101) = .50$	.78	
ID diagnosis	0.0%	0.0%	0.0%	-	-	
Special Ed	15.2%	6.1%	2.9%	$\chi^2(2, N = 101) = 3.75$	.15	
IEP	6.1%	9.4%	2.9%	$\chi^2(2, N = 100) = 1.26$	.53	
Clinical*						
GAD diagnosis	3.0%	3.0%	2.9%	$\chi^2(2, N = 101) = .002$	1.0	
MD diagnosis	0.0%	0.0%	0.0%	-	-	
ADHD diagnosis	15.6%	0.0%	11.4 %	$\chi^2(2, N = 100) = 5.23$	.073	
ODD diagnosis	0.0%	0.0%	0.0%	-	-	
CD diagnosis	3.1 %	0.0%	0.0%	$\chi^2(2, N = 100) = 2.15$	.34	

Note. N = 101; 2C = two-component WMT; $1C_A$ = adaptive one-component WMT; $1C_{NA}$ = non-adaptive one-component WMT; SES = total family income in US Dollars; Gender = percent male; Handedness = percent right handed; *Parent-reported variables (yes/no): LD = Learning Disorder; ID = Intellectual Disorder; Special Ed =

special education, remedial services, special class or school; IEP = Individualized Education Program currently or previously; GAD = Generalized Anxiety Disorder; MD = Major Depressive Disorder; ADHD = Attention Deficit Hyperactivity Disorder; ODD = Oppositional Defiant Disorder; CD = Conduct Disorder.

Table 4

ANOVAs comparing two-component, adaptive one-component, and non-adaptive one-component WMT groups included in the final analysis on training related variables

	ANOVA			**2C (n = 33)**	**$1C_A$ (n = 33)**	**$1C_{NA}$ (n = 35)**
	F	*p*	η_p^2	M (SE)	M (SE)	M (SE)
Pre-test parent ratings						
BRIEF-WM	$F(2, 98) = 1.92$	.15	.038	16.8 (.80)	14.8 (.80)	15.0 (.77)
ADHD	$F(2, 98) = 2.32$	.10	.045	10.7 (1.47)	6.21 (1.47)	8.37 (1.43)
Post-training motivation						
Interest/Enjoyment	$F(2, 98) = .48$	.62	.010	24.2 (1.44)	26.2 (1.44)	24.9 (1.40)
Effort/ Importance*	$F(2, 97) = 2.60$	.080	.051	28.2 (.72)	29.1 (.73)	30.5 (.70)
Sight word reading						
TOWRE	$F(2, 98) = .88$	.42	.018	82.5 (1.57)	80.1 (1.57)	79.9 (1.53)
Compliance						
Trained Days	$F(2, 98) = .24$	.79	.005	24.8 (.10)	24.8 (.10)	24.9 (10)

Note. 2C = two-component WMT; $1C_A$ = adaptive one-component WMT; $1C_{NA}$ = non-adaptive one-component WMT; BRIEF-WM = parent-rated WM on the Behavior Rating Inventory of Executive Functions; ADHD = total parent-rated ADHD symptoms on DuPaul ADHD Rating Scale (Parent Rating Form); Interest/Enjoyment = Intrinsic Motivation Inventory Interest/Enjoyment Scale; Effort/Importance = Intrinsic Motivation Inventory Effort/Importance Scale. * $1C_A$, N = 32 for Effort/Importance Scale of IMI.

Table 5

Pre, post, and follow-up outcome measure descriptive statistics for participants included in the final analysis

