Far-transfer effects of working memory training on a novel problem solving task.

by

Sharon Chan
BSc, University of Toronto, 2011

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

in the Department of Psychology

University of Victoria

Supervisory Committee

Far-transfer effects of working memory training on a novel problem solving task

by

Sharon Chan
BSc, University of Toronto, 2011

Supervisory Committee

Dr. Ulrich Mueller, Department of Psychology
Supervisor

Dr. Michael Masson, Department of Psychology
Departmental Member

Abstract

Supervisory Committee
Dr. Ulrich Mueller, Department of Psychology
Supervisor
Dr. Mike Masson, Department of Psychology
Departmental Member

The goal of this study is to assess the far-transfer effects of strategy-based working memory (WM) training to a novel problem solving task. Far-transfer refers to the application of trained skills to an untrained situation and is especially important because it deals with the generalization of learning to novel contexts. However, previous working memory training studies have produced little evidence for far-transfer. In the current study, children were trained in two strategies, phonological rehearsal and semantic categorization. These strategies have been suggested to increase the efficiency in processing and encoding of information and are invoked to explain developmental increases in WM capacity. Sixteen 6-to 9-year-olds were randomly assigned to each of four training conditions: semantic and rehearsal training, semantic training only, rehearsal training only, and treated control group. The treated control group performed significantly worse on the problem solving task compared to the three training groups. Surprisingly, the treatment groups did not differ significantly from each other. There was no statistically significant difference in receiving combined training of both strategies compared to only one strategy and furthermore, neither strategy resulted in better performance compared to the other strategy. Future directions for WM training and the implications for cognitive interventions are discussed.

Table of Contents

Tables

Figures

Introduction

The goal of this study is to assess the far-transfer effects of strategy-based working memory (WM) training to a novel problem solving task. Far-transfer refers to the application of trained skills to an untrained situation and is especially important because it deals with the generalization of learning to novel contexts. In this thesis, I am interested in how training of carefully selected memory strategies will transfer to the much higher order process of problem solving. That is, to what extent will training in WM strategies improve children's performance in a problem solving task, given that WM is a component of problem solving? In exploring this question of far-transfer, I will first define WM and then review the existing evidence for transfer effects in training WM. Approaches to training and assessment of training have been as variable as their results. To navigate through this extensive literature, I first frame the discussion of training studies into two phases: design and assessment. I will first introduce the two main categories of training studies (core and strategy-based) and then discuss the general findings of training studies within the context of near- and far-transfer effects. Finally, I will describe the current study, which adopts a strategy-based approach to training and measures both near and far transfer effects. Fig. 1 provides a visual breakdown of the literature review.

Baddeley's Model of WM

Working Memory (WM) is a cognitive system with limited capacity for the temporary storage and simultaneous processing and manipulation of information. Baddeley's model of working memory is one among many, although being one of the most influential, it will be adopted in this paper. The construct has traditionally been separated into two domain-specific systems including verbal and nonverbal (or visual) components. Training programs and WM tasks have often targeted processes belonging to toward one system or the other, and each is described in detail below.

Verbal WM. This first component of WM is responsible for phonological short term memory (PSTM), the ability for an individual to remember a small amount of verbal or acoustic information over a short period of time. Verbal information is held in the phonological store. In order to prevent rapid decay, the articulatory rehearsal mechanism constantly refreshes items verbal repetition. The structural limit on PSTM ranges between three and five items, but can be expanded by strategies such as chunking, where groups of items are stored as a single unit rather than individual items (Cowan, 2005).

Phonological short term memory is commonly measured by span tasks in which participants are presented a list of words, letters, or digits and asked to immediately repeat the list in the correct serial order (Henry, 2012). The list length is gradually increased to determine the longest list that can be recalled reliably. The number of items in the longest list is the corresponding 'memory span'. Span tasks have been useful in predicting other cognitive abilities, such as reasoning, memory, and comprehension (e.g., Ackerman, Beier, & Boyle, 2005; Kane, Hambrick, & Conway, 2005).

Nonverbal WM. This second, independent component of WM deals visuospatial short term memory (VSSTM), the ability for an individual to remember visual and spatial information over a short period of time. Visuospatial information is help in the "visuospatial sketchpad". More recently, there has been evidence to show a double dissociation between spatial memory (involving locations of items) and visual memory (shape/object features) within the visuospatial sketchpad (Darling, Sala, & Logie, 2007; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999). Separate tasks have been developed to measure spatial memory and visual memory.

The Corsi Block Tapping task is commonly used to measure spatial memory. It tests ability to remember a series of spatial locations presented in sequence. The experimenter points to a series of blocks, arranged in different spatial locations one at a time. The participant then points to the same blocks in the same sequential order as presented (Henry, 2012). The Visual Patterns Test (Della Sala, Gray, Baddeley, & Wilson, 1997), or Pattern Span, is another task designed to measure visual memory. This task involves pictures of partially filled in grids containing an even number of coloured and blank squares. After a short delay, the participant indicates which of the squares were previously filled in. The complexity of the grid is incrementally increased to reflect a visual span.

The central executive. A supervisory system known as the central executive is responsible for coordinating and regulating the activities of the two slave systems described above (Baddeley, 2000, 2007). Specifically, it deals with focusing, dividing, and switching attention between the phonological loop and visuospatial sketchpad. An additional component of the working memory model is the episodic buffer, described as a multimodal temporary store that is able to combine information from the slave systems, long-term memory, and perceptual input into a meaningful unit or ‘episode (Baddeley, 2000, 2007). Unlike the slave systems, the episodic buffer deals with information from different modalities (e.g. auditory, visual, spatial, or kinesthetic) and helps organize such information into meaningful chunks. It also allows for a link between long term memory and information being processed in working memory. For example, the episodic buffer links phonological information to long term memory to make sense of sentences, or links numbers to long term memory to recognize familiar phone numbers (Henry, 2012). The central executive and episodic buffer are important because in everyday life situations are seldom limited to only one domain of WM. In fact, WM has a vital role in a great number of pragmatic outcomes.

The importance of WM

WM forms the basis for many other higher order cognitive processes; including executive functioning (EF). EF is an umbrella term for a set of mental processes involving the separable but related components of working memory, attentional flexibility, and inhibitory control. EF manages higher order processes such as reasoning, planning and the execution of plans, and problem solving (Miyake & Friedman, 2012). Given our understanding of EF, working memory is an important component of problem solving, a term here defined as the generation of adaptive, goal-directed behaviours in novel or ambiguous situations, often in the face of interference and conflicting stimuli (Zelazo & Müller, 2011). WM has also been linked to many outcomes that fall under the broad definition of problem solving including: arithmetic (Fuchs et al., 2005) and reading competence (Nevo & Breznitz, 2011), syllogistic reasoning (Gilhooly, Logie, & Wetherick, 1993), daily functioning (Cohen & Conway, 2007) and school readiness (Müller, et al. 2008). On the other hand, deficits in WM have been implicated in developmental cognitive disorders (Melby-Lervåg, Lyster, & Hulme, 2012; for a review, see Gathercole & Alloway, Willis, & Adams, 2006). Given the relevance of WM, there has been prolific research in conducting training studies with the goal of improving working memory capacity (WMC). In the following section, previous findings on WM training literature will be discussed.

Previous research on WM training studies

Assessing the effectiveness of training. Cognitive training is a controversial topic in general. Research on WM training in particular has generated much skepticism with regard to the longevity and generalizability of training effects. There are two dimensions to address in the assessment of training. The first dimension is cross-contextual transfer, or the generalizability of training effects to new and untrained situations. The second component deals with temporal transfer, or the maintenance of training effects over time (i.e. at long term follow ups). Although literature has supported reliable-short-term gains in training (Melby-Lervåg & Hulme, 2012), the evidence for long-term gains has been limited. These two dimensions of training are not mutually exclusive but can occur in combination with each other. For example, a far-transfer effect can be long term or short term in its duration, as can a near-transfer effect.

The cross-contextual dimension. Cross-contextual transfer is highly relevant to our understanding of how training effects can be transferred to real world environments. Although both cross-temporal and cross-contextual considerations are important, the focus of this thesis is on cross-contextual transfer. Producing a sustained change in behavior is an ideal outcome of training, but the duration of change may sometimes be secondary to the nature of change itself. Training which produces improvements in only a narrowly constrained set of outcomes may not be viewed as successful as training which produces improvement across multiple outcomes, for example in a higher order cognitive process or in domains that are different but related to the trained construct. Alternatively, one might ask to what extent training can facilitate insight learning or help children think in qualitatively different ways, or generalize to situations with greater cognitive demands. At minimum, we expect training to transfer to tasks outside the initial trained task. This leads us to make the distinction between near-transfer and far-transfer effects.

Near-transfer training effects are only seen in tasks very similar to the initial training task. Training with different span tasks is an example of near-transfer effects. In span tasks, participants are presented a list of words, letters, or digits and asked to immediately repeat the list in the correct serial order. In one study, participants who were trained to rehearse a series of images of concrete nouns also demonstrated improved digit and letter memory (Comblain, 1994). Results suggest that training of the rehearsal strategy may transfer across contexts provided the stimuli can be verbally encoded.

While there is some cross-contextual transfer in that memory performance on untrained stimuli improves in addition the trained stimuli, far-transfer effects would go beyond the limited transfer seen in this example. Specifically, for far transfer to be the case we might expect for example, that training on word spans improves retention of verbal information after reading passages, not merely word lists. In other words, in far-transfer we would expect to see training effects in post-tests that are related to the construct trained in the initial task, but which are also qualitatively different or more complex. A broader example of far-transfer is that one might train a working memory strategy and measure performance on a problem solving task that has a working memory component. Far-transfer effects are arguably more meaningful because they show application of trained skills to situations that are not only new but have inherently different cognitive demands from the trained task. In this way they are more likely to represent changes in core processes rather than simple practice effects. Thus, the assessment of training effects should ideally involve far-transfer tasks.

Near and far transfer effects in previous research. Reviews of training studies by different authors have arrived at dramatically different conclusions with regard to the generalization of training effects, perhaps coloured by a personal stance on the subject. The ideal far-transfer effect would be that training improves functioning in a pragmatic sense; in daily life, or for those whose baseline WM performance is below developmental norms. A review of training studies by Morrison and Chein (2011) suggested that cognitive training yielded broad cognitive benefits and could serve as a remediating intervention for individuals with low WMC who show impaired academic performance or functioning in everyday life. However, these authors admitted that their interpretation of the results was hampered by the sheer variability that existed across studies and that there was no systematic method of comparing effectiveness of training across studies.

A methodologically more rigorous meta-analysis of 200 WM training studies was conducted by Melby-Lervåg & Hulme (2012). Results were systematically compared in terms of near-and far-transfer effects. The authors concluded that WM training did not produce gains in far-transfer tasks even when assessments took place immediately after training, a finding in stark contrast to that reported by Morrison and Chein (2011). Other researchers have also concluded that generalization of skills to novel contexts is notoriously poor when use proper controls and randomization (Wass, Scerif, & Johnson, 2012). For example, in some studies measuring far-transfer training effects of nonverbal ability (e.g., Jaeggi, Bushkuehl, Jonides, & Shah, 2011; Nutley, Soderqvist, Bryde, Thorell, Humphreys, & Kingberg, 2011), significant training effects could not be replicated when treated control groups were used as opposed to untreated controls (Melby-Lervåg & Hulme, 2012). Overall, the results of their meta-analysis suggest that there is very poor evidence for far-transfer effects even when assessments take place immediately after training.

This concern over the use of treated control groups in far-transfer training studies has also been voiced by other researchers. Shipstead, Redick, & Engle (2010) has suggested that, "as of yet, results are inconsistent and likely to be driven by inadequate controls and ineffective measurement of the cognitive variables of interest" (p. 245). It is apparent that future training studies must strive to include proper treated control groups and use random assignment for validly interpretable results.

The temporal dimension. The second dimension in evaluating training effects is time. Performance in near- or far-transfer tasks can be assessed immediately after training or after a delay. Additionally, performance in a near- or far-transfer task can be measured at long-term follow ups to determine the extent of remaining effects at various time points after initial training. It is also conceivable to imagine more than one time point of training. After all, how much learning can be internalized after only one session of training? An analogy that every student is familiar with is that paced studying where information is consolidated over several spread-out time points is more effective in retaining information than cramming for an exam (Romano, et al., 2005). Training may have similar limitations depending on whether it is administered in a distributed or massed way. An unrealistic expectation that has been adopted too often in training research, has been to expect long-term-far transfer effects after intensive bursts of training. Nevertheless, one can begin to appreciate the complexity and multidimensionality of training, with a spectrum of outcomes ranging from immediate-near-transfer effects to long-term-far-transfer effects and everything in between.

