Spring 4-5-2015

Current Methods and Future Directions Regarding Working Memory Training Research

Emily G. Wright
emily.wright@otterbein.edu

Recommended Citation

CURRENT METHODS AND FUTURE DIRECTIONS REGARDING WORKING MEMORY
TRAINING RESEARCH

Emily G. Wright

Department of Psychology

Otterbein University

Westerville, Ohio 43081

5 April 2015

Submitted in partial fulfillment of the requirements for

graduation with Distinction

Advisory Committee:

Meredith Frey, Ph.D.   _____

Distinction Advisor   Advisor's Signature

Cynthia Laurie-Rose, Ph.D.   _____

Second Reader (or Co-Advisor if applicable)   Second Reader's Signature

Abstract

Cognitive training is the process through which individuals perform a series of computerized tasks over a period of weeks for the purpose of improving a variety of cognitive abilities. Cognitive training is important in commercial, clinical, and educational fields alike, considering the possibilities of improving and sustaining cognitive abilities in both typically developing and cognitively deficient populations. The present review assesses cognitive training paradigms targeting working memory. Working memory is a predictor of academic achievement and is closely related to mechanisms of higher cognition. In particular, the present review focuses past studies which have investigated the effects working memory training on improved performance in fluid intelligence. At conclusion, future directions for cognitive training research are presented.

Table of Contents

Current Methods and Future Directions Regarding Working Memory Training Research

The proposed benefits of cognitive training have sparked an interest in clinical, educational, and commercial populations alike. Commercially available training programs like Lumosity, or the more clinically-researched Cogmed, present visually-appealing, game-like tasks to train a wide array of cognitive functions that have produced mixed results of both near- and far-transfer to working memory (WM), fluid intelligence (Gf), and various executive functions following several weeks of daily training (Klingberg, 2010). Transfer, as defined by Morrison and Chein (2011), is the event in which improvements in one cognitive ability result in improvements in other cognitive areas. For example, near-transfer occurs when individuals complete WM training using a particular task and, following training, exhibit additional improvements in untrained tasks that also measure WM. Alternatively, far-transfer occurs when individuals complete WM training and exhibit improved performance in untrained tasks that do not target WM, when pre- and post-test scores are compared. In the present case, WM training and improvements that lead to improvements from pre- to post-test in Gf measures would be considered far-transfer (Jaeggi, Buschkuehl, Jonides, & Shah, 2011)

Single-task training paradigms using Dual n-Back and Complex Span Tasks (CSTs) have been shown to produce transfer effects that allow researchers to better understand the most effective qualities of training (Redick, Broadway, Meier, Kuriakose, Unsworth, Kane, & Engle, 2012). Due to the established relationship between working memory and fluid intelligence, past studies have concluded that researchers should not focus on whether cognitive training works in improving performance in WM, but rather, whether those benefits can transfer to untrained tasks and, further, under what conditions these transfer effects can be optimized (Jaeggi et al., 2011).

Working memory, as defined by Chein and Morrison, is the retention of information over a brief period of time and is of central importance for a wide range of cognitive tasks and for academic achievement (2011). WM improvement has become the focus of many paradigms because of WM's practical applications in daily life, as well as the finding that WM is a strong predictor of individual differences in Gf and executive functioning (Klingberg, 2010). WM has been linked to Gf through measures of far-transfer following specific training paradigms (Jaeggi et al., 2011). It is therefore suggested that Gf, the ability to reason and to solve problems independently of previously acquired knowledge, can be altered by experience when under specific condition (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). The debate surrounding the degree to which WM training can produce far-transfer effects to Gf has been illustrated in much of the literature that will be presented.

Given the established relationship between working memory and higher cognition, the primary objective of this review is to evaluate current literature with the goal of identifying how cognitive training programs can be optimized to significantly improve WM capacity, along with determining what training regimens and conditions may result in better far-transfer effects to untrained tasks. This review will serve to identify ways in which future research can aid in constructing more effective cognitive training programs to benefit populations with known WM deficiencies and improve ease and success of cognitive performance in one's daily life (Chein & Morrison, 2010). A proposal containing suggestions for future study on cognitive training will be provided in conclusion to address these issues.

Working Memory and Higher Cognition

Baddeley and Hitch (1974) proposed a model for working memory to provide a more accurate illustration of short-term memory and has since become a widely used platform for WM study. The original model was comprised of three primary components: The central executive, which serves as the supervisory system, and two slave subsystems, the phonological loop, and the visuospatial sketchpad. Later, a third subsystem was introduced called the episodic buffer (Repovs & Baddeley, 2006). This model describes short-term memory as an active process and allows for explanation of more complex cognitive processes.