		2C			**$1C_A$**			**$1C_{NA}$**		
		Pre	Post	Follow	Pre	Post	Follow	Pre	Post	Follow
Near Transfer										
Recall Latency	N	33	33	31	33	33	32	35	35	31
	M	8309	8227	7947	7723	7678	7725	8631	8108	7995
	SD	2338	1648	1732	1611	1915	2214	1534	1722	2034
Total Correct	N	33	33	31	33	33	32	35	35	31
	M	21.2	20.8	22.1	22.5	20.4	23.3	22.8	22.1	22.3
	SD	3.98	5.04	5.85	5.50	4.99	5.08	4.78	5.56	5.97
Total Errors	N	33	33	31	33	33	32	35	35	31
	M	9.85	15.0	12.3	7.08	8.96	9.81	8.74	10.9	9.94
	SD	7.40	9.42	7.62	4.36	5.19	5.50	4.35	6.19	5.50
Far transfer										
gF	N	33	33	33	33	33	33	35	35	35
	M	4.79	5.83	6.35	5.06	5.70	6.39	5.57	5.87	6.40
	SD	1.72	1.87	2.01	1.89	2.00	1.82	1.92	1.68	1.75
Reading Total	N	33	33	32	32	33	33	35	34	35
	M	228	227	231	228	226	230	226	228	227
	SD	9.48	11.2	9.87	10.8	12.9	10.2	10.1	10.5	13.0
Literature Comp.	N	33	30	32	32	32	33	35	34	35
	M	228	225	231	227	225	230	224	228	224
	SD	9.65	12.5	11.0	12.9	15.5	12.8	9.64	11.3	12.6
Non-fiction Comp.	N	33	30	32	32	32	33	35	34	35
	M	228	226	229	228	226	229	224	226	227
	SD	10.5	12.7	12.2	11.5	12.9	11.1	11.1	11.8	15.6
Vocabulary	N	33	30	32	32	32	33	35	34	35
	M	228	230	232	230	227	232	228	231	231
	SD	10.8	12.3	10.1	11.7	14.7	11.3	13.6	12.0	14.3
Math Total	N	33	33	33	31	33	33	35	35	35
	M	242	242	247	243	244	249	242	244	250
	SD	12.2	13.1	14.6	15.0	14.9	15.3	13.6	13.1	15.5
Number sense	N	33	30	33	31	32	33	35	35	35
	M	242	245	246	244	247	251	242	248	252
	SD	13.4	14.2	16.6	19.1	17.7	21.5	14.4	17.2	19.8
Computation	N	33	30	33	31	32	33	35	35	35
	M	243	242	245	241	243	246	242	244	248
	SD	15.1	15.7	16.5	15.0	17.7	13.1	15.8	14.2	15.2
Algebra	N	33	30	33	31	32	33	35	35	35
	M	240	243	250	241	247	250	243	244	251
	SD	12.8	14.1	16.3	17.3	16.7	16.1	14.5	14.2	16.8
Geometry	N	33	30	33	31	32	33	35	35	35
	M	241	241	247	243	244	247	241	243	250
	SD	14.6	15.8	16.4	17.3	14.9	16.7	15.6	16.7	18.6
Data Analysis	N	33	30	33	31	32	33	35	35	35
	M	243	241	244	243	242	248	244	244	248
	SD	13.4	14.6	15.0	15.4	13.7	15.7	15.3	14.1	15.5

Note. 2C = two-component WMT; $1C_A$ = adaptive one-component WMT; $1C_{NA}$ = non-adaptive one-component WMT; Recall latency values represent milliseconds.

Table 6

ANOVAs comparing two-component, adaptive one-component, and non-adaptive one-component WMT groups on pre-test SM capacity, gF, and academic growth measures

	ANOVA				**2C**	**$1C_A$**	**$1C_{NA}$**
	F	*p*	η_p^2		M (SE)	M (SE)	M (SE)
Near Transfer							
Recall Latency	$F(2, 98) = 2.07$	.13	.041	N	33	33	35
					8309 (323)	7723 (323)	8631 (314)
Total Correct	$F(2, 98) = 1.07$	.35	.021	N	33	33	35
					21.2 (.84)	22.5 (.84)	22.8 (.81)
Total Errors	$F(2, 98) = 2.10$	.13	.041	N	33	33	35
					9.85 (.96)	7.08 (.96)	8.74 (.94)
Far transfer							
gF	$F(2, 98) = 1.59$	.21	.031	N	33	33	35
					4.79 (.32)	5.06 (.32)	5.57(.31)
Reading Total	$F(2, 97) = .81$	.45	.016	N	33	32	35
					228 (1.77)	228 (1.79)	226 (1.72)
Literature Comp.	$F(2, 97) = 1.03$	.36	.021	N	33	32	35
					228 (1.88)	227 (1.91)	224 (1.82)
Non-fiction Comp.	$F(2, 97) = 1.41$	.25	.028	N	33	32	35
					228 (1.92)	228 (1.95)	224 (1.87)
Vocabulary	$F(2, 97) = .22$	.81	.004	N	33	32	35
					228 (2.12)	230 (2.15)	228 (2.05)
Math total	$F(2, 96) = .009$	.99	.001	N	33	31	35
					242 (2.37)	243 (2.45)	242 (2.30)
Number Sense	$F(2, 96) = .19$	.83	.004	N	33	31	35
					242 (2.74)	244 (2.82)	242 (2.66)
Computation	$F(2, 96) = .053$	.95	.001	N	33	31	35
					243 (2.67)	241 (2.75)	242 (2.59)
Algebra	$F(2, 96) = .41$	.66	.009	N	33	31	35
					240 (2.59)	241 (2.68)	243 (2.52)
Geometry	$F(2, 96) = .13$	.88	.003	N	33	31	35
					241 (2.76)	243 (2.84)	241 (2.68)
Data Analysis	$F(2, 96) = .026$	.97	.001	N	33	31	35
					243 (2.57)	243 (2.65)	244 (2.49)