Short and long term effects in previous research. The meta-analysis by Melby-Lervåg & Hulme (2012) also systematically compared training effects in terms of immediate and long-term transfer effects. Three categories of training effects were evaluated for their temporal transference: verbal, visuospatial, and attentional shifting (i.e., between two conflicting channels of information where one channel must be inhibited while the other is acted upon). Results of the meta-analysis showed that short-term improvements in verbal WM were not sustained after a delay averaging 9 months. A modest training effect was found for visuospatial WM that remained for an average of 5 months, but this finding was limited by insufficient power due to small sample size of studies in this area. Lastly, a small to moderate training effect was found for attentional shifting immediately after training, but this effect was reduced to zero at long-term follow-up. Thus, with the use of more stringent evaluation criteria, long-term training effects evaporated.

Methods of training. Careful assessment of training is undoubtedly useful in capturing an honest picture of training effectiveness. However, assessment comes after the fact, and while it can contribute to our hindsight knowledge of training methods, a discussion on the approach to training deserves a section of its own. If training is not as effective as desired, one must consider what aspects of training itself needs to be changed. A primary factor to consider is the design of training. In general, cognitive training adopts one of two approaches: core-based and strategy-based.

Core based training. Core training methods are designed to target domain-general WM mechanisms. Such training would not be associated with a particular type of information or sensory modality, but would aid in the overall encoding, maintenance, and retrieval of information. The processes targeted would involve control and allocation of attention, monitoring what information is processed by the central executive, reducing interference from irrelevant sources of information, and controlling the activation of domain specific strategies (Morrison & Chein, 2011).

A number of criteria have been proposed to characterize core training programs to ensure that domain-general processes are elicited. The training task should 1) limit the use of domain-specific strategies, 2) be less susceptible to automatization, 3) target multiple modalities, 4) require the maintenance of information in the face of interference, 5) reinforce rapid encoding and retrieval demands, 6) adapt to participants' varying levels of proficiency, and 7) evoke a high cognitive workload (Morrison & Chein, 2011). A core training paradigm that meets all these criteria will necessarily be complex and pose a considerable challenge to implement. More importantly, a major drawback of using such a multifaceted training task is the difficulty in pinpointing which specific aspects of training are effective, and which aspects of training affect specific components of WM. Arguably, this is a profound limitation because it means that at most we would be able to describe a change, but not explain the mechanisms behind that change or the specific processes at work. Without knowledge of the specific mechanisms, we would not be able to explain why some components of training are more effective than others, or fine tune training for different individuals.

Core training methods are also susceptible to the fallacious assumption that improvements in performance indicate improvements in core processes. Since core-based training tasks are designed to elicit core processes, one might assume that training would far-transfer to untrained tasks that meet the same core criteria. However, improved performance may still only be restricted to the trained task unless this can be proven otherwise through improved performance in an untrained task. Although core-based training tasks are designed to "avoid automatization", practice effects are inevitable and true changes in core processes must be teased apart by having different pre- and post- training measures. However, if both the trained task and the post-training task must be core tasks, and both tasks must be different, there is the risk that both tasks are not validly measuring the same construct. Second, it is difficult to gauge how different to make the post-test task while still validly measuring the same construct that is tapped in by the pre-test task. To circumvent this issue of administering two different core-based tasks, one might train specific strategies and measure transfer to tasks that measure core-based processes at post-test.

Strategy-based training. An alternative to core training programs is domain-specific or strategy-based training. This approach focuses on training specific strategies that help encode, maintain and retrieve specific information over a delay period (Morrison & Chein, 2011). Participants are explicitly taught to use the strategy of interest and then encouraged to use and refine their mastery of specific skills in practice. Two types of strategies have received notable attention in the literature: those that increase reliance on, and facility with, articulatory rehearsal (e.g., Turley-Ames & Whitfield, 2003), and those that train elaborative encoding (e.g., Carretti, Borella, & De Beni, 2007; McNamara & Scott, 2001).

Elaborative coding involves making associative relationships between stimuli to aid recall. Subjects might be trained to group items into chunks, or to use mnemonic techniques such as devising mental stories and use of imagery to increase saliency of task relevant information. For example, in remembering a list of words, children might create a story between the words on the list to help them remember specific items or to group items that are associated with each other (e.g., cat-dog). Using mnemonics, an association is made between task-relevant information and information already held in long-term memory to facilitate retrieval. As previously mentioned, articulatory rehearsal involves rehearsing items out loud or silently in order to maintain and refresh information in memory.

The relation between rehearsal and working memory capacity in particular has been well supported in the literature. Morrison & Chein (2011) suggest that rehearsal improves WM performance by directing attention away from less effective strategies or by increasing the quality or efficiency of covert rehearsal mechanisms that support maintenance of information. The use and practice of rehearsal improves WM task performance in both developmentally delayed (Conners, Rosenquist, Arnett, Moore, & Hume, 2008) and typically developing children and adults (Ford, Pelham, & Ross, 1984; Ornstein & Naus, 1983).

According to Flavell, Beach, and Chinsky (1966), the increased usage of rehearsal throughout childhood corresponds with children's developmental increases in WMC. That is, as children age, they are more likely to use a rehearsal strategy to remember items, and they are likely to remember more items because they use rehearsal strategy. Flavell and colleagues propose that children undergo a progressive linguification of their world such that they increasingly apply language to more situations. The lack of this ability in younger children to flexibly use language explains their inability to apply language skills in a strategic sense, as is required in a rehearsal. The authors also suggest that children are unable to synthesize the two processes within rehearsal: the initial recoding of visual stimuli to verbal, followed by the rehearsal of recoded items. In any case, it is apparent that children's use of rehearsal strategy is closely related to their increasing WMC, possibly with language ability as a mediator.

As illustrated with the literature on rehearsal and WM, an advantage of strategy-based over core based training is that training effects have been found more consistently and these changes are more easily interpretable and measurable. Another motivation for adopting a strategy-based training paradigm comes from the evidence that children with higher WMC compared to their peers may in fact have higher WMC because of their patterns of strategy use. Although children with higher WMC may be benefitting from a combination of factors, it is clear that individual differences exist in children's selection and implementation of strategies, and these individual differences in strategy use account for significant variance on WM tasks (Dunlosky & Kane, 2007; Engle, Cantor, & Carullo, 1992; Friedman & Miyake, 2004; Kaakinen & Hyönä, 2007; McNamara & Scott, 2001; Turley-Ames &Whitfield, 2003).

A possible caveat with strategy-based training is that it lacks generalizability and is constrained to situations that are similar to the trained task. One may ask whether strategy-based training can produce far-transfer effects because the strategies being trained are so specific. However, strategy-based training may still be a better alternative to core training in producing far-transfer training effects, an idea expanded on below.

Bailey, Dunlosky and Kane (2008) highlight an important dilemma that concerns the cause and effect relation of strategy use and WMC. The strategy-as-cause hypothesis claims that strategy use is the direct cause of individuals obtaining, for example, higher span scores. Strategies allow a subject to retain more information, resulting in a higher span score. Alternatively, the strategy-as-effect hypothesis suggests that having higher WMC allows one to be more strategic, which in turn contributes to better performance on WM tasks. There is some agreement that strategy use may be both cause and effect, but whereas some researchers favour the strategy-as-effect hypothesis (see Dunlosky & Kane, 2007 for review), I believe there is a valid argument to revisit the idea of strategy-as-cause, particularly in the case of training children who do not use mature strategies.

My argument is that certain strategies are reflective of important conceptual shifts in development and must be incorporated into a “cognitive toolkit”, preceding permanent changes in WMC. This feedback loop between strategy use and WM is particularly important in childhood because certain conceptual shifts may allow for increased efficiency in processing and encoding. These shifts must occur and, by extension, certain strategies would be incorporated into the intuitive cognitive toolkit before permanent changes in WMC can occur. One might additionally theorize that select combinations of strategies trained together would be more potent. It might also be useful to consider whether certain characteristics of the trained group make them more or less suitable for certain strategies. In other words, we might consider what interindividual differences are important in determining which strategies are more or less effective for those individuals.

It follows that if one could design a specially tailored strategy-based training program that is strongly theory driven, it may be possible to produce far-transfer training effects as measured in a core-based post-test. Thus a strategy-based training program could be a better alternative to a core-based training program by circumventing the issue of designing two complicated core tasks that fulfill the opposing needs of being similar enough to expect transfer, but different enough to ascertain “far” transfer. Instead, the trained strategies could be embedded within the post-test so that use of the strategies improves overall performance under conditions of additional cognitive demands and interference. If training effects are maintained despite additional demands and interference, they will be more likely to endure in other situations and in the real world. Finally, the post-test itself could meet the criteria for a core-based task as outlined earlier in this paper, thus giving us greater confidence that strategy-based training can result in improvements in core processes.

A theory on good strategies. A review of the literature suggested that particular strategies produced more robust or consistent improvements in performance. WM has two components, storage and simultaneous processing or manipulation of information. In my view, successful strategies are able to adaptively limit the cognitive resources allocated to one component of working memory, the processing component, in order to optimize the function of the other component, storage. The less effortful the processing/encoding, the more cognitive resources are freed up for storage. In examining the relations between task demands and storage in tests of WMC, Towse & Hitch (1995) found that each involved qualitatively different attentional loads. In a similar vein, I propose that three pivotal cognitive shifts occur in childhood that increase the efficiency in processing and encoding of information and that may explain developmental increases in WMC.

Developmental shifts in WM processing and encoding. The first is a shift from unorganized to increasingly organized processing of information. The second is a shift from nonverbal to predominantly verbal encoding of information. Lastly, there is a change in consolidation mechanisms from passive maintenance to active refreshing. Whether or not a child has undergone these shifts may explain individual differences in WM performance during development.

1. Unorganized to organized. Converging lines of evidence support the idea that organization of information aids in WM performance during both retrieval and encoding processes (Schleepen & Jonkman, 2012; Shiffrin & Atkinson, 1969; Tulving, 1962). Grouping and organization of information facilitates delayed retrieval in particular (Lange, Guttentag, & Nida, 1990) and consolidation into long-term memory (Schleepen & Jonkman, 2012; Tulving, 1962). However, individual differences in the ability to intentionally and spontaneously apply organizational strategies is strongly associated with working memory capacity, which in turn is determined in large part by genetics and viewed to be stable within the individual (Schleepen & Jonkman, 2012; Bailey, Dunlosky, & Kane, 2008).

Interestingly, a recent study suggests that not all grouping strategies are equal: categorization based on semantic features has been shown to improve memory performance more than those based on perceptual features or personal associations (Schelble, Therriault, & Miller, 2012). In this study, semantic categorization (e.g., think about all types of dogs, or all types of mammals) was found to be more effective in improving performance compared to other strategies such as groupings made on the basis of animals' perceptual features, or the use of free associations and personal associations (e.g., animals that the child liked, disliked, feared).

Children with higher WMC were also found to use the classification strategy more often independently and spontaneously. This relation between semantic categorization and higher WMC has been replicated in other studies. For instance, children who effectively use semantic categorization have been reported to have superior WMC (McNamara and Scott, 2001), and older children's ability to use the semantic categorization is mediated by their WMC (Schleepen & Jonkman, 2012). This relation between WMC and semantic categorization has also been found for adults, with higher WMC individuals being more likely to use a semantic grouping strategy compared to lower WMC adults in verbal fluency tasks (Rosen & Engle, 1997). The direction of cause and effect between these two variables is less clear. Do children who use semantic categorization develop higher WMC? Or do children with higher WMC selectively choose to use this strategy? Or perhaps, a more reciprocal relation holds such that development in one scaffolds development in the other.

Perhaps the most important finding in these studies is that the use of semantic classification mediated the effect on the relation between WM and retrieval performance (Schelble, Theerriault, & Miller, 2012). When participants used this strategy, it made a stronger contribution in predicting retrieval performance than did children's individual WMC. Furthermore, although children with higher WMC have been found to be more strategic than children with lower WMC in free recall tasks, presenting participants with retrieval cues which prompted better strategy selection eliminated this difference (Unsworth, Brewer, & Spillers, 2013). These finding suggests that effective organizational strategies such as semantic categorization could compensate for lower WMC in a demanding retrieval task.

2. Non-verbal to verbal encoding. Unlike visual memory, rehearsal is particularly useful when there is a delay between the presentation of information and recall, as information can be maintained and refreshed in the phonological loop. Despite its usefulness, younger children do not spontaneously use this strategy, instead relying on visual WM. The use of rehearsal is contingent on information being preferentially recoded from visual information to verbal information. Younger children who rely instead on visual short term memory, do not spontaneously engage in this recoding process. A study involving children between the ages of 4-11 found that only children around the age of 7 or older could effectively apply phonological strategies when committing pictures to short term memory, and that both older children and those with higher WMC tended to use verbal strategies more frequently (Sanefuji, et al., 2011). Another study found that children younger than 5 years do not show any evidence of phonological coding, and 4 year olds do not demonstrate any strategy use at all when administered a picture span task (Henry, Messer, Luger-Klein, & Crane, 2012). Rehearsal develops gradually starting from about 5-6 years of age but adult-like performance/utilization of phonological rehearsal is not apparent until about 6-8 years of age as noted by phonological similarity and articulatory suppression effects occurring only in older but not younger children (Henry, Messer, Luger-Klein, & Crane, 2012; Lehmann & Hasselhorn, 2007; Henry, Turner, Smith and Leather, 2000).