The central executive is described by Baddeley and Hitch (1974) as being the control system of limited attentional capacity that is responsible for the manipulation of information within WM. The central executive mediates the three subsystems and is involved whenever information within these subsystems must be manipulated for dual-task performance. In complex cognitive abilities, the central executive is mostly involved as a source of attentional control, focused attention, and attentional switching; therefore, the central executive relies on information provided by the three subsystems to guide these processes

The first of the three subsystems to be described is the phonological loop. This system is comprised of two components: A phonological store, which holds memory traces in acoustic or phonological form for a few seconds, and the articulatory rehearsal, which retrieves and rearticulates the contents of the phonological store to refresh memory traces. While speech, for example, enters the phonological store automatically, information from other modalities must be recoded by the articulatory rehearsal before it can enter the phonological store. Since this process occurs in real time, the capacity of the phonological store is limited by the number of items that

can be articulated before the memory traces fade away. This limitation was exhibited, for example, through the finding that immediate memory for word sequences declines as word length increases, for memories decay faster than the words can be rehearsed (Repovs & Baddeley, 2006). The initial assumptions of this system seem to have withstood continued empirical testing and have since indicated that they are well-capable of explaining phenomena related to verbal working memory (Baddeley & Hitch, 1974).

Baddeley and Hitch further described the second slave subsystem, the visuospatial sketchpad, as being responsible for maintaining and manipulating visual and spatial information, a process that is crucial for performing an array of cognitive tasks (1974). Furthermore, the visuospatial WM system can be broken down into purely visual and purely spatial subcomponents, with each containing separate and independent passive storage, representations, mechanisms of maintenance, and manipulation. Both subcomponents have been shown to be closely related to forms of visual attention. The encoding of information in visual working memory has been shown to be significantly affected by both bottom-up perceptual features, and by top-down influences based on previous experience, such as category learning (concept formation). While visual working memory is closely related to perception and visual imagery, spatial working memory shows closer connection to attention and action.

The newest of the slave subset systems is the episodic buffer (Repovs & Baddeley, 2006). The episodic buffer represents a separate storage system of limited capacity. It is episodic, for it holds information that is compiled from a range of systems, including other working memory components and long-term memory, scenes or narrative episodes. It is considered a buffer, because it serves as an intermediary between the other subsystems. The integration and maintenance of information within the episodic buffer depends on a limited capacity attentional

system, namely the central executive. The retrieval of information is based on conscious awareness. Together, with the ability to create and manipulate novel representations, the episodic buffer providing the basis for planning future action based on mental modeling of outcomes.

This model is important in explaining how WM functions provide the ability to maintain and manipulate information necessary to perform complex cognitive tasks. Furthermore, these subsystems, and the mediation performed by the central executive, are targeted by a variety of WM training programs. By identifying which specific components of these subsystems are targeted by a single paradigm, researchers will develop a better understanding of how the targeting of certain components will not only produce performance gains throughout training, but will also produce transfer to other cognitive abilities that utilize these similar components.

When discussing the multitude of training tasks and programs used to target WM performance, many researchers divide components of WM systems into two factors: domain-specific and domain-general (executive) (Turley-Ames & Whitfield, 2003). The domain-specific factors of WM include strategies tied to the maintenance and management of various types of information. Often discussed is the domain-specific strategy of "articulatory control." This strategy involves the use of inner speech mechanisms to maintain representations of linguistic items that have been verbally coded (Richmond, Morrison, Chein, & Olson, 2011). In contrast, domain-general factors of WM include mechanisms that control attention, gate the flow of information in and out of WM buffers, such as visual or episodic buffers, reduce the interference of sources of information irrelevant to current WM processing, and govern the engagement of domain-specific strategies. In general, domain-general processes include those that are not associated with a particular type of information or sensory modality, but that nonetheless aid in the encoding, storing, and retrieval of information from WM (Morrison & Chein, 2010).

Both domain-general and domain-specific factors are involved in the link between WM and higher cognition, but the executive attention aspects of domain-general processing, more than domain-specific factors, seem to drive WM validity in many higher cognitive skills (Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). Considering such, the greatest generalizations from cognitive training might be expected when the training task targets the aforementioned domain-general processes. Approaches to WM training can often be classified according to their primary targeting of either domain-general or domain-specific factors of the WM system. The two most common types of training paradigms are strategy training and core training.

Strategy training is often intended to promote the use of domain-specific strategies that allow participants to remember increasing amounts of information of a particular type (McNamara & Scott, 2001). The primary aim of most strategy training paradigms is to increase performance in tasks that require the retention of information over a delay (Morrison & Chein, 2010). In strategy training sessions, participants are often given a series of introductions to the strategy mechanisms that experimenters wish the participants to employ to succeed in the task, and then provide them with practice trials to encourage the strategy of interest. The procedures that make up strategy training are known to promote rehearsal, thus improved WM performance by shifting trainees away from less effective strategies, or by increasing the quality of efficiency of covert rehearsal mechanisms that support maintenance of WM (Shipstead, Hicks, & Engle, 2012).