Note. 2C = two-component WMT; $1C_A$ = adaptive one-component WMT; $1C_{NA}$ = non-adaptive one-component WMT; N = sample size; Recall latency values represent milliseconds.

Table 7

Mixed ANCOVAs comparing two-component, adaptive one-component, and non-adaptive one-component WMT groups on SM capacity, gF, and academic growth measures at post-test and follow-up

		Mixed ANCOVA (Controlling for pre-test scores)				**2C**	**$1C_A$**	**$1C_{NA}$**
		F	*p*	η_p^2		M (SE)	M (SE)	M (SE)
Near Transfer								
Recall Latency	Time	*F*(1, 90) =	.80	.001	N	31	32	31
	Condition	*F*(2, 90) = .23	.80	.005	Post	8226 (291)	7801 (290)	7889 (294)
	Time x Condition	*F*(2, 90) = .31	.74	.007	Follow	7912 (343)	7903 (342)	7847 (346)
Total Correct	Time	*F*(1, 90) = 8.73	.004	.088	N	31	32	31
	Condition	*F*(2, 90) = .067	.94	.001	Post	21.4 (.72)	20.3 (.71)	21.4 (.73)
	Time x Condition	*F*(2, 90) = 1.49	.23	.032	Follow	21.9 (.96)	23.3 (.94)	22.6 (.96)
Total Errors	Time	*F*(1, 90) = 4.35	.040	.046	N	31	32	31
	Condition	*F*(2, 90) = 4.10	.020	.084	Post	13.5 (.90)	9.88 (.89)	10.3 (.89)
	Time x Condition	*F*(2, 90) = .62	.54	.014	Follow	11.9 (1.07)	10.3 (1.05)	9.85 (1.06)
Far Transfer								
gF	Time	*F*(1, 97) = 1.85	.18	.019	N	33	33	35
	Condition	*F*(2, 97) = .12	.89	.003	Post	6.0 (.30)	5.7 (.30)	5.7 (.30)
	Time x Condition	*F*(2, 97) = .26	.77	.005	Follow	6.5 (.30)	6.4 (.30)	6.2 (.30)
Reading	Time	*F*(1, 94) = 2.52	.12	.026	N	32	32	34
Total	Condition	*F*(2, 94) = 1.05	.35	.022	Post	226 (1.18)	225 (1.18)	229 (1.15)
	Time x Condition	*F*(2, 94) = 4.09	.020	.080	Follow	230 (1.21)	230 (1.21)	229 (1.18)
Literature	Time	*F*(1, 90) = 2.27	.14	.025	N	29	31	34
Comp.	Condition	*F*(2, 90) = .090	.91	.002	Post	224 (1.79)	224 (1.72)	229 (1.65)
	Time x Condition	*F*(2, 90) = 7.19	.001	.140	Follow	230 (1.78)	230 (1.71)	226 (1.64)
Non-fiction	Time	*F*(1, 90) = .36	.55	.004	N	29	31	34
Comp.	Condition	*F*(2, 90) = .89	.42	.019	Post	224 (1.81)	226 (1.74)	228 (1.68)
	Time x Condition	*F*(2, 90) = .038	.96	.001	Follow	228 (2.01)	229 (1.93)	230 (1.87)
Vocabulary	Time	*F*(1, 90) = .007	.93	.001	N	29	31	34
	Condition	*F*(2, 90) = 1.33	.27	.029	Post	231 (1.87)	227 (1.81)	231 (1.73)
	Time x Condition	*F*(2, 90) = .91	.41	.020	Follow	232 (1.59)	231 (1.54)	232 (1.47)
Math Total	Time	*F*(1, 95) = 1.09	.30	.011	N	33	31	35
	Condition	*F*(2, 95) = 1.30	.28	.027	Post	242 (1.14)	243 (1.18)	244 (1.11)