The phonological similarity occurs when serially presented sequences of phonologically similar items are harder to recall than dissimilar sequences (Conrad, Baddeley & Hull; 1964) due to the constituent items having similar and hence confusable acoustic codes within the phonological store (Salame & Baddeley, 1986). For example, trying to remember words that share acoustically confusable onset (e.g. <u>mo</u>ss, <u>ma</u>ll, <u>mo</u>p, <u>mo</u>th, <u>mo</u>ck) or coda (e.g., ba<u>ck</u>, bag;

bat, bad. The articulatory suppression effect occurs when a subject is required to continually repeat an irrelevant word of phrase during an immediate verbal working memory task, and performance in the latter is impaired since the target stimuli cannot be rehearsed in the articulatory loop; it is suppressed since the articulatory system is preoccupied with the distracting utterances. Both phenomenon can be used to illustrate the lack of rehearsal use in children because only when an individual is relying on a rehearsal mechanism to maintain items in memory, do phonological similarity and interfering utterances create impairments in recall performance. The lack of such impairments suggests that children are not actively using rehearsal and are thus not impeded when other processes tap into the phonological store. Thus, a "profound change in the use of verbal mediation" has been suggested to take place after 5 years of age (Al-Namlah et al., 2006, p. 117). As a result, existing research on this strategy has focused on adults or children who are old enough to use verbal encoding spontaneously and independently.

One theory explaining younger children's difficulty with using phonological strategy is that the recoding process is cognitively effortful and taxing on younger children's lower WMC (Sanefuji, et al., 2011). As WMC improves with age, older children become more efficient in the recoding process and the use of the phonological strategy becomes less taxing. Research examining children's use of rehearsal as a memory strategy has shown it to be particularly beneficial for children with lower WM spans. When comparing children's usage of rehearsal, imagery, and semantic strategies, it was found that children with lower WM span benefitted more by using a rehearsal strategy compared to children with higher WM span (Turley-Ames & Whitfield, 2003).

3. Passive maintenance to active refreshing. A developmental shift also occurs around 7 years of age from passive maintenance to active refreshing of information in WM. According to the task-switching model (Hitch, Towse, & Hutton, 2001; Towse & Hitch, 1995), younger children fail to implement maintenance activities while performing a concurrent task, resulting in time-based decay of the memory trace (Camos & Barrouillet, 2011). Instead, they passively hold items in memory without any attempt at active maintenance. Thus, their ability to hold items in memory is greatly affected by the duration of the delay period between presentation and recall. Older children have an increased capacity to control attention and monitor cognitive processes, allowing them to allocate attention during processing to reactivate, or refresh memory traces in real time (Camos & Barrouillet, 2011).

Current study

Current study: what is being trained?

To summarize the literature reviewed above, there are three cognitive developmental shifts which lead to more efficient processing and encoding of information, and which may explain both intra- (between) and inter- (within) individual differences in developing WMC. These include: 1) a change from unorganized to increasingly organized processing, 2) a change from nonverbal to predominantly verbal encoding, 3) and a change in consolidation mechanisms from passive maintenance to active refreshing. This study will attempt to train children in the use of specific strategies that facilitate the use and practice of the first two cognitive shifts. First, children will be trained to use semantic categorization to facilitate unorganized to organized processing. Second, children will be trained to use phonological recoding and rehearsal strategies to facilitate nonverbal-to-verbal encoding.

With regard to the third shift of passive maintenance to active refreshing, it would also be possible to ask whether or not a child could be trained to self-monitor and use active maintenance. Aside from articulatory rehearsal, verbal material can be maintained by another mechanism known as attentional refreshing or active maintenance (Barrouillet, Bernardin, & Camos, 2004). According to this view by Camos, Mora, & Oberauer (2011), attention maintains memory traces in an active state, because directing attention to a target strengthens the target's representation in working memory. When the trace of an item is activated in memory, its activation is maintained as long as the item is within the focus of attention. This trace beings to fade as soon as the focus of attention is switched, but prior to the trace being lost, it can be reactivated by focusing attention on it again.

However, this may not be necessary to examine in the current study. Although verbal rehearsal and active refreshing both contribute to the maintenance of verbal information in short term memory, adults have been shown to adaptively choose one technique over the other in certain situations. In one study comparing the use of these two strategies (Camos, Mora, & Oberauer, 2011), it was found that when adults were asked to remember "phonologically confusable" material, they preferred to use active refreshing to lessen the impact of phonological characteristics (a phonological similarity effect), whereas when the task required greater attentional demand, they preferred the less attention-demanding mechanism of rehearsal. In this study, efforts were made to avoid using stimuli that could evoke a phonological similarity effect. The problem solving task that children will be tested on will also require greater attentional demand. Furthermore, active maintenance involves the conscious monitoring and refocusing of attention on the task, a requirement that may be difficult, if not impossible to meet for children who are still developing metacognitive abilities. Therefore, training efforts will focus on improving verbal encoding and rehearsal as opposed to active refreshing.

Interestingly, the literature suggests that rehearsal and semantic categorization may have overlapping developmental trajectories in childhood. This idea comes from the observation that both strategies appear and are mastered around the same age in childhood. Semantic categorization is a relatively late-occurring memory strategy with its use being absent in 6-7 year olds and not approaching adult-like performance until the ages of 8–12 years (Schleepen & Jonkman, 2012). Similarly, rehearsal develops gradually starting from about 5-6 years of age and adult-like performance is not apparent until about 6-8 years of age as previously mentioned. Although there is some overlap in the timing of their development, it appears that semantic categorization emerges slightly later in development as compared to rehearsal. It is possible that the verbal encoding processes described in rehearsal may precede the ability to use other verbal strategies, and specifically bootstrap the ability to use a semantic strategy. For example, being able to characterize objects by abstract properties (e.g. having legs, being a tetrapod or bipedal creature, and being terrestrial) may be facilitated by storing those properties verbally in long term memory, particularly when they are hard to visualize (e.g. an animal that barks).

Considering the overlap in middle childhood regarding the emergence of these strategies, and the idea that one may bootstrap the other, one might ask whether training both strategies at the same time may be more potent than training only one strategy. More importantly, can combined training of these two strategies, which have been shown to be independently useful in increasing WMC result in a far-transfer effect and improve performance in a problem solving task? I believe it would be reasonable to expect such a far-transfer given that problem solving relies on executive functioning, and working memory forms the basis of EF. As previously mentioned, WM is also related to more specific types of problem solving such as arithmetic and syllogistic reasoning.

Current study: “who” is being trained?

The participants in this training study were carefully chosen with several considerations in mind. First, younger children demonstrate reliably better training effects than older children (Melby-Lervåg & Hulme, 2012). Furthermore, it has been recently suggested that cognitive training applied to younger children leads to significantly more widespread transfer of training effects and therefore efforts should concentrate on early intensive training (Wass, Scerif, & Johnson, 2012). The general consensus is that training might be more effective while neural systems are more plastic (Melby-Lervåg & Hulme, 2012; Wass, Scerif, & Johnson, 2012).

Additionally, the recruited age group reflected the period in which target strategies first emerge and are actively being used, but with varying levels of proficiency (6-9 years). Individual differences in children's baseline mastery of each strategy should be expected, not only between age groups, but especially within age groups. Therefore, age was not a variable of interest in this study efforts were made to include a range wide enough to accommodate the individual differences in development while still targeting the age bracket for which training would be most relevant. Rather than looking for age effects, which are already very clearly established for WM, I am interested to see whether performance differs according to training method. An age cut off of 6 was chosen as previous research has found that children at 6 years of age distinguish between solutions arrived by some form of reasoning as opposed to no reasoning (Amsterlaw, 2006). This metacognitive awareness forms the basis for children's ability to assess their own problem solving as good or bad, a prerequisite in knowing how to selectively choose and implement a more effective strategy over a previously used one.

Research questions and outline

Three main research questions were addressed in the current study. The primary question was whether strategy-based training was able to produce a far-transfer effect. To answer the first question, a study was designed to train the specific strategies of semantic organization (S) and rehearsal (R) under the premise that both S and R strategies would improve efficiency and processing in order to free up mental resources in WM for storage, with the result that WMC could be increased. By extension, increased WMC would produce improvements in problem solving performance.

Previous findings on S and R training have shown robust success in increasing WMC with older children and adults. However, these studies have typically examined their production in children who were already spontaneously using the strategies, and did not focus on facilitating their use in younger children who may not be spontaneously or independently using strategies, or using them to full extent. The current study aimed to do this through its target age group. Furthermore, while both strategies have been individually trained with success, no prior study has looked at the combined effects of training both against one, and in the context of a far-transfer post-test. Another possibility was that one strategy would be more effective than another. Specifically, I expected that a semantic strategy would be more useful to children in the far-transfer post-test than a rehearsal strategy based on the finding that children with higher WMC tend to use deeper encoding strategies as opposed to surface ones (Dunlosky & Kane, 2007; Friedman & Miyake, 2004).

To address the lack of evidence in extant training studies for a far-transfer effect, the current study measured the transfer effects of strategy-based training in a novel problem solving task that children had no prior experience with, and which had qualitatively different cognitive demands from the trained tasks. Performance on this problem solving game was expected to improve with the use of trained strategies, as task demands were embedded within the game, but children had to choose to use these strategies on their own accord, and under conditions of increased cognitive demand and interference. The problem solving task incorporated several constraints to elicit planning behaviour, involved confusing visual stimuli to add interference, and gave no explicit instruction on what problem solving method to use. Therefore children were free to use or not use trained strategies as they found appropriate.

In addition, careful attention was placed on the use of appropriate control tasks in response to previous concerns over the use of untreated control groups. Careful consideration was given to include control tasks that corresponded to semantic and rehearsal training phases.

These control tasks were comparable in time, type of stimuli involved, and level of mental stimulation as the training tasks. All groups therefore spent approximately the same time interacting with the researcher throughout the session.

Secondly, in addition to measuring the far-transfer effect, the study will also attempt to replicate a near-transfer effect using the free recall task. The free recall will be administered twice to all groups, first as a pre-test, and secondly as a post-test following the problem solving game, at the end of the session. Different categories and stimuli will be used to test whether the clustering strategy transfers to untrained stimuli, but the task itself will remain essentially the same. Given that near-transfer effects are common in the WM training literature, a near-transfer effect is not only unsurprising, but to be expected. It is expected that semantic training should improve category clustering but this study will also compare the effects of combined training and rehearsal training on the free recall performance.

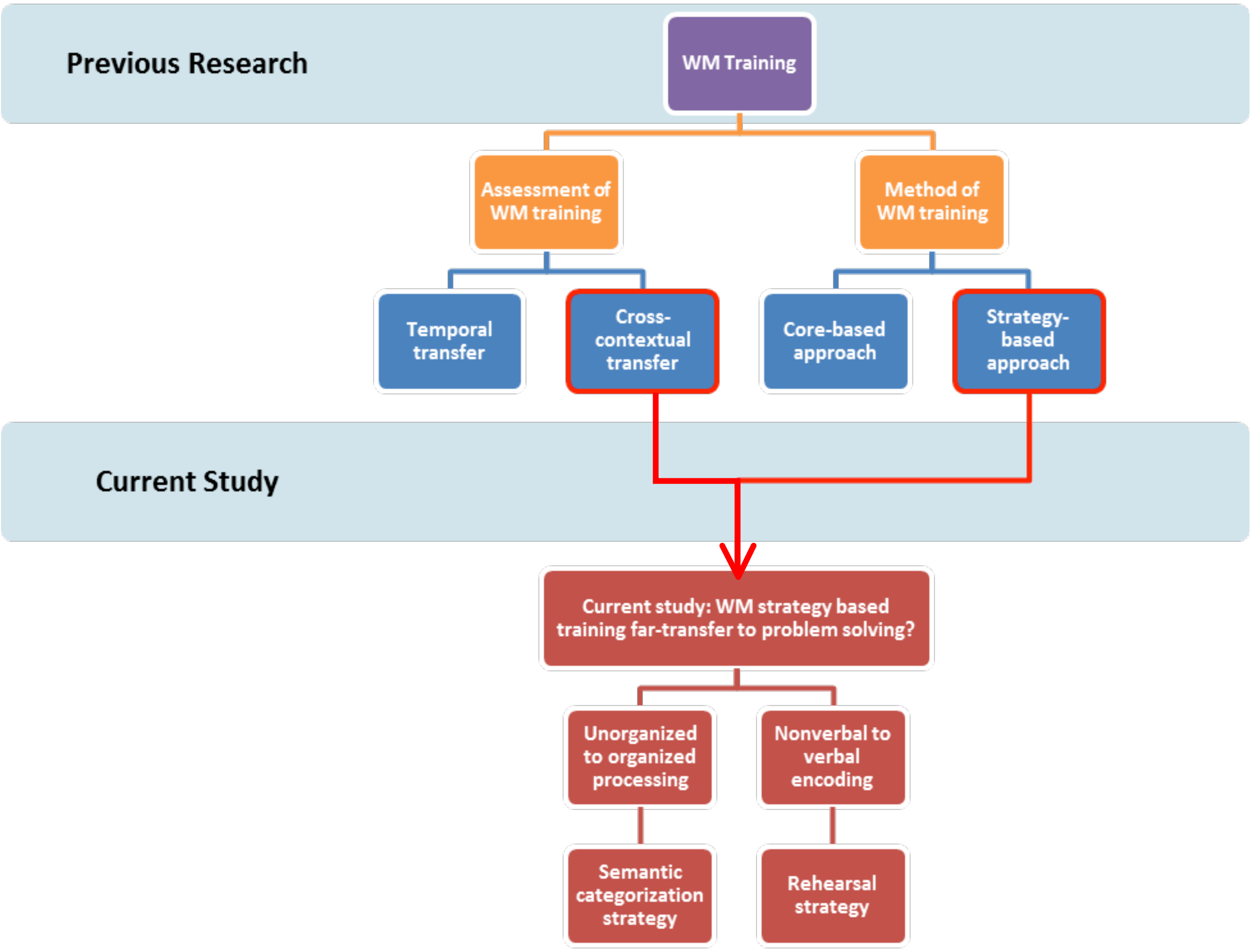
Third, with regard to individual differences in responsiveness to training, previous research has found that children with lower WM span benefitted more by a rehearsal strategy compared to children with higher WM span (Turley-Ames & Whitfield, 2003). Therefore another question for this study will be whether children who have lower baseline WMC benefit more from training than those with higher initial WMC. In order to test this question, pre-test measures will be used to differentiate low from high functioning WM groups post-hoc and differences in problem solving performance based on this factor will be analyzed.