In a study performed by Turley-Ames and Whitfield, researchers aimed to determine how strategy training influences WM span performance (2003). Additionally, Turley-Ames and Whitfield aimed to determine what impact strategy instruction had on correlation between WM

span scores and reading ability as a form of higher cognitive function. Participants completed two versions of the Operation Span Test. Before testing, participants were given 12 practice trials to familiarize themselves with the task. After completing one version, half of the participants were given instructions that contained strategy information (a rehearsal strategy) to be used during the second Operation Span Test. In this test, participants solved simple math problems while remembering unrelated words that followed each problem. After presentation of 2–6 problem-word sequences, participants were asked to recall the words presented in the preceding set. There were three trials each for the 2, 3, 4, 5, and 6-sequence sets, totaling 15 trials. Both versions of the test were fundamentally the same, but contained different problems and words to be remembered.

After completing the pre- and post-operation span measures, the Nelson-Denny reading test was administered. This test is composed of two subtests, a vocabulary subtest and a reading comprehension subtest. The vocabulary subtest contained 80 multiple-choice questions while the reading comprehension subtest contained 7 passages and 38 multiple-choice questions. Results indicated the change in WM span scores from pre- to post-test was greater for those assigned to the rehearsal condition. However, it was unclear whether certain individuals benefited more from rehearsal strategy instruction. The marginally significant interaction between pre-WM span scores and the rehearsal condition suggested that individuals with low span performance may have benefited more from rehearsal strategy. If this is indeed true, then is it likely that low-performing individuals might have improved because the rehearsal strategy helped them manage the demands of the task by reallocating their WM resources to better accommodate the dual task at hand. It is also likely that the rehearsal strategy allowed previously low-performing participants to increase the amount of time processing the words to be remembered.

Regarding transfer to reading ability, Turley-Ames and Whitfield discovered that both pre- WM span scores and post-WM span scores (with rehearsal strategy) predicted reading ability; however, the interaction between post-WM span scores and post-test condition failed to achieve significance, suggesting that post-WM span scores in the rehearsal condition were not a better predictor of reading ability. Researchers noted that, while a single strategy was not introduced to the control group, it was likely that individuals naturally used a variety of strategies to manage the difficulty of the task; thus, WM span scores represent individual differences in both capacity and strategy.

The primary expectation of strategy training is that studies should yield improved performance only with tasks that are similar to those of the trained strategy (near transfer, and should not generalize with far transfer to untrained tasks; however, strategy training may lead to generalized benefits when practices are confined to a particular task and information type (Turley-Ames & Whitfield, 2003). Due to the rarity of strategy training use, and the prediction that generalized transfer will not be consistently observed under most conditions, strategy training's most principal value is that it may contribute to the enhancement of skills that rely directly on WM and are conducive to a trainable strategy (Morrison & Chein, 2010).

Core training, in contrast, involves the repetition of demanding WM tasks that target domain-general mechanisms (Klingberg, Forssberg, & Westerberg, 2002). In general, core training paradigms often aim to limit domain-specific strategies and automization, while, in turn, including tasks that span multiple modalities, enforcing rapid WM encoding and retrieval demands, and offering high cognitive workloads (Morrison & Chein, 2011). Core training paradigms are often structured under the "kitchen-sink" approach. This approach compiles several tasks featuring diverse stimulus types that impact multiple aspects of cognitive

functioning (attention, memory, fluid intelligence, spatial ability, etc.) and often span multiple modalities (Shipstead, Hicks, & Engle, 2012). The diversity of the tasks comprising core programs increases the chance that one task, or a combination of tasks, will produce the desired training gains, and near- and far-transfer; therefore, it is the ideal long-term goal for researchers to determine a battery of equally effective tasks that contribute to training gain and cognitive improvement in an additive fashion, ultimately leading to a significantly greater transfer effect with reliable, maximum efficiency (Klingberg, 2010).

The most prominent and troubling drawback to many core training paradigms is that the use of multiple tasks, all differing in stimuli and targeted cognitive processes, presents a substantial difficulty in identifying which precise components of the training paradigm are most responsible for present cognitive improvements, specifically which exact mechanisms of WM were improved. To minimize the ambiguity of transfer sources, some researchers use a more stripped down regimen featuring only one task in the experimental training group (Jaeggi et al., 2011; Verhaeghen, Cerella, & Basak, 2004). This approach allows researchers to assume that the single task is responsible for any training effects that are observed, although, due to the complexity of most single task, core training paradigms, it is often still unclear exactly which mechanisms of WM are responsible for present training gains. To address this issue, some training paradigms are built around a single component or mechanism of WM processing. For example, Dahlin, Neely, Larsson, Backman, and Nyberg implemented a "letter updating task" during which participants were shown lists of letters with an unknown length and asked to repeat the last four letters in the series. This task was designed specifically to emphasize the updating mechanism of WM, an index specific component process of WM, and produced observable training gains (2008).