	Time x Condition	*F*(2, 95) = .23	.80	.005	Follow	247 (1.34)	248 (1.38)	250 (1.30)
Number	Time	*F*(1, 91) = 2.91	.091	.031	N	30	30	35
Sense	Condition	*F*(2, 91) = 1.30	.28	.028	Post	245 (2.19)	244 (2.20)	248 (2.03)
	Time x Condition	*F*(2, 91) = .68	.51	.015	Follow	247 (2.61)	250 (2.62)	253 (2.42)
Computation	Time	*F*(1, 91) = .20	.65	.002	N	30	30	35
	Condition	*F*(2, 91) = .72	.49	.016	Post	242 (2.13)	242 (2.13)	244 (1.97)
	Time x Condition	*F*(2, 91) = .35	.71	.008	Follow	244 (2.00)	246 (2.00)	248 (1.84)
Algebra	Time	*F*(1, 91) =.054	.82	.001	N	30	30	35
	Condition	*F*(2, 91) = .81	.45	.018	Post	244 (1.62)	247 (1.62)	242 (1.50)
	Time x Condition	*F*(2, 91) = 1.86	.16	.039	Follow	252 (1.95)	251 (1.94)	250 (1.80)
Geometry	Time	*F*(1, 91) = .46	.50	.005	N	30	30	35
	Condition	*F*(2, 91) = 1.02	.37	.022	Post	241 (1.85)	242 (1.85)	244 (1.71)
	Time x Condition	*F*(2, 91) = 1.04	.36	.022	Follow	247 (2.10)	246 (2.10)	250 (1.94)
Data Analysis	Time	*F*(1, 91) = .85	.36	.009	N	30	30	35
	Condition	*F*(2, 91) = .87	.42	.019	Post	241 (1.95)	242 (1.95)	244 (1.81)
	Time x Condition	*F*(2, 91) = .47	.62	.010	Follow	244 (2.08)	248 (2.08)	248 (1.93)

Note. 2C = two-component WMT; $1C_A$ = adaptive one-component WMT; $1C_{NA}$ = non-adaptive one-component WMT; N = sample size; Mean scores from ANCOVAs are adjusted to control for corresponding pre-test scores.

Figure Captions

Figure 1. Flowchart of participants from recruitment to final analysis

Figure 2. Average span length achieved on common and critical exercises by two-component and adaptive one-component participants with Profile 1 and Profile 2. Error bars reflect standard errors of the mean.