Lastly, in addition to the above research questions, it will also be important to establish the construct validity of the novel problem solving task which was developed for this thesis. The task meets the criteria for a problem solving task, and particularly one that has high executive functioning demands because it 1) requires a working memory component 2) requires inhibitory control in selecting incorrect stimuli and in adaptively choosing to implement more effective strategies, and 3) requires flexibility in being able to alternate between strategies, and to apply them to new sets of stimuli. The task also meets several criteria of a core processing task, which will be noted in the discussion. If a transfer effect is found, it is meaningful to know that it has occurred under these conditions.

In terms of the duration of training, there has been no unified approach to the time invested and no evidence to look at the relation between time and quality of training. Previous training studies have all been time-intensive, involving multiple sessions. For example, one study found that training children with ADHD led to a gradual increase in performance on practiced and nonpracticed, visualspatial memory and reasoning tasks (Klingberg, Forssberg, & Westerberg, 2002). The training regimen involved approximately 25 minutes of daily practice on WM tasks for a duration of 24 days. Another study with children with low WMC yielded improvements in children's mathematical ability 6 months after training, but contrary to expectations, no improvement in verbal working memory capacity (Holmes, Gathercole, & Dunning, 2009). The training lasted for 35 minutes a day using a computerized WM training task for 20 days.

For the purpose of a Master's thesis I designed a training phrase that was comparatively shorter. Despite a short training session, particular attention was given to the structure of the training, including careful selection of strategies, targeting a specific developmental period when training should produce greatest impact, and combining strategies previously shown to be independently effective with older samples.

Fig 1. *Framework of thesis.*



Methods and Procedures

Participants

A power analysis using G*power 3.1 (http://www.gpower.hhu.de/en.html) was conducted to determine the minimum sample to find a medium effect size (0.25) for an interaction between the repeated measures factor and the between-subject factor using a mixed measures ANOVA, given $\alpha = 0.05$, four groups, and 3 repeated measures. The correlation among the repeated measures derived from preliminary analysis was 0.6. The minimum sample size suggested was 48 ($1-\beta = 0.95$).

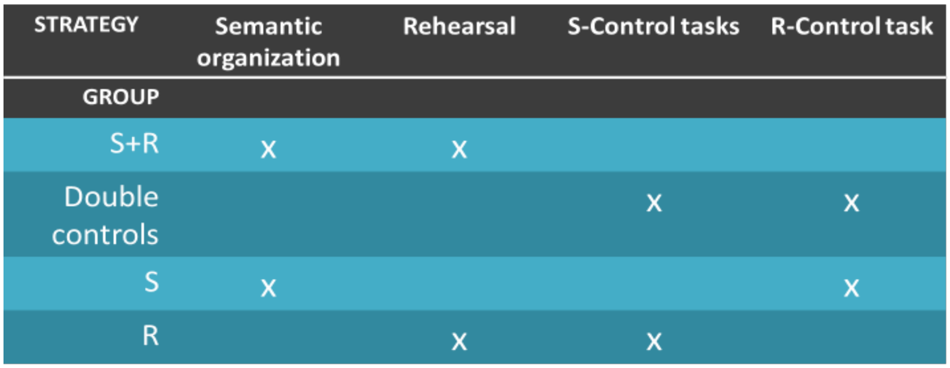
After data had been collected, the plan to analyze the data using repeated measures ANOVA was dropped and thus a second power analysis was conducted. The rationale for dropping the repeated measures ANOVA was that I was not interested in looking at change over trials, and did not have specific hypotheses about how performance in the problem solving task would change across trials. The purpose for having three trials was to more reliably capture children's performance at a single time point (e.g., much in the same way 3 trials were administered for the matrices task at any given span). In other words, the three trials were best represented as a single, cohesive task at one time point. However, the one-way ANOVA required a much larger sample size (n=230) given the same parameters of a medium effect size and 95% confidence interval. In my view, it was more important to select the statistical test that would more clearly answer my question of far-transfer, and given that this was not in essence a repeated measures design because I was not interested in change across different time points, nor was I interested in measuring a treatment effect over time, I decided to run a univariate ANOVA keeping in mind the limitation of small sample size.

Approximately 600 flyers were distributed to children in public and private schools within Victoria, B.C. and interested children and parents contacted the researcher for participation. The response rate from recruitment was approximately 10%. Sixty-five typically developing children aged 6-9 years agreed to participate and proceeded with the study. One child was excluded from the study because of inability to understand and follow the instructions. Sixty-four remaining children completed all the pretests, training phases, and post-test measures over a single 1.5 hour period.

There were four training conditions (see Table 1). The S+R group received training in both S and R strategy. Double controls received only the filler tasks (described below). The S group received the semantic training and the rehearsal control tasks. The R group received the rehearsal training and the semantic control tasks. The treated control groups received equal time and similar type of stimulation as the experimental groups to reduce other unintentional confounds associated with the training phase such as time spent interacting with the experimenter or fatigue.

Sixteen children were randomly assigned to each condition with the constraint that children in the different conditions had a similar mean age and gender distribution. There were no significant differences between the mean ages across groups (see Table 2).

Table 1. *Training conditions*.



STRATEGY GROUP	Semantic organization	Rehearsal	S-Control tasks	R-Control task
S+R	x	x		
Double controls			x	x
S	x			x
R		x	x	

Table 2

Participant demographics

Groups	*n*	$\bar{x}$ *age*	*s*	Male	Female
S+R	16	7.5 (0.97)	0.97	9	7
Controls	16	7.38 (0.96)	0.96	7	9
S	16	7.19 (0.91)	0.91	7	9
R	16	7.44 (0.96)	0.96	9	7
Total	64	7.38 (0.93)	0.93	32	32

Measures and tasks

A. Pretests

1. Digit-forward and backward span tasks. In the digit-forward span, children were asked to recall lists of digits (Henry, 2012). The child was asked to immediately repeat the list in the correct serial order. In the backward-span, children were asked to recall the list in order backward from presentation (last presented to first presented). Sequences were read at a rate of one number per second, and the task was discontinued after two consecutive scores of 0.

A score of 0 or 1 was given for each trial and points were summed for a total score. Longest Digit Span Forward/Backward was also recorded as the number of digits recalled on the last trial correctly answered.

2. Visual short term memory (VSTM) task. (Logie & Pearson, 1997; recall version).

The child was presented with a matrix pattern drawn on white cards in which half the squares, chosen randomly, were coloured red. The pattern was displayed for 2 seconds and then removed, followed by a further 2 second delay where the child was shown a blank white card. The child was then given an empty version of the same matrix and asked to point to which squares were previously coloured red.

Trials began with a four-square pattern (2-span) including two red squares and two white squares. Three trials of the 2-span patterns were administered. If at least two out of three patterns were correctly identified, the number of squares in the pattern was increased to show three red and three white squares (3-span). The child was then given three trials with these six-square patterns and successful identification on two out of three trials lead to an increase to four red and four white squares, and so on. The task was discontinued when the child failed to replicate two out of three patterns correctly for a given span (see Appendix A p. 83 for instructions and Appendix B p. 89 for stimuli).

An average score for matrices was calculated by taking the sum of the number of red squares in the three most complex patterns correctly recalled, and divided by three. For example, if a child successfully recalled all red squares for all three trials in the 10-square patterns, but was successful on only one trial in the 12-square patterns, then the span score would be (5+5+6)/3 = 16/3 = 5.33.

3. Free recall pretest. The method was adapted from Black and Rollins (1982) who presented items to be recalled as pictures on cards (see Appendix A p. 83 for instructions and Appendix B 90 for stimuli). Free recall tasks are traditionally administered to adults in written lists but given that some children would not be actively using verbal strategies and had limited reading skills, pictures were used instead. Coloured photos were used instead of black and white line drawings to provide a more realistic and ecologically valid representation of objects (Moreno-Martinez & Montoro, 2012). Four categories were chosen for use in the free recall task, for a total of twenty items (insects, fruits, vehicles, and furniture). Items did not reoccur across different phases of the study (e.g., different items were used in the post-test and the free recall post-test). High frequency items previously used with children from each category were preferentially selected (Rabinowitz, 1984; Snodgrass & Vanderwart, 1980). Children were given as long as they needed to study the cards. They were given up to 3 minutes for recall; during recall they were prompted once, “Can you remember anymore?”

Two performance indices were recorded for this task: the number of correctly retrieved cards and the adjusted ratio of clustering (ARC) (Gerjuoy & Spitz, 1966; Roenker, Thompson, & Brown, 1971), a measure which quantified the extent of category clustering. ARC was obtained through a category clustering calculator developed by Senkova and Otani (2012) using only correct cards. Items that children listed twice or more during the recall were scored as repeats. As repeated items could not be technically classified as 'errors' and the Senkova and Otani (2012) calculator did not incorporate repetitions, these items were excluded from analyses. Intrusions (i.e., recalled items that belonged to one of the target categories but which were not on the list to be recalled) were rare and also excluded.

B. Training tasks

The goal of training was to increase children's familiarity and hone their correct use of strategies. Therefore, accuracy was not recorded or analyzed for the training tasks, another major reason being that in most of the tasks (e.g., category inclusion decisions, rehearsal training), children were encouraged to make attempts until they arrived at the correct answer. For example, in the category inclusion training task in which children had to make decisions about which items did or did not belong with others in group, some children had trouble arriving at the answer that an "ant" did not belong in a group of mixed animate and inanimate objects because it "could not fly". This response was met with the prompt, "think of what these objects can do" to help children arrive at the correct answer on their own.

1. Semantic Categorization training tasks. These tasks encouraged children to organize information on the basis of their common abstract properties. Children were first trained to think in terms of categories, and then to apply them strategically. Training involved two phases. In phase A, Category inclusion decisions, children had to make decisions about which object in a group did not belong with the others. For example as depicted in Figure 2, they were shown a butterfly, beetle, spider, and banana. Children were asked, "Which does not belong?" followed by, “What do the other ones have in common?” to which the correct answer would have been, in this case, the banana, because it was not a bug/insect.

In phase B, a trained free recall was administered. Using the same cards from the free recall pretest, children were first prompted to sort the cards by their categories (fruit, insects, furniture, vehicles). Next, they were explicitly instructed to use the strategy to focus on similarity among items (“If you study the cards that are similar together, such as all the fruit together, it will be easier to remember them”). This general explanation method was found to be most effective in encouraging children to adopt the category clustering strategy, as reported by Black and Rollins (1982). Children were again given 3 minutes for recall.

Figure 2. *Example of a category inclusion trial administered to child.*

(Correct answer: banana, the others are bugs/insects.)

2. Semantic Control tasks. In the first control task, children were given a regular deck of playing cards and asked to find all the cards that shared an arbitrary colour or shape (e.g., all the red hearts, black clubs, etc.). In the second control task, the experimenter randomly selected a few cards from the free recall deck and children were asked to make up a story about the items.