Compared to the results of strategy training, core training seems to produce clear improvements in tasks directly involved in the retention and retrieval of temporarily stored information (Morrison & Chein, 2011). In general, core training studies report that participants exhibit significantly improved performance on the trained WM tasks presented (Klingberg, Fernell, Olesen, Johnson, Gustafsson, Dahlstrom et al., 2005). Core training is increasingly being gauged by assessment in its generalization in clinical, school-aged, and other special populations that emphasize the practical application of cognitive training. In clinical populations, the value of core training has been expressed through training effects demonstrated as far-transfer in laboratory, everyday memory, and quality of life measures (e.g. self-reports of elderly individuals' improved abilities to navigate their homes) (Klingberg, 2010). By strengthening the domain-general processes previously discussed, core training paradigms should yield improvements not only in near-transfer, but also to these more disparate cognitive measures. It's intuitive, then, to state that core training will increase performance in other, untrained cognitive tasks that are reliant on WM capacity, such as fluid intelligence, reading comprehension, and cognitive control (Jaeggi et al., 2008).

Relationship between Working Memory (WM) and Fluid Intelligence (Gf)

WM and Gf share a common capacity constraint and can be expressed either as the number of items that can be held in WM or as the number of interrelationships among elements in a reasoning task (Holmes & Gathercole, 2013). The reason for a common capacity limitation is assumed to derive from the primary relation of WM and Gf through attentional control processes (Jaeggi et al., 2008). Carpenter Just and Shell proposed that individual differences in typical tasks that measure Gf (such as Raven's Advanced Progressive Matrices test) account from the ability to derive abstract relations and to maintain a large set of possible goals in WM

(1990). It may be possible, according to Jaeggi et al., to produce transfer effects from a trained task to a reasoning task in which performance relies, to a large extent, on the same processes; that is, there may be potential for transfer after training on WM (2008).

To further evaluate these claims, Jaeggi et al. continued investigation through a study in which researchers evaluated both short- and long-term effects of cognitive training, including a 3 month follow-up testing session (2011). Elementary and middle school children in the experimental group trained on a video game-like working memory task (a spatial single n-back task) for one month, five times per week, for 15 minutes per session. In the spatial single n-back task, participants were presented with a sequence of stimuli at one of six possible spatial locations and asked to press a key when the currently presented stimulus was at the same location as the one n items back in a series and another key if the currently presented stimulus was not located in the same area as the one n items previously presented in the series. Each stimulus was presented randomly for the 15 trials (rounds). This task was adaptive in nature; that is, if the participant committed four or more errors in three consecutive rounds, the n-back level of the next round would be decreased by one, while participants who committed three or fewer errors in a single round would observe an increase of one in the n-back level of their next round. After the end of each round, a feedback screen was presented to indicate movement onto the next level of n-back presentation, with n varied between 1 and 4.

An active control group was formed and trained for the same duration using a knowledge- and vocabulary-based task that required use of crystalized intelligence (Gc). In this self-paced task, participants were presented questions in the middle of a screen, offered four potential answers, and asked to select the appropriate responses. Following the answer selections, a feedback screen appeared to inform each participant whether his/her answer was

correct, along with additional factual information in some cases. The nature of this control group is unique, for many studies do not feature active control groups that train on a knowledge-based task to simulate the training structure (Ranganath, Flegal, & Kelly, 2011). Both the control and experimental group's tasks were provided narrative and graphical themes to provide context to the instructions provided for each task and maximize motivation and compliance throughout training (e.g., space, haunted house, pond, and pirate ship themes). Posttest questionnaires later indicated that children found both training tasks to be equally motivating.

Pretest, posttest, and follow-up measures were administered to assess matrix reasoning using two different tasks: the Test of Nonverbal Intelligence (TONI) and Raven's Standard Progressive Matrices (SPM). Two reasoning tasks were chosen to minimize the presence of inconsistent results, as are sometimes yielded during research on training and transfer. Follow-up measures were administered 3 months following the posttest period. To examine whether individual differences in training gain might affect Gf transfer, researchers split results of the experimental group at the median to form two subgroups for a comparative analysis: A large training gain group and a small training gain group.

Results indicated that no significant improvement in Gc was observed in the active control group, while significant improvement on the trained task in the experimental group was revealed. A repeated-measured ANOVA indicated that there was a significant interaction [$F(1, 30) = 43.37$; $p < 0.001$] between session (the mean n-back level obtained in the first two training sessions vs. the last two training sessions) and group (low training gain vs. high training gain). Post hoc tests indicated that the increase of performance was highly significant for the group with high training gain, whereas the group with small training gain showed no significant improvement. In a comparison of transfer to Gf between the two training subgroups and the

control group, researchers found that only participants in the large training gain group (those participants above the median in WM training improvement) showed transfer to measures of Gf, with no significant differences between the small gain group and the active control group. Furthermore, a positive correlation was present between improvement on the training task and improvement on Gf, suggesting that the greater the gain in the training task, the greater the transfer to Gf. The group differences in Gf gain remained throughout the training hiatus, as reported from the 3-month follow-up testing session.