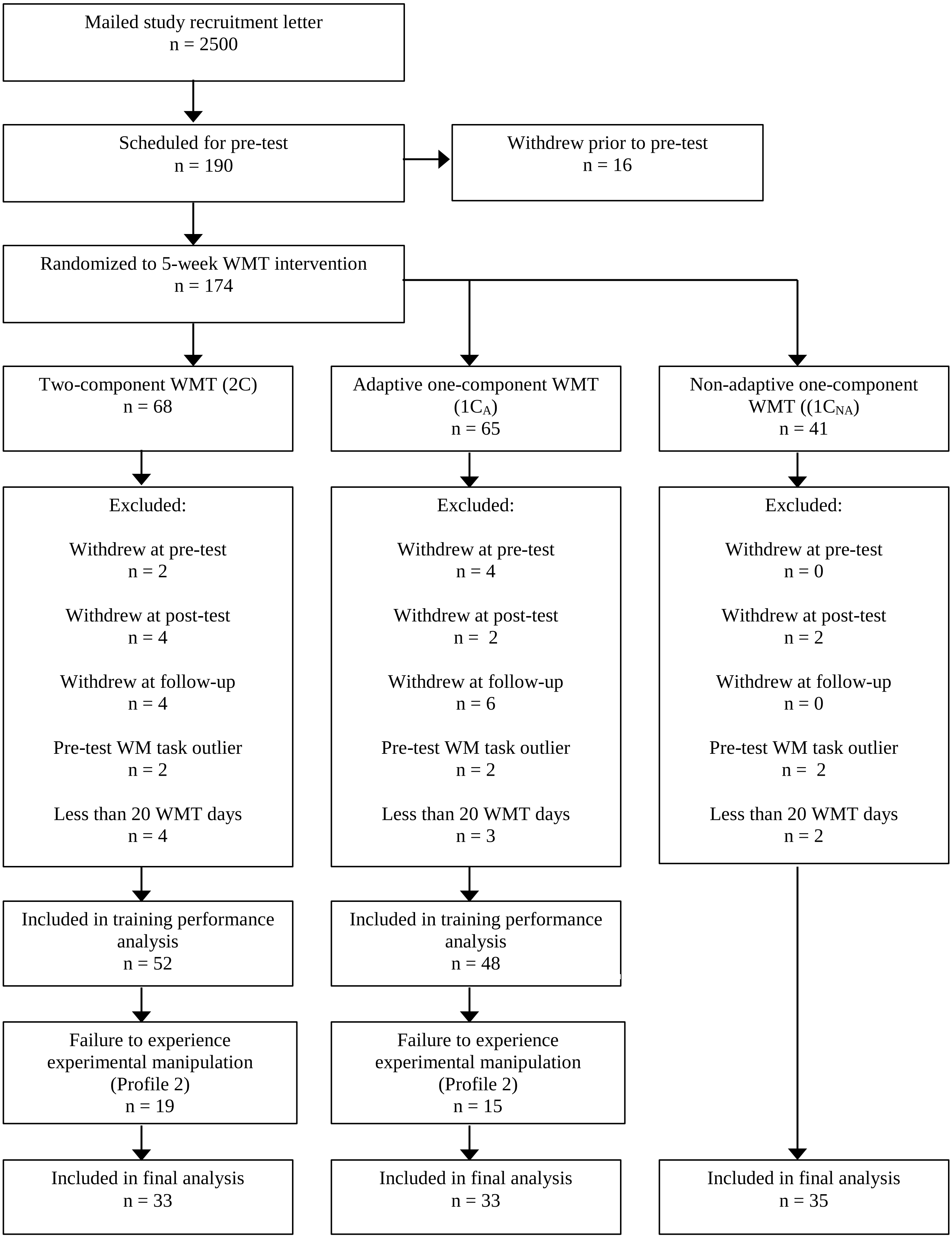

Mailed study recruitment letter
n = 2500
Scheduled for pre-test
n = 190
Withdrew prior to pre-test
n = 16
Randomized to 5-week WMT intervention
n = 174
Two-component WMT (2C)
n = 68
Adaptive one-component WMT
(1CA)
n = 65
Non-adaptive one-component
WMT ((1CNA)
n = 41
Excluded:
Withdrew at pre-test
n = 2
Withdrew at post-test
n = 4
Withdrew at follow-up
n = 4
Pre-test WM task outlier
n = 2
Less than 20 WMT days
n = 4
Excluded:
Withdrew at pre-test
n = 4
Withdrew at post-test
n = 2
Withdrew at follow-up
n = 6
Pre-test WM task outlier
n = 2
Less than 20 WMT days
n = 3
Excluded:
Withdrew at pre-test
n = 0
Withdrew at post-test
n = 2
Withdrew at follow-up
n = 0
Pre-test WM task outlier
n = 2
Less than 20 WMT days
n = 2
Included in training performance
analysis
n = 52
Included in training performance
analysis
n = 48
Failure to experience
experimental manipulation
(Profile 2)
n = 19
Failure to experience
experimental manipulation
(Profile 2)
n = 15
Included in final analysis
n = 33
Included in final analysis
n = 33
Included in final analysis
n = 35

Figure 1.