3. Phonological rehearsal training. The goals of the R tasks were to train children in 1) recoding visual information into verbal information, and then 2) maintaining that verbal information by rehearsal. In the first phase of training children were asked to label pictures, and then asked to rehearse those words until they were ready to list them without referring to the pictures. In the second phase of training, word lists were presented orally only, and children had to recall items after a short delay (See Figure 3 for two phases of training; see Appendix C p. 91 for word lists).

Figure 3. *Rehearsal training.*

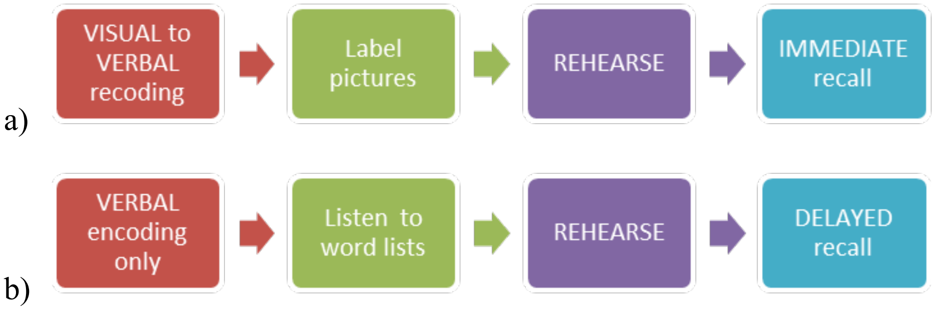


Figure 4. *Example of a 4-span from a-phase given to child.*

4. Rehearsal control tasks. Children were asked to read from a picture book with the help from the experimenter, or engage in a discussion about what they did on the weekend/upcoming activities at home or school for the duration of 10 minutes.

C. Post-tests

1. Problem solving game. The task involved three adjacent rooms, two troll houses where the child collected items, and bear's house. The goal of the game was to retrieve all the cards on bear's list (correct cards) while making as few errors as possible and using as few tokens as possible. (See Appendix A p. 85 for game instructions and Appendix D p. 92, 93, 94 for example trial and stimuli lists.) Each list had 24 target items, and there were a total of 48 cards. One troll housed all animate objects (terrestrial and aquatic animals) and the other troll housed all inanimate objects (school supplies, things to wear) although this was not told to the child. The child was first introduced to the bear, and then to each of the trolls. In the troll's rooms, they were shown the troll's money bank and box of cards.

While collecting the cards, several constraints were in place to elicit strategy use and problem solving. The first constraint was that the child had 6 tokens to pay the trolls. Each time a troll's box was opened, a coin had to be forfeited. Once all the coins had been used, the trial was terminated whether or not all cards had been collected. The second constraint was that children were only allowed to take away 7 cards or fewer at any given time. A conspicuous number 7 placed on the side of the basket, which was carried between rooms, served as a visible reminder of this rule at all times. This constraint ensured that children did not walk away with the entire deck of cards at once.

Children did not have access to the bear's list while in the Troll's rooms and had to remember which items to retrieve. Figure 6 shows the setup of the rooms. They were permitted to consult the list again when they returned to the bear's house to place their collected cards. None of the children were given any explicit instruction that they should use a particular strategy or that the items could be sorted into categories.

The rules of the game can be summarized as follows: 1) Only 7 cards were allowed in the basket at one time. The child was allowed to have 7 or fewer cards in the basket but not more. 2) Upon return to the bear's house, the cards had to be placed on the shelves in the same order as shown in the list. This rule was designed to help children keep track of which cards they still needed to get as it was helpful for them to leave a gap for the missing cards. During the collection, however, children could collect cards in any order they liked. The freedom in collecting cards in any order allowed children to make plans about the best way to collect as many cards as they could, ideally using a memory strategy to help them group cards together. 3) The child had 6 coins to finish the list. The trial ended when all coins were used regardless of how many cards had been collected. 4) The child was told that in order to win the game, all correct cards had to be retrieved while making as few mistakes as possible and using as few of the coins as possible.

For each trial, the following outcomes were noted: the number of errors, number of correct cards retrieved, number of tokens used, and the number of cards retrieved for each token. Total scores were computed as follows: total number of correct cards retrieved over three trials (/72), total number of errors over three trials, and total tokens used (/18). The highest median span for each trial was determined from the number of cards retrieved per token for each trial. Mean highest span was not used because mean was too susceptible to the number of 'clean up' trips children made. These clean up trips involved a child trying to fulfill a mental chunk by collecting the remaining one or two cards, and were not representative of their actual WMC. For example, if a child typically remembered 4-5 cards, and spent one trip collecting only one card, it was likely the child had wanted to cross check accuracy for a certain chunk before retrieving the last card for fear of making a mistake. A final performance index was calculated from the aggregate scores over three trials. This ratio was expressed as total errors subtracted from total correct cards (calculated first), divided by number of tokens (C-E)/T.

2. Post-training free recall. A post-training free recall task was administered to all groups after the game. The post-test used a different set of stimuli and categories (body parts, nature, instruments, kitchen utensils; see appendix D p. 93 for full list) from the pre-test.

Figure 5. *Room set up for problem solving game*.

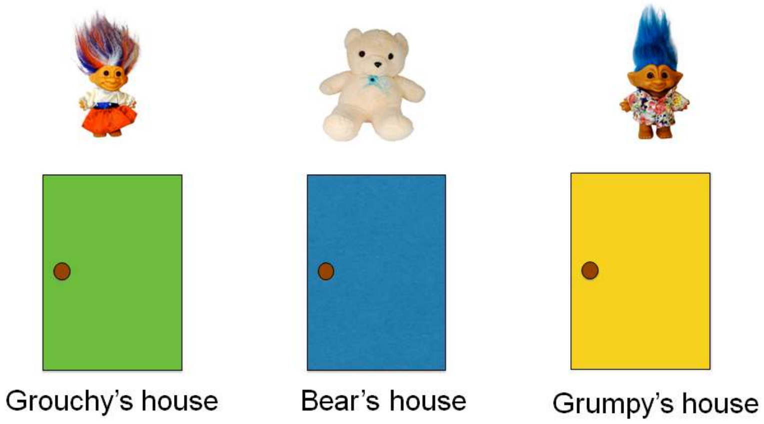


Figure 6. *Initial set up in bear's house*.

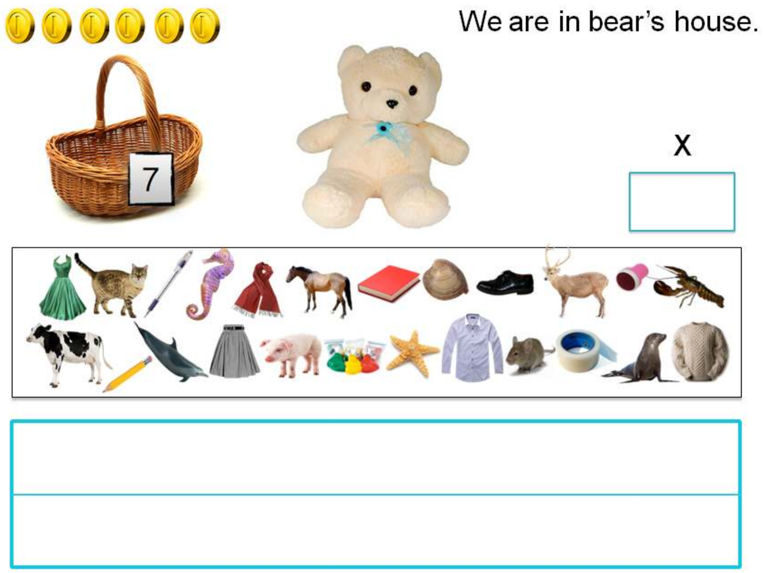
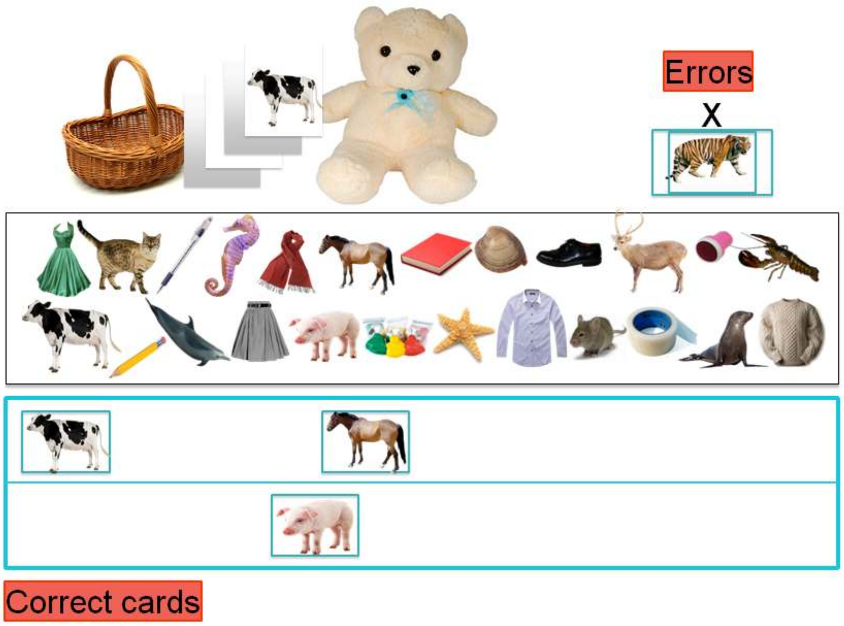


Figure 7. *Set up of Grumpy's house*.

Figure 8. *Example of a possible trip made by child*.



Results

Pretests

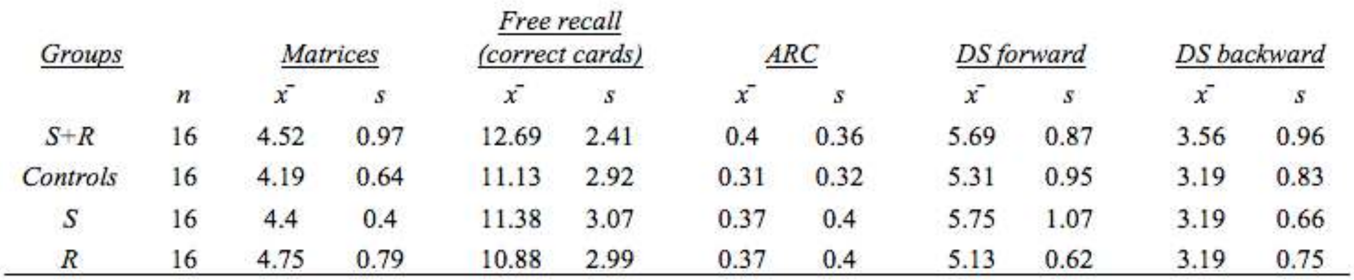
Prior to analyzing training effects, group performance on pretests was examined to ensure that there were no significant group differences in working memory. Given that digit span scores are not expected to be normally distributed (Babikian, Boone, Lu, & Arnold, 2006), Kruskal-Wallis tests were used to test group differences on the digit forward and backward span total and longest span. The test revealed that the scores for both forward χ^2 (3)=2.96, p = .4) and backward χ^2 (3)=3.05 p=.384) total scores were not significantly different across groups.

Shapiro-Wilk normality tests suggested that matrices W(64)=.97, p=.109), number of cards W (64)=.97, p=.14 and ARC W (64)=.57, p=.57 at pretest of the free recall were normally distributed. Box's M suggested that equal covariance matrices of the dependent variables can be assumed across groups F (9, 41255.3) = .84, p = .58, and Levene's test showed that variance of matrices F(3, 60)=1.32, p =.28 and number of correctly recalled cards F(3, 60)= .252, p =.86 were assumed equal across groups. A MANOVA was conducted to test the between-group differences for matrices and the pretest free recall. Results using Pillai's trace showed no significant between-group differences for these measures F(1.38, 120)= 1.38, p=.23.

Thus, these findings suggest that there were no significant baseline differences between groups on the pretests of visual and verbal WM, as well as in the tendency to use a clustering strategy. In other words, no group had any WM advantage compared to other groups prior to training.