Jaeggi et al.'s findings suggest that Gf transfer is critically dependent on the amount of improvement during WM training. Some children, as it was shown, failed to improve on the training task (small gain group) and therefore failed to show transfer to the untrained Gf reasoning tasks, despite there being no significant difference between groups on initial WM performance measures prior to the start of training. Along with the finding that some children reported in the posttest that they had difficulty coping with the increased difficulty in the task, researchers suggested that children with large training gains improved more in Gf because they started off with a lower ability and therefore had more room for improvement in the two measures. Furthermore, children with small training gains may have been performing at their ceiling working memory capacity at the start of training. Such factors may explain why there are more significant transfer effects present in WM deficient populations (Klingberg et al., 2005).

Jaeggi et al.'s conclusion could further be supported by the finding that participants in the large training gain group did not end up with significantly higher Gf scores at posttest; however, later analysis indicated that there were no significant group differences between the high initial Gf performance and those with low initial Gf performance in terms of transfer. Additionally, there was no correlation between Gf gain and initial n-back performance during the first two

sessions. Thus, preexisting ability does not seem to be a primary explanation for differences in transfer, whereas the degree of improvement in the trained task, along with perceived difficulty of the trained task, seems to be critical in producing transfer to Gf.

Jaeggi et al. concluded that their findings add to the present body of literature by demonstrating how transfer effects persist over time post-training, but not without having boundary conditions to transfer effects. The findings described present a valuable view on the influence of individual differences on the variability between participants who succeed and fail to produce transfer. This study suggests that, like Gc, there is substantial evidence for environmental influence on Gf improvement. Furthermore, Jaeggi et al. provided evidence that the relationship between WM and Gf supports the transfer between WM and Gf , especially during single-task core training. It is critical that future studies evaluate a variety of WM tasks that may produce similar far-transfer to untrained tasks, to better understand the degree to which improvement in WM predict Gf transfer.

Cognitive Training Paradigms

Cogmed

Cogmed is a WM training protocol heavily researched and updated through its use in studies by Klingberg, initially in collaboration with Forrssberg, and Westerberg with the aim of evaluating WM training effects in ADHD children (2002). In general, Cogmed has been associated with many demonstrations of transfer in clinical populations featuring WM deficits (Morrison & Chein, 2011). In earlier versions of the task battery, Cogmed has been found to produce training benefits that have extended to measures of cognitive control through Stroop testing, and fluid intelligence using Raven's Colored Progressive Matrices (RCPM) in healthy

young adults (Klingberg, Forssberg, & Westerberg, 2002). The program, as updated and used in

a 2005 study, featured a series of visuospatial WM tasks requiring participants to remember the

position of objects in a 4x4 grid, as well as perform verbal tasks involving remembering

phonemes, letters, and digits (Klingberg et al., 2005). Each task in the program was adaptive in

nature, adjusting difficulty automatically on a trial-by-trial basis to match the WM span of the

child in each task.

This follow-up study by Klingberg et al. (2005) also evaluated WM training effects in

ADHD children and required participants in the treatment condition to train on 90 WM trials for

40 minutes each day on the adaptive Cogmed program. Alternatively, the comparison condition

was required to train for the same duration, but on a non-adaptive version of the Cogmed battery

where each task was performed at low level of difficulty consistently through training. Following

5-6 weeks of training, both conditions engaged in a post-intervention testing period and a follow-

up testing period 3 months later. Sufficient compliance, for the purpose of data analysis, was

defined to be 20 or more days of program use. In the pre- and post-intervention assessments, the

span-board task, where participants were asked to replicate a shape pattern in forward or

backward presentation order, was used to measure visuospatial WM, the digit-span task was used

to measure verbal WM, the Stroop interference task was used to measure response inhibition

(cognitive control), and Raven's Colored Progressive Matrices was used to measure nonverbal

reasoning ability. This battery of measurements includes tasks used in other comparative studies

by Klingberg et al. in both healthy and ADHD populations of children (2002; 2010).

The treatment group that trained using high-intensity, adaptive tasks improved

significantly in WM performance in relation to the comparison group on the main outcome

measure, the visuospatial WM-targeting, non-practiced span-board task, and expressed sustained

effect at follow-up. Treatment effects were also determined for response inhibition (Stroop task), verbal WM (digit-span), and complex reasoning (RCPM). Although similar in nature, the span-board task differs from the trained visuospatial WM tasks with respect to the types of stimuli used (blocks on a board versus circles lighting up and disappearing on a screen), stimulus organization (10 irregularly configured blocks versus 14 blocks positioned in a 4x4 grid), and response mode (physically pointing versus interacting via computer mouse). Considering these respective differences, Klingberg et al. concluded that improvement in the span-board task is therefore evidence of near-transfer in visuospatial WM.

Following training, parents reported that there was a significant reduction in inattention and hyperactivity/impulsivity in their children during the follow-up session 3 months following post-test. To further support the effectiveness of visuospatial WM training on ADHD symptoms, comparisons between the effect size of the treatment effect (0.93) to the effect sizes of past studies testing visuospatial WM improvements following stimulant medication intake (0.7-1.2) have shown that the effect of training on visuospatial WM tasks was fully comparable with that of medication (Bedard, Martinussen, Ickowicz, & Tannock, 2004).