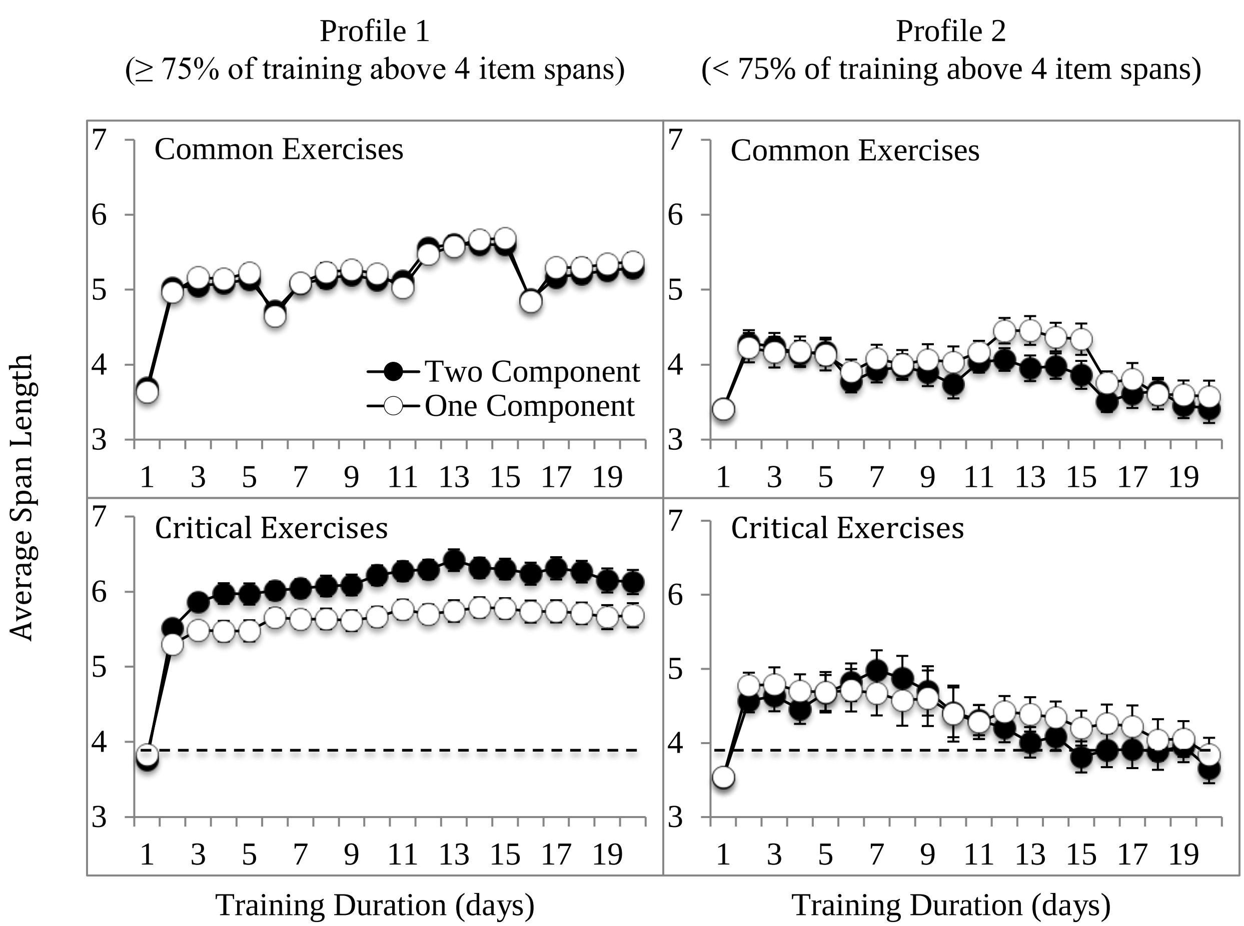
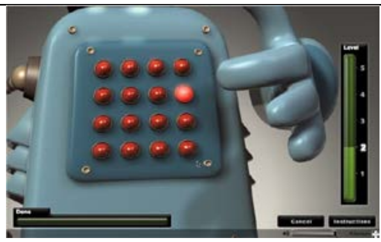
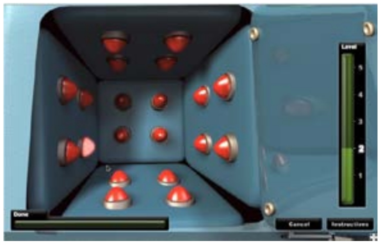
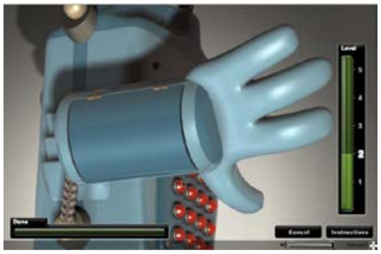
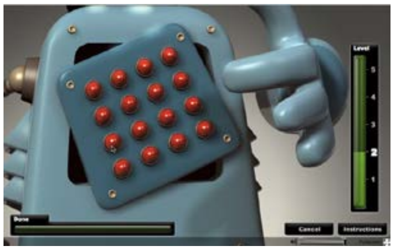
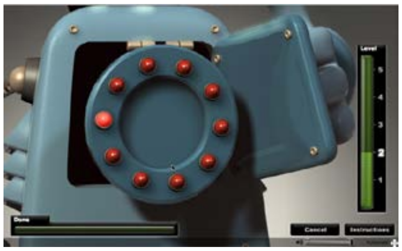
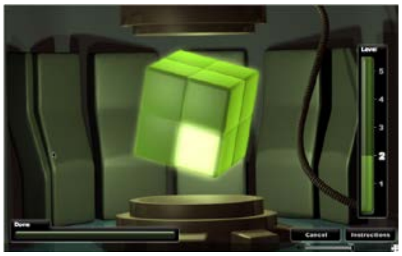