Table 3

Group means for pretest measures



Groups		Matrices		Free recall (correct cards)		ARC		DS forward		DS backward	
	n	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
S+R	16	4.52	0.97	12.69	2.41	0.4	0.36	5.69	0.87	3.56	0.96
Controls	16	4.19	0.64	11.13	2.92	0.31	0.32	5.31	0.95	3.19	0.83
S	16	4.4	0.4	11.38	3.07	0.37	0.4	5.75	1.07	3.19	0.66
R	16	4.75	0.79	10.88	2.99	0.37	0.4	5.13	0.62	3.19	0.75

Between subjects far-transfer training effects

For the problem solving task, an aggregate performance score (C-E)/T was computed using the total scores ([total correct – total errors]/total tokens used) across three trials of the task. The performance score met the assumption of normal distribution according to the Shapiro-Wilk test (p=.11) with acceptable skewness (-.57) and kurtosis (-.04). Levene's test showed that there was a violation of the assumption of homogeneity of variance, $F(3,60)=2.94, p=.04$. However, the F test is fairly robust against inequality of variances when sample sizes are equal (Northwestern University, 1996). Consequently, I used the ANOVA to analyze the training effects. Using the aggregate performance score as the dependent variable and training condition as independent variable, a univariate ANOVA was conducted to examine whether training conditions affected problem solving performance. Results revealed a significant main effect for condition, $F(3, 60)= 3.04, p<.05, \eta^2_p= 0.14$. In response to the first research question, the results suggested that strategy-based WM training was successful in producing far-transfer effects to a novel problem solving task. Table 4 shows the descriptive statistics for the performance scores, and table 5 shows the specific contrasts in scores between groups.

Post-hoc analyses using least significant difference (LSD) revealed statistically significant differences in problem-solving performance between the double control group and the

S+R group (mean difference in performance was -.98, $p<.01$), between the double control group and the R group (mean difference = -.75, $p<.05$), and between the double control group and the S group (mean difference = -.73, $p<0.05$). In other words, the double treated control group performed significantly worse on the problem solving task compared to all the training groups (see row 1 table 5).

Surprisingly, the three treatment groups (S+R, S, R) did not differ significantly from each other (see row 2 table 5). In other words, there was no statistically significant difference in receiving combined training of both strategies compared to only one strategy. Furthermore, the R and S groups did not differ significantly from each other.

In addition to the performance scores (C-E)/T, the mean difference between groups on the number of correct cards retrieved across trials (C-E) is reported in table 5. This variable was not subjected to an ANOVA because it did not meet the assumption being normally distributed. However, this variable provides valuable information about the difference in performance between groups. Table 5 shows that, on average, the group receiving both S and R collected 14.63 cards more than the double controls (over the three trials). The rehearsal group collected on average 11.75 cards more than the double controls, and the semantic group collected 8.69 cards more than controls.

Figure 9 shows a comparison of the mean number of correct cards collected per trial for each group. The chart suggests that performance decreased for all groups over time. However, the S+R group performed better than all other groups and their performance appears to deteriorate much less than the double controls.

Table 4

Group means for problem solving outcomes

Groups	*n*	*Total performance*		*C-E1*		*C-E2*		*C-E3*		*Total Correct*		*Total Errors*		*Tokens*	
		$\bar{x}$	*s*	$\bar{x}$	*s*	$\bar{x}$	*s*	$\bar{x}$	*s*	$\bar{x}$	*s*	$\bar{x}$	*s*	$\bar{x}$	*s*
S+R	16	3.79	0.75	21.88	3.59	21.5	4.13	19.88	4.26	64.94	12.69	4.75	4.95	16.81	1.28
Controls	16	2.81	1.21	18.81	5.09	15.69	8.31	14.13	8.1	60.44	12.12	11.8	9.31	17.56	0.89
S	16	3.54	1.1	20.88	2.94	19.25	5.43	17.19	6.4	65.38	6.3	8.44	8.03	16.63	1.71
R	16	3.56	1.76	21.88	2.25	19.81	4.65	19	5.2	67.44	4.15	6.75	6.36	17.25	1.13
Total	64	3.42	1.02	20.86	3.7	19.06	6.1	17.55	6.42	64.55	9.66	7.94	7.63	17.06	1.31

Note: Total performance = (C-E)/T

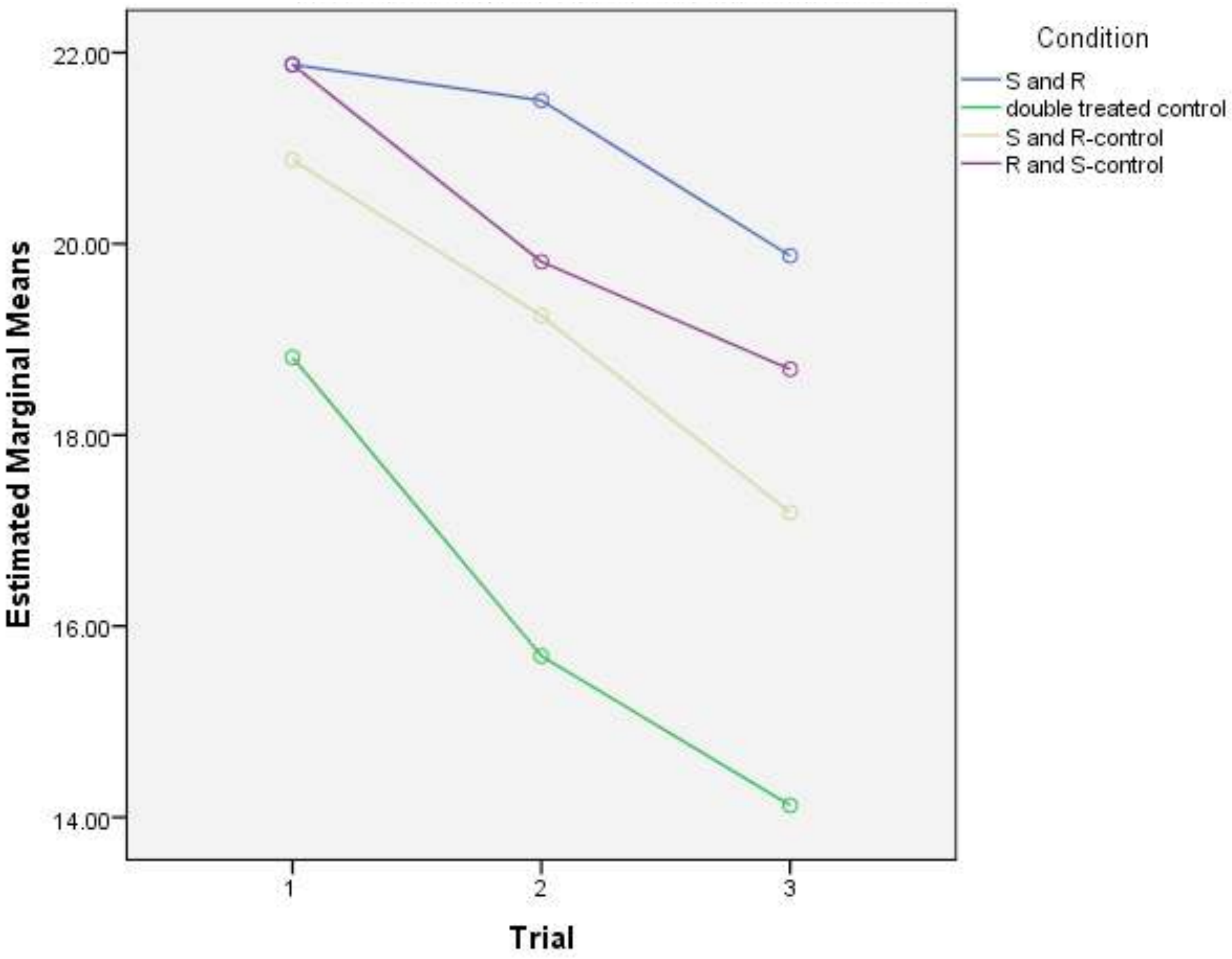
Table 5

Mean differences between groups on problem solving outcomes over all trials

Condition (I)	*Condition (J)*	*Cards correct (I-J)*	*Total performance (I-J)*	*p*
Controls	S+R	-14.63	-0.98*	0.006
	S	-8.69	-0.73*	0.038
	R	-12.06	-0.75*	0.035
S+R	S	5.94	0.25	0.48
	R	2.56	0.24	0.5

*significant at $p < .05$

Figure 9. *Means for problem solving performance (C-E) across trials.*



Between subjects near-transfer training effects

The near-transfer effect was assessed independently of the far-transfer effect of problem solving using a 2x4 mixed repeated measures ANOVA. In this analysis, I compared the performance on the pre-training free recall to the performance on the post-training free recall across conditions. Results showed no significant differences in the number of correctly recalled items across time (pre to post) and no significant interaction between time and condition. However, upon closer inspection (see Figure 10) it was apparent that performance was not contingent on whether participants received training or not, but on whether the training phase included the semantic strategy.

Thus, the variable of condition was recoded into two groups depending on whether training included semantic training or not. Using this newly created variable as between-subjects factor, a 2x2 repeated measures ANOVA revealed a significant interaction between the within-subjects variable of time (pre- to post- test) and the newly created training variable $F(1,62) = 4.19$, $p < .05$, $\eta^2_p = 0.06$. However, one should be cautious in interpreting this finding, because Box's M suggested that the assumption of homogeneity of variance was violated, and although the variable was normally distributed at pretest, it was not at post-test.

With this being said, the mean of the group with no semantic training (controls, R) was 11 (*SD*=2.91) at pretest and 10.47 (*SD*=3.35) at post-test. The mean of the group who had received semantic training (S+R, S) was 12.03 (*SD*=2.8) at pretest and 13.16 (*SD*=4.78) at post-test. Despite this seemingly marginal change, it is important to note the opposite direction of change for each group. As shown in Figure 11., those children who received semantic training did better at post-test, whereas those children who did not receive semantic training actually did worse at post-test.

Another outcome variable of interest in the near-transfer analysis was the adjusted clustering ratio (ARC), which measured the extent of category clustering in free recall. A separate 2x2 repeated measures ANOVA using ARC as dependent variable revealed no significant interaction between the recoded training condition (S or no S) and ARC pre- and post-test scores. However, there was a main effect of the within-subjects factor of time $F(1, 62)= 23.57$, $p <.001$, $\eta^2_{p=} 0.275$, suggesting that all participants, regardless of training condition, significantly increased their use of a clustering strategy on the post-test free recall (see Figure 12). However, ARC was not normally distributed, and thus results must be interpreted with caution.

ARC scores were significantly correlated with the number of correct items retrieved in the post-test free recall ($r = .46$, $p <.001$). Interestingly, whereas the above results suggest that practice effects alone may have induced a slightly greater tendency to use clustering strategy, only training in the semantic strategy produced statistically significant changes in the performance on the free-recall post-test.

Lastly, as the free recall was a measure for near-transfer effects and the problem solving task was a measure for far-transfer effects, it was also informative to determine how performance on these measures related to each other. The number of correctly recalled items on the free recall post-test was significantly correlated with total problem solving performance (r=.56, $p <.001$). Therefore, near-transfer effects seem to be highly correlated with far-transfer effects.

Figure 10. *Correctly retrieved items on free recall by condition.*

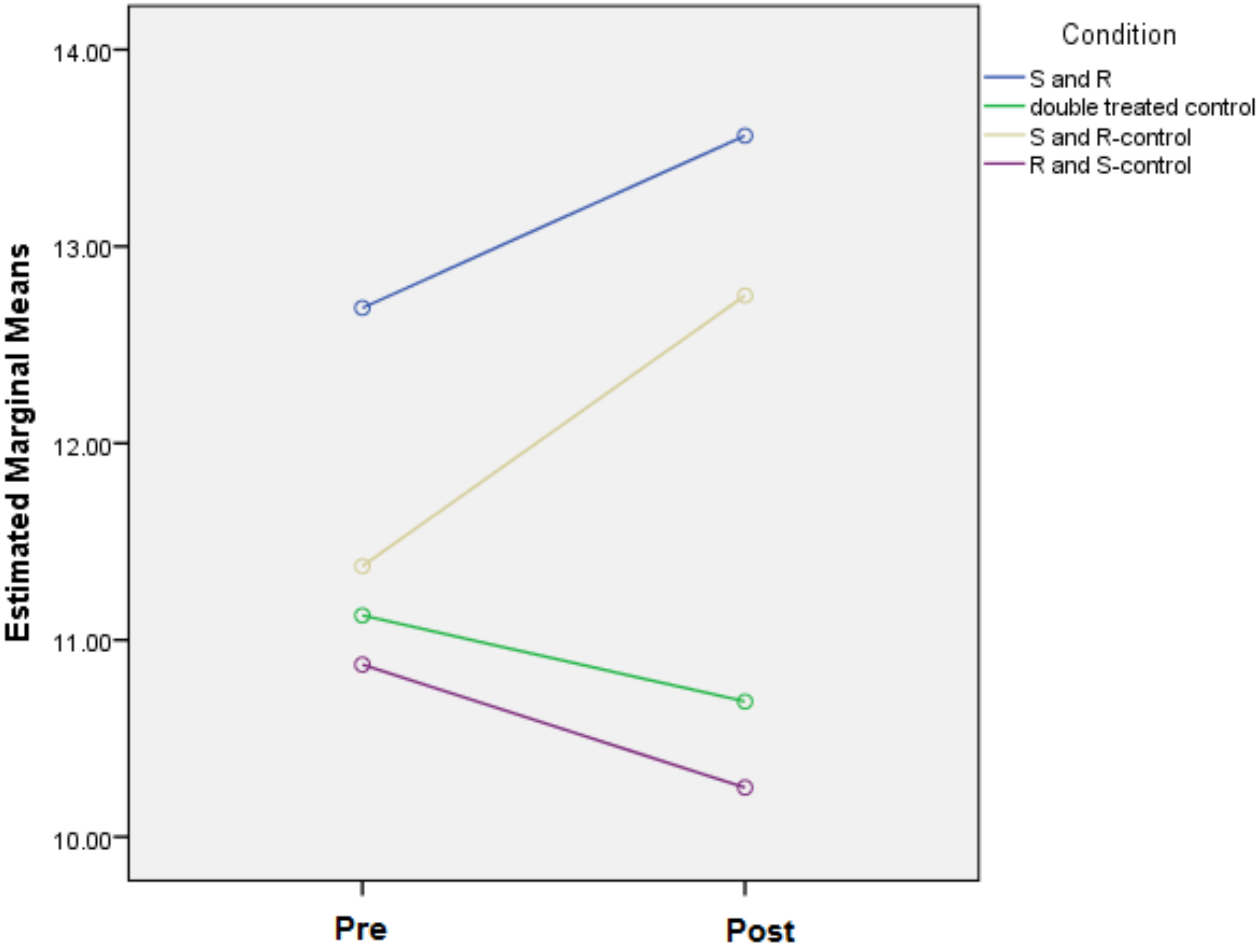


Figure 11. *Correctly retrieved items on free recall by recoded condition (S or no S).*