Of the secondary outcome measures, which Klingberg et al. stated should be interpreted cautiously, group differences were found for the Stroop task and Raven's task, just as they were found in the preliminary study with ADHD children (Klingberg et al., 2002) and in a study of WM training in adults (Olesen, Westerberg, & Klingberg (2004). With the consistency of these group differences across studies, the results indicate that WM training produces far-transfer to measures of response inhibition and fluid intelligence. Additionally, this improvement in fluid intelligence and reasoning ability after WM training is not only consistent with Jaeggi et al.'s findings (2011), but also with work conducted by Engle, Tuholski, Laughlin, and Conway

(1999), which suggested that WM is necessary for reasoning ability, as visuospatial WM correlates highly with performance on Raven's tasks.

Klingberg and his associates further explain that generalization of WM training to improvements on these measures could be a result of each system's commonalities through overlapping neural systems. Neuroanatomically, areas of the prefrontal cortex and the parietal lobe used for WM and reasoning overlap, while the same areas in the superior part of the prefrontal cortex and the parietal cortex underlie development of both visuospatial WM capacity and performance on the Stroop task (response inhibition). Furthermore, Oleson, Westerberg, and Klingberg (2004) supported this explanation by finding that training of WM tasks increase brain activity in multimodal areas of both the prefrontal and parietal cortices. With this in mind, Klingberg et al. suggested that Cogmed training may benefit, as a method of cognitive therapy, individuals who have suffered from traumatic brain injury and stroke.

n-Back Task

Jaeggi et al. has not been the only team to extensively use an n-back paradigm to evaluate WM training effects. Redick, Shipstead, Harrison, Hicks, Fried, Hambrick, Kane, and Engle (2013) compared typically developing young adults on either a dual n-back task (training group), or adaptive visual search program (active placebo-control group) over a series of 20 training sessions. An additional no-contact control group was formed and completed no training. Redick et al., in alignment with aforementioned research, explained that far-transfer should only occur if the processes that are improved via WM training are the same processes that are shared between WM capacity and the target transfer construct. With the dual n-back task (Jaeggi et al., 2008),

there is evidence that the processes involved in successful dual n-back performance overlap with the processes needed to solve reasoning and fluid intelligence tests.

The dual n-back task requires participants to respond to the identity of aurally presented letters and the location of visually presented squares, with both types of stimuli presented simultaneously (Redick et al., 2013). Participants are instructed to decide whether the current stimuli (letter and/or square) match the ones that were present *n*-back, with *n* varying between 1 and 4 across experimental blocks for all trainees. The dual n-back task has been shown to correlate with tests of Gf at much higher rates than with other measures of WM (Jaeggi, Buschkuel, Perrig, & Meier, 2010; Redick et al., 2012). It is unclear as to why dual n-back success is weakly related to other performance measures of WM, but history has shown that far-transfer is possible in effects related to Gf.

Redick et al. aimed to critically examine the effectiveness of dual n-back training, as they confirmed that Jaeggi et al.'s (2008) results are so potentially important that replicating them across varying samples and laboratories is necessary, especially since the training of reliable near- and far-transfer effects has yet to be definitively produced throughout literature. Considering the practical application and benefits of increased intelligence following a relatively brief training regimen, Redick et al. developed the primary goal of this study to replicate these dual n-back training results by showing transfer to measures of Gf. In order to evaluate intelligence transfer, researchers administered 17 transfer measures to assess Gf, multitasking, WM, crystallized intelligence, and perceptual speed, instead of only using particular tasks like Raven's APM. For each of the aforementioned constructs, both verbal and nonverbal assessments were administered.

It is important, as Redick et al. noted, to provide an array of measurements to rule out explanations based on task-specific processes and evaluate the efficacy of dual n-back training at the level of these constructs. Additionally, Redick and his associates determined that the Gf measure used should not be comprised of a matrix reasoning task. This would ensure that any observed transfer effects are not a result of visuospatial materials present in both the training and transfer tasks. The purpose of providing a battery of transfer tasks is not due to the expectation that all tasks will provide evidence of transfer, but is to evaluate the assumption that, if Gf is truly improved through WM training, then tests showing the greatest demand of g (general intelligence) should show the most transfer, while tests with the lowest demands of g should show minimal transfer. This allows researchers to view Gf transfer on a spectrum across construct-focused tasks.

Redick et al. asserted that the best control groups are active, performing tasks that are as adaptive and challenging as the true training groups, but do not focus heavily on WM capacity. The aim of this design is to equate that motivations, beliefs, expectations, and efforts of the active-control group to the dual n-back (or other training paradigm) group, leaving the key difference between groups to be their WM capacities following training. To achieve this structural equality, while minimizing the WM efforts of the control group, Redick et al. chose an adaptive visual search task to serve as the placebo control. This allowed researchers to separate transfer effects due to improving WM through relevant training from pure placebo effects. Jaeggi et al. (2008) concluded that Gf transfer is dosage-dependent-a relationship that Redick et al. aimed to investigate. To do so, researchers administered transfer tasks as pre-, mid-, and post-tests. This allows for the dose-response relationship to be evaluated within-subjects, rather than between subjects.