Profile 1
(≥ 75% of training above 4 item spans)
Profile 2
(< 75% of training above 4 item spans)
Average Span Length
Common Exercises
Two Component
One Component
Critical Exercises
1 3 5 7 9 11 13 15 17 19
3 4 5 6 7
Training Duration (days)

Figure 2.

Appendix

Working Memory Exercises

Exercise	Type	Domain	Description	
Visual Data Link	Critical	Spatial	A series of lamps will light up on a four by four grid. After a brief delay, the participant must click on the lamps in the same order that they lit up.	
Data Room	Critical	Spatial	A series of lamps in a 3D room will light up. After a brief delay, the participant must click on the lamps in the same order that they lit up.	
Input Module without Lid	Critical	Verbal	A series of digits will be read out loud and simultaneously light up. The participant clicks on the numbered buttons in the forward order. *Commercially available Cogmed requires backward recall.	
Input Module with Lid	Critical	Verbal	A series of digits will be read out loud. Participants cannot see the numbered buttons as they are read. The numbers will appear when it is the participant's turn to click on the numbered buttons in the forward order. *Commercially available Cogmed requires backward recall.	

Rotating Data Link	Common	Spatial	A series of lamps will light up on a four by four grid. After a brief delay, the participant must click on the lamps in the same order that they lit up. However, before the participant responds, the entire panel will rotate 90 degrees.	
Rotating Dots	Common	Spatial	A series of lamps will light up on a continuously moving rotating circle. After a brief delay, the participant must click on the lamps in the same order they lit up.	
3D Cube	Common	Spatial	A series of panels will light up on a 3D cube. The cube rotates to show which panels are lit up. After a brief delay, the participant must click on the panels in the same order that they lit up.	
Asteroids	Common	Spatial	A series of continuously moving asteroids will light up in particular order. After a brief delay, the participant must click on the asteroids in the same order that they lit up. As the participant increases through span levels, the number of asteroids in the display will increase.	

Space Whack	Common	Spatial	Monsters will randomly appear in craters. Before they emerge, they let out a little cloud of gas. The participant needs to remember the sequence of the gas clouds in order to be able to hit the monsters on the head when they appear. It is important that the Participant waits until all gas clouds have shown and is prepared to hit each monster by starting with the pointer above the first crater.	
Sorter	Common	Verbal	A series of boxes will light up on a four by four grid. As each box lights up, a number will be revealed and then disappear. After a brief delay, the participant must click on the boxes in numerical order (and not necessarily in the order that they lit up).	
Stabilizer	Common	Verbal	A series of letter will be read aloud. When a letter is read, it will be displayed in the middle circle, and at the same time, a corresponding lamp will light up. After all the letters have been read, one of them will be displayed again in the middle circle. The participant must click the lamp that came for that particular letter.	

Decoder	Common	Verbal	A series of letters will be said aloud. The participant needs to recall the letters in the forward order from list of three presented letters. For example, the program says “D, P, E” The first letter is ‘D’ and the participant must select that letter from the three options under the first lamp. At the next lamp, ‘P’ must be selected and ‘E’ from the choices under the third lamp.	

Note. Exercises descriptions and images have been adapted from Cogmed training materials.