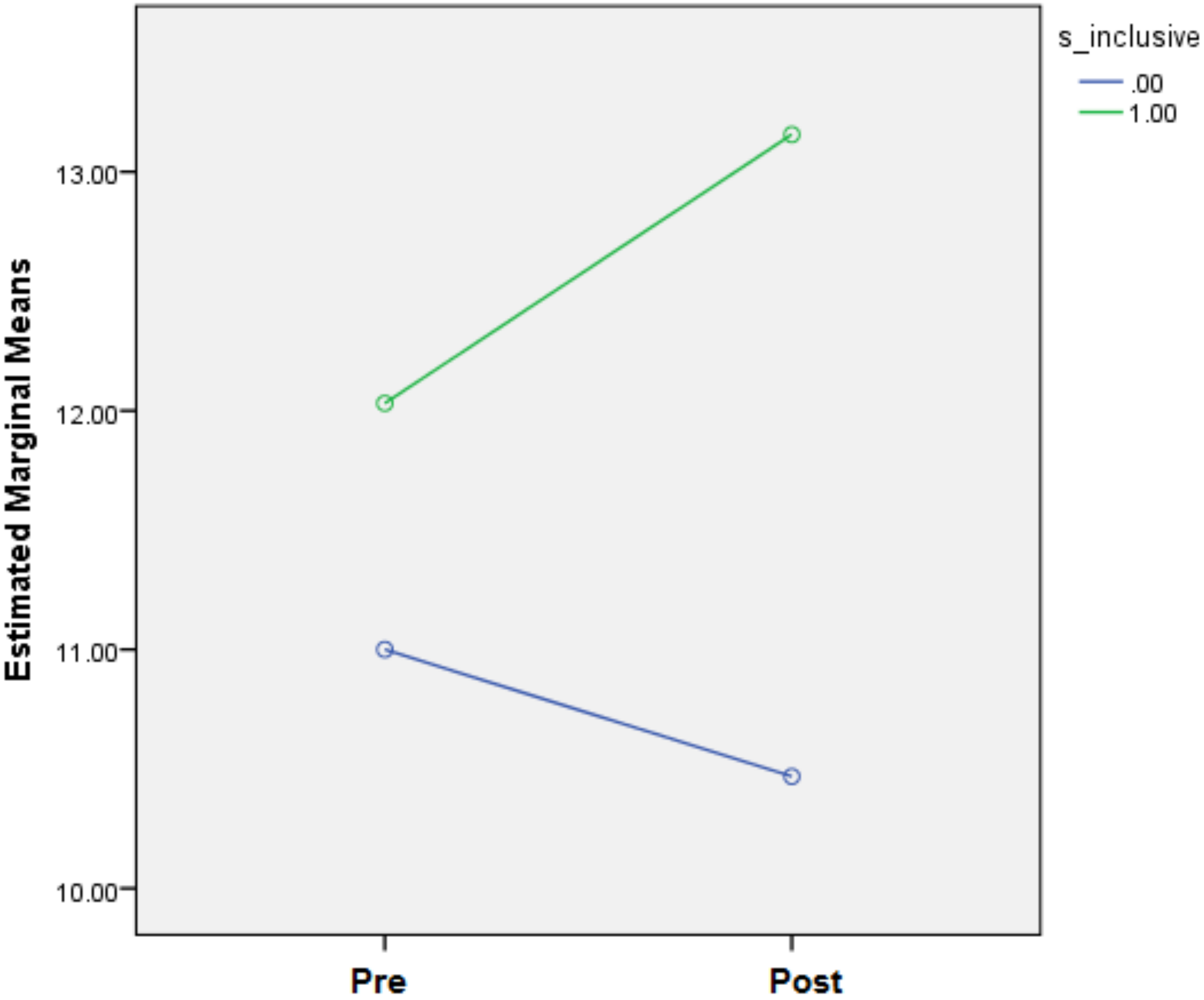
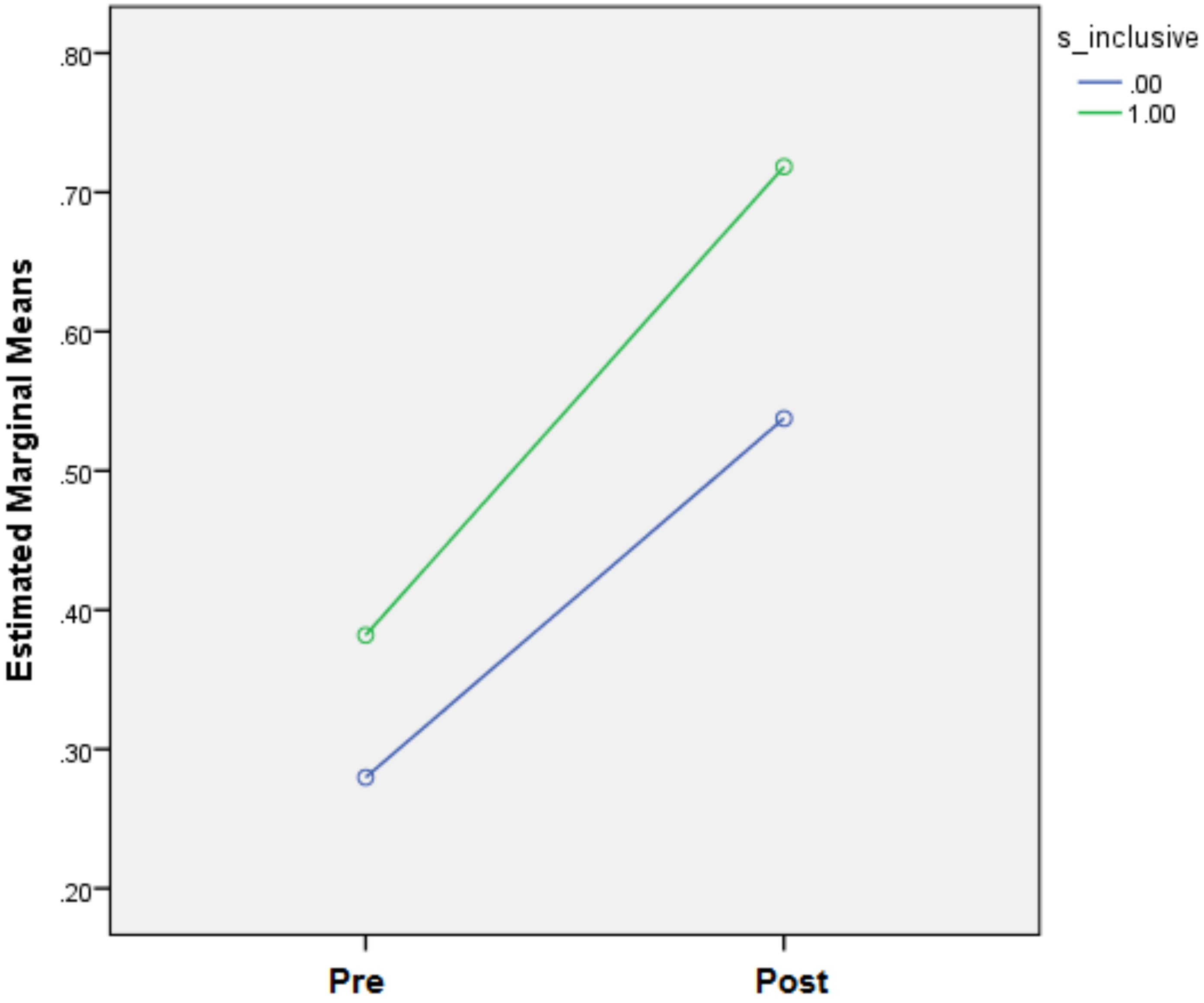


Fig 12. *ARCs on free recall by recoded condition (S or no S)*.



Individual differences in treatment effect

The last research question concerned the issue of whether children with lower baseline WMC would benefit more from training than those with initially higher working memory capacity. In order to explore this question, children were divided into two equal groups of low and high WM based on standardized scores derived from the pretest measures (digit span forward total, digit span backward, matrices, and free recall). Standardized scores were calculated for each of the separate measures and then summed to produce an overall standardized score for each participant. The median of the aggregate score was used to divide the participants in the low and high WMC groups.

A 2x4 between-subjects ANOVA was conducted using condition and baseline-WM (low and high) as between-subject factors. The main effect of baseline-WM was significant $F(1, 63)= 20.68, p<.05, \eta^2_p =0.25$. Low WM participants did worse on the problem solving task (*mean* =2.91, s=.94) compared to high WM participants (*mean*= 3.94, s=.93). However there was no significant interaction between condition and baseline WM. In other words, children's problem solving performance scores did not differ between the different training conditions based on their baseline working memory.

A separate 2x4 ANOVA was conducted on condition and baseline-WM on the total number of errors made in the problem solving task. Results showed that low WM children made significantly more errors on the problem solving task $F(1, 62)= 11.25, p <.01, \eta^2_p=0.92$. The low WM group made an average of 10.91 total errors on the problem solving task (s=8.14) and the high WM group made an average of 4.97 (s=5.83) errors on the task. The difference in total errors was confirmed to be significant with an independent samples t-test t(62)=3.36, p=.001

(equal variances assumed). However based on the ANOVA, there was no significant interaction between baseline-WM and training condition on the total number of errors.

However, the error variable was not normally distributed; thus results must be interpreted with caution. However, the performance variable and error variable were highly correlated (r=-.86, p <.001), and given that low WM groups performed significantly worse on the problem solving task based on performance scores, it is not surprising that this group also makes significantly more errors.

An alternative method to examine whether an interaction existed between baseline WMC and training condition was to conduct a multiple regression. Group 2 (double controls) was used as a comparison group and the remaining conditions were dummy coded. The aggregate standardized score based on WM pre-tests was used as the proxy variable for baseline WM. A multiple regression was performed with total performance as the dependent variable and the dummy coded training groups (except for group 2) and baseline WM as predictors. The model was significant, accounting for 33.3% of the variance in problem solving performance, (R^2 adjusted = .333), F(4, 59)= 8.87, p<.001. A second multiple regression was performed adding in the interaction terms between baseline WM and the dummy coded training groups. Results showed that by adding in the interaction terms, the new model accounted for 33.9% of the variance in problem solving performance, (R^2 adjusted = .339), F(7, 56)= 5.61, p<.001. Results indicated that this model was not a significant improvement from the previous model, and none of the interaction terms were significant.

Problem solving construct validity

The premise that WM training should improve problem solving is contingent on the idea that the problem solving task involves a WM component. Because the problem solving task was a newly developed task, it was essential to examine the correlations between the pretest WM measures and problem solving performance. All WM pretests including the DS forward (r=.28, p<.05), DS backward (r=.34, p<.01) and matrices (r=.31, p<.05) were found to be significantly correlated with the problem solving task. Furthermore, a stepwise multiple regression analysis was used to determine whether the pretest measures significantly predicted participants' problem solving performance. Results indicated that two predictors explained 15.5% of the variance in performance (R^2 adjusted= .149), F(2, 61)= 6.79, p <.01. Matrices significantly predicted problem solving performance (β=.36, p<.01), as did the backward span task (β=.37, p<.05).