Four possible transfer outcomes were assessed in this study. The first was that the processes trained through dual n-back training produce transfer effects to Gf improvement (whereas visual search practice does not). The second was that visual search would yield improvement similar to a no-contact group, with the training group yielding even greater transfer effects. This would indicate that either placebo effects occurred or that the visual search task did inevitably facilitate WM, thus Gf, improvements. The third was that the dual n-back and visual search training groups would increase Gf equivalently relative to the no-contact control group, suggesting that the processes being trained through the dual n-back regimen are not specifically responsible for improving intelligence. Additionally, this may suggest that cognitive training gains and subsequent transfer are all products of placebo responses. Finally, the fourth possible result was that no group would show differential improvement on the tasks, showing consistence with Jaeggi et al.'s (2008) study, but with a longer duration of testing (20 sessions vs. 8 sessions).

Redick et al. concluded with three primary findings: 1) Subjects improved with practice on both the dual n-back and visual search tasks; 2) Training groups showed no transfer to any ability measures including Gf; 3) Dual n-back trainees reported subjective improvements in cognitive abilities despite the apparent lack of objective support for improvement. In addition, researchers were successful in creating an adaptive, active-control training regimen (visual search) that yielded the same amount of experimental contact as the experimental regimen (dual n-back), as well as similar scores of self-reported effort and enjoyment. Despite performance improvements, however, no positive transfer to any abilities provided evidence against Jaeggi et al's (2008) conclusion that training is dosage-dependent.

Redick et al. inferred that this lack of transfer from WM to Gf could have been observed because WM transfer effects are actually not commonly observed, as the file-drawer problem prevents WM training studies with nonsignificant results from being published, a notion further supported by Shipstead, Hicks, and Engle (2012). Redick et al. closed with a statement considering the amount of training improvement, specifically that of dual n-back training, to be only one variable that may affect transfer. Pretest ability level, sample size, the number and length of training sessions, transfer tests administered and their method of administration, trainee motivation, and the spacing between training sessions are all potential factors that may alter the prominence of training effects to a general array of cognitive abilities.

Complex Span Tasks

Redick, Kane, and Engle have worked extensively with Complex Span Tasks (CST) as WM measures, in addition to their work using n-back tasks as WM training paradigms. CST have been shown to trigger the dynamic WM system that involves both the storage and processing of information (Redick et al., 2012). This is in contrast to simple span tasks that measure the storage aspect of short-term memory capacity only. Kane et al. (2004) and Unsworth & Engle (2007) have both indicated that CST performance is strongly related to performance on measures of Gf. In addition, use of the Symmetry Span and Reading Span CSTs have shown that individual differences in WM capacity are domain-general, due to the observation that the storage components of these tasks account for similar variance in verbal and spatial ability tests (Redick et al. 2012). Automated versions of CSTs have been created for quick administration, automatic scoring, and computerized applications. Examples of these automated tasks were used in the aforementioned Redick et al. (2013) study using a dual n-back task as the primary training

task; however, few researchers have considered CST as primary training tasks rather than near-transfer effect measures.

On the surface, dual n-back and CSTs seem to tap into the same WM mechanisms; however, there seems to be little evidence for a relationship between these two tasks. In Redick et al.'s (2013) study, n-back training did not transfer to improvements in CST performance. This suggests that WM training is task-specific, such that improvement in one type of WM measure does not lead to improved performance in another category of WM task (Redick & Lindsey, 2013). This suggests that these two tasks do not tap into the same mechanisms, as it would be assumed that n-back practice would increase performance in the same mechanisms found in CSTs. Additionally, Redick and Lindsey stated that, while dual n-back training is capable of identifying the neural substrates of WM (dual n-back tasks are much more easily and practically completed in neuroimaging machines), CSTs are better suited for identifying individual differences due to higher reliability. This conclusion asserts Kane, Conway, Miura, and Colflesh's (2007). suggestion that researchers should not discuss WM training results interchangeably when either dual n-back or CSTs are employed as training paradigms individually.

Future Directions for WM Training Research

Evidence has shown that, to some degree, WM training is effective in not only improving WM performance, but transferring those improvements to measures of Gf and other pertinent cognitive abilities. It is necessary that future research focuses on pinpointing the aspects of training paradigms, including the WM mechanisms targeted during training, that best produce transfer to Gf, in effort to create reliable transfer after each training session. It is important that

future research moves beyond the desire to show that broad change can be realized through a training period on a limited set of tasks, according to critics Shipstead, Hicks, and Engle (2012). Thus, researchers should implement paradigms that are based on specific aspects of WM and transfer effects that should logically follow such training

Researchers should begin by integrating active controls into their designs to more strongly rule out confounding variables. This active control should feature an experience that is closely matched to the training group, but include tasks that do not target WM mechanisms. Non-adaptive training variants, less intense forms of training, or a placebo training group may all serve as active controls (Morrison & Chein, 2011). When using an active control, forming a no-contact comparison group is also encouraged. There has also been little study, but much speculation, on the possibility of sustaining training effects through periodic maintenance training (Jaeggi et al., 2011). Klingberg et al. (2005) has suggested the use of everyday activities with high-WM load, such as mathematics and other demanding academic activities, be used in maintenance of WM effects, especially in classroom settings. Future studies should also determine whether performance gains are maintained over extensive periods of time, providing evidence that training produces plastic brain changes (Smith, Housen, Ruff, Kennison, Mahncke, et al., 2009).

Researchers should offer a self-report on cognitive performance through a post-training survey, as they are rarely administered in cognitive training studies but are essential in evaluating attitudes towards training procedures and effects (Brehmer, Westerberg, & Backman, 2012). The use of more formal measures targeting various subscales of perceived performance based on workload, such as the NASA-TLX scale (Baldwin & Reagan, 2009), would be encouraged for extensive evaluation of attitudes towards task performance and demands. When reporting data,

Redick and Webster (2014) strongly recommend presenting the pre- and post-test values for transfer tasks, instead of only reporting pre- to posttest change scores. This allows researchers and critics to determine if the pattern of significant results allows a strong conclusion or if there is ambiguity in the transfer results. Furthermore, researchers should provide full data for each participant, especially in small samples, as supplemental material for more closely analyzing training gain and transfer.

Finally, researchers should perform direct comparative studies between transfer effects following training on either a dual n-back or CST. To prevent further interchangeability among transfer results using these two task types, future comparisons would allow researchers to evaluate the separate mechanisms employed through training using these paradigms. In the proposed study, it would be hypothesized that the CST would succeed in producing far-transfer where the dual n-back has previously failed (Redick et al., 2013). CST tasks are rarely used as the sole training paradigm, so future study would aid in both evaluating the differing mechanisms targeted from those of dual n-back training, while also investigating CSTs independent success as a WM training paradigm. In general, if reliable transfer can be produced, researchers may move away the question of "Does cognitive training work?" and towards evaluations of conditions that provide optimal transfer, for whom training works best, and what underlying neural mechanisms are exercised through training. If researchers can answer these questions, WM training paradigms will progress to greatly and consistently benefit clinical, educational, and consumer populations alike.

References

Baddeley, A. D., & Hitch, G. Working memory. *The psychology of learning and motivation.* New York: Academic Press, 1974. Vol. 8.

Baldwin C. L., Reagan I. (2009). Individual differences in route-learning strategy and associated working memory resources. *Hum. Factors* 51, 368–377

Bedard A.C., Martinussen R., Ickowicz A., Tannock R. (2004), Methylphenidate improves visual-spatial memory in children with attentiondeficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 43: 260–268.

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.

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.

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*(5882), 1510–1512.

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

Holmes, J., & Gathercole, S. E. (2013). Taking working memory training from the laboratory into schools. *Educational Psychology*, 1-11.

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., 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., Jonides, J., & Shah, P. (2011). Short- And long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25), 10081-10086.

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

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

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.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (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. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317-324.

McNamara, D., & Scott, J. (2001). Working memory capacity and strategy use. *Memory & cognition, 29*(1), 10–17.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise And challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, 18(1), 46-60.

Olesen, P., Westerberg, H., Klingberg, T. (2004). Increased prefrontal and parietal brain activity after training of working memory. *Nat Neuroscience, 7*, 75–79.

Ranganath, C., Flegal, K., & Kelly, L. (2011). Can cognitive training improve episodic memory?. *Neuron*, 72(5), 688-691.

Redick, T.S., Broadway, J.M., Meier, M.E., Kuriakose, P.S., Unsworth, N., Kane, M.J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment, 28*, 164-171.

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

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142, 359-379.

Redick, T. S., & Webster, S. B. (2014). Videogame interventions and spatial ability interactions. *Frontiers in Human Neuroscience*, 8, 183.

Repovs G., Baddeley A. (2006) The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience* 139, 5–21.

Richmond, L. L., Morrison, A. B., Chein, J. M., & Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychology and Aging*, 26(4), 813-822.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective?. *Psychological Bulletin*, 138(4), 628-654.

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Working memory training remains a work in progress. *Journal of Applied Research in Memory and Cognition*, 1(3), 217-219.

Smith, G. E., Housen, P., Yaffe, K., Ruff, R., Kennison, R. F., Mahncke, H. W., et al. (2009).Results from the improvement in memory with plasticity-based adaptive cognitive training (IMPACT) study. *Journal of the American Geriatrics Society*, 57(4), 594-603.

Turley-Ames, K., & Whitfield, M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language*, 49(4), 446-468.

Unsworth, N., & Engle, R. W. (2007). 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, 1038–1066.

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: How to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(6), 1322-1337.