Discussion

The following discussion is organized into corresponding sections with the results. The first section discusses the results pertaining to the differences between training groups in their performance on the far-transfer problem solving task. The second section discusses the results pertaining to the differences between groups in their performance on the near-transfer free recall post-test. The third section discusses the interaction between two between-subjects factors, baseline WMC and training condition. Finally, given that the problem solving task was a newly developed task, a few comments on its validity as a problem solving measure are provided. Lastly, the final section outlines several limitations of the current study and concludes with a few final remarks on training studies overall.

Between subjects far-transfer training effects

The primary goal of this training study was to determine whether strategy-based training could produce a far-transfer effect in problem solving. Results suggested that strategy-based WM training did transfer to a novel problem solving task in which the task demands of the strategy training were embedded. The double control group performed significantly worse on the problem solving task compared to all three training groups. Surprisingly, the three treatment groups did not differ significantly from each other. In other words, there was no statistically significant difference in receiving combined training of both strategies compared to only one strategy.

With regard to the expectation that combined training would be better than training of one strategy, a possible explanation for why there was no statistically significant difference in problem solving performance is that the post-test measure might not have been sensitive enough to capture potentially stronger effects of the combined training. Under conditions of even higher cognitive demand or at a long-term follow-up, the training of two strategies may show a greater training advantage. It is possible as well that the training phase itself was too short. As mentioned in the literature review in the introduction, similar strategy-based training studies were much more time intensive, spanning multiple sessions over weeks. Training effects might have been amplified if there had been multiple training sessions to help foster learning.

Alternatively, children may not have benefited more from the combined training than from the single-strategy training because they were exercising (i.e., focusing their attention on) only one strategy in the combined condition even though both were taught. Being able to monitor and evaluate the effectiveness of different strategies in different situations is a vital antecedent to the selection of superior strategies, but completing these actions simultaneously is cognitively demanding (Touron, Matthews, & Hines, 2010; Whitebread, 1999). It has been suggested that if a task is cognitively demanding, children with limited cognitive resources will make less adaptive strategy choices (Imbo, Duverne, & Lemaire, 2007). Furthermore, children with lower WM would be more likely to experience what is known as a "utilization deficiency" (Gaultney, Kipp, & Kirk, 2005; Bjorklund, Schneider, Cassel, & Ashley, 1994). This occurs when a child is instructed to use a particular strategy and seems to comprehend the steps in its application, but still fails to implement the strategy in production due to limited cognitive resources. According to one process model, there is a constraint on the available processing capacity in WM which limits the amount of attention that can be distributed over concurrent tasks (Lovett & Anderson, 1996). Younger children who have lower WM may not always be able to effectively implement strategies that they are trained in. Therefore, the effortful task of alternating and deciding between two strategies may actually impede the ability to use both, and if one strategy is easier to implement, a child may default to using only one.

A second possibility was that one strategy would be more effective than the other. I had expected that a semantic strategy (S) would be more useful to children than a rehearsal strategy (R) based on the finding that children with higher WMC tend to use deeper encoding strategies as opposed to surface ones (Dunlosky & Kane, 2007; Friedman & Miyake, 2004). Furthermore, in free recall tasks, participants are usually able to recall much longer lists of items compared to their longest span in digit span tasks due to the associations made through clustering. In free recall, one does not rely exclusively on memory because items from one category prime activation for other items in the same category. The expectation was that children who were taught to use this clustering strategy would be able to apply it to the problem solving task and outperform those that simply used rehearsal. However, the current findings did not support this idea; there were no significant differences between the training groups that received semantic training only and rehearsal training only.

As to why this occurs, recall that in cognitively demanding tasks, children tend to use less effortful strategies. If rehearsal is less effortful than semantic categorization, the prior may have been fully implemented in the problem solving task, but the latter was not because it required deeper processing. In a cognitive taxing situation where many other demands interfered with the child's ability to focus on a single task, the ability to use a more complex strategy to its full extent may have been impaired. The problem solving task involved many overarching game rules and sub-goals in addition to memorizing items for each trial. The near-transfer effects found in this study lend credence to this theory. When comparing R-only and S-only groups in their performance of the free recall post-test, those that received semantic training outperformed those that received rehearsal training. In the much less cognitively demanding free recall task, it is possible that children were able to use the semantic strategy more effectively and thus this strategy was shown to be superior to rehearsal in helping children recall items. However, in the case of the problem solving task which had many other embedded task demands, the more effortful semantic strategy may have been used incompletely or inaccurately leading to no significant differences between those groups in performance.

Another theory as to why the S group did not outperform the R group is that the categories in the problem solving task themselves may not have been transparent enough in order to elicit a detailed categorization strategy from children. The groups of aquatic and terrestrial animals may have been too similar and perceived as a single large category. Using a clustering strategy based on this single large category would not have been as helpful because within any given list, half of all the "animals" were incorrect cards. School supplies may also have been too broad of a category. Without making this connection, items may not have appeared to belong to any coherent category.

Lastly, semantic categorization is a good strategy for recalling a greater number of items because of primed associations, but arguably not any less prone to errors. In fact it may actually impede performance in a task that requires high distinction for accuracy because of this same priming effect. During free recall tasks, participants occasionally make intrusion and repetition errors. Intrusions errors occur when the subject names an item that belongs to a target category, but was not actually an existing item on the list (e.g., a dog in a list of animals that had no dog, but perhaps had a cat). Repetitions occur when the subject names the same item more than once. In a free recall task, this does not affect accuracy, but in the problem solving task, committing an item to memory that has already been collected, or one that was on a previous list but not current list, essentially "wastes" a mental slot that could have otherwise been used for a correct card to be collected. Accuracy was also an important component of the game; if clustering both increased recall but also the propensity for intrusion and repetition errors, the overall performance score would be jeopardized. This is especially true because items of the same category may have been a target on a previous list, but not on a current list, (e.g., cat it one list, tiger on another) creating intertrial interference that clustering would only exacerbate. Therefore if the semantic strategy impeded accuracy or was incompletely implemented, the non-significant difference in performance between R-only and S-only groups could be explained despite S strategy being expected to improve recall more than rehearsal.

It should be noted however, that in spite of the possible interference factors described above, that the S-group still performed significantly better than the double treated controls, and this suggests that the strategy was still overall successful in improving performance in a far-transfer task.

Interestingly, results showed that problem solving performance decreased for all groups over time. A possible explanation is that proactive interference increased as trials progressed, making it difficult to distinguish target items of the current list from target items of a previous list. Proactive interference occurs in a situation in which prior information inhibits the ability to recall something new, such that performance declines on subsequent trials when study material is drawn from repeated categories (Sullivan, 2009). Consider the following characteristics of the trials: first, it was important to randomize the use of each item so that the presence on a subsequent list could not be predicted based on its appearance or lack of appearance in a previous trial. For example, a bear appearing in T1 would not have indicated whether it would or would not appear in T2. However, because there were a limited number of items, some items did repeat across trials. This was necessary to ensure that a child was actively using working memory to decide whether an item was a current target, and not adopting the incorrect assumption that, “Since this item has appeared in a previous list, it should not appear again on a new list”.

Due to the unpredictability of repeated items, as trials progressed, children would have increased difficulty in distinguishing between target items from a previous trial and a current trial. This interference from previous trials is known as proactive interference. On T2, there is proactive interference from T1. On T3, there is proactive interference from T1 and T2. Proactive interference, combined with mental fatigue and decreasing motivation over time, may explain decreased performance in the problem solving task across the three trials.

However, it is important to note that despite the presence of proactive interference across trials, the training groups performed significantly better than the control group and this relative difference is maintained for all groups compared to the controls. In other words, training groups appeared to deteriorate much less than the double controls. The S+R group in particular, deteriorated the least compared to all other groups, although this last difference was not statistically significant (but it may have been with a more sensitive task).

Between subjects near-transfer training effects

The goal of administering the pre- and post- measures of the free recall was to replicate findings of previous studies: that strategy-based training is effective in near-transfer tasks immediately after training (Melby-Lervåg & Hulme, 2012). The post-test free recall was completed at the end of session after the problem solving task, and it should be noted that many children experienced fatigue and were not highly motivated to complete the last task. However, given that the control tasks ensured all groups received approximately equal mental stimulation and time in the experiment, the level of fatigue should have been similar across conditions with no single group at a relative advantage or disadvantage.

Given the high variability in individual children's level of attention and motivation by the end of the testing session, the first analysis using the original training variable as the predictor revealed that there was no between-group change in post-test free recall scores. However, a significant interaction was revealed when groups were recoded into two groups, one that received semantic training and one that did not. Those who received S training did better on the free recall post-test, whereas those who did not receive S training did worse. Importantly, the change was in different directions for the groups. A possible interpretation of this result is that without targeted training, mental fatigue would have resulted in decreased performance from pre to post-test, but targeted training mitigated this and so that performance improved rather than deteriorated in the post-test. Instead of replicating previous findings on near-transfer effects, the results suggested that targeted training may also be able to mitigate effects of undesirable environmental factors on a task at hand, such as reduced attention and motivation, fatigue, or boredom.

The adjusted ratio of clustering (ARC), a measure for the extent of clustering used in the free recall task, also revealed an interesting relation with the number of items correctly recalled. ARC scores increased from pre to post-test regardless of training condition, but improvement in the free recall, as measured by correctly recalled items, was only observed for the groups that received semantic training. I suggest that individuals' increased use of clustering at the post-test is attributable to a practice effect (assuming that the difference is a true difference given the violated assumption of normality underlying the ANOVA). It is possible that in situations characterized by decreased attention or fatigue, practice effects alone are not enough to sustain a near-transfer effect in task performance, but targeted strategy training may sustain improved task performance under these non-ideal conditions.

On a side note, performance on the free recall at post-test was significantly correlated with performance on the far-transfer problem solving task, lending credence to the idea that semantic categorization, which evokes deeper processing of information, may have been helpful in the more demanding problem solving task. In future research, it would be helpful to administer a rehearsal post-test as well to determine which strategy is more closely correlated to problem solving performance.

Individual differences in treatment effect

Although it was expected based on previous findings that children with lower baseline working memory would benefit more from training (Turley-Ames & Whitfield, 2003), results did not show any significant interaction between training condition and baseline working memory on the dependent measures of problem solving performance and total errors. Contrary to expectations, training did not help children with lower WMC more than children with higher WM in making fewer errors on the problem solving task, or score higher on overall performance.

In retrospect, perhaps it is not surprising that none of the training conditions were more helpful to children with lower baseline WMC. With regards to the idea of two strategies over one, if it were more cognitively effortful to implement two strategies in general, then this would be especially challenging for children with lower baseline WMC. One might argue that they would be even less likely to effectively use two strategies instead of one because of the limitations in WM. Children with lower WMC are also more susceptible to utilization deficiencies and at the same time, in cognitively demanding tasks, children tend to use cognitively less effortful strategies. With these two factors in tandem, children with lower WMC who were taught the more effortful semantic strategy may have failed to use it in a cognitively demanding problem solving task, or finding it too effortful, may have defaulted to a simpler rehearsal strategy. This phenomenon occurs despite the semantic strategy being more useful to children with higher baseline WM resources. This presents a conundrum for the current and many other training programs: the group being targeted in training may also have the characteristics of those who are most susceptible to utilization deficiencies.

Given that this was a typically developing sample of children, it would be expected that most children would be performing at an average level of WM for their developmental level. One might argue that a differential response to training in children with low and high WM could only become apparent when there are sufficiently large samples of children with low and high WMC. In other words, to see this effect one would have to collect samples not of average children, but of those with particularly low WM scores at baseline. This analysis is beyond the scope of the current study, and future studies should enroll a larger sample to test for this possibility.

An alternative view is that children with lower baseline WM do not benefit more from training than children with higher baseline WM. That is not to say they do not benefit at all, but there is no reason to expect that given their initial limitations, that they would be more apt at using the training than children with high WMC. Results of the current study seem to support this idea, and, in fact, suggest that children, regardless of their initial levels of WMC, benefit from WM training. This is an encouraging thought rather than a disparaging one, as it suggests that cognitive intervention may be broadly beneficial.

Furthermore, in addition to baseline levels of WM, performance on a problem solving task depends on many other factors. For example, I would expect that children’s individual differences in being able to self-regulate attention and frustration are particularly important to a novel problem solving task to keep track of rules and formulated plans. The capacity to control attention is a major contributor to differences in WMC (Barrett, Tugade & Engle, 2004). Children who are prone to getting frustrated with a challenging task may be less inclined to focus their attention and efforts on the task. Research, for example, has suggested that cognitive ability contributes to the control of emotional responding in that individuals with higher WMC are better able to suppress negative emotions as well as respond to stimuli in an unemotional manner (Schmeichel, Volokhov, & Demaree, 2008). In a challenging task that evokes frustration, this emotional control may be important in keeping the child on task and motivated to keep trying despite initial difficulties.

Problem solving construct validity

An interesting finding was that the digit forward task was not significant predictor of problem solving performance in the multiple regression analysis. Digit span tasks measure verbal short-term memory, and being able to hold more information in verbal short-term memory would conceivably improve performance on the problem solving task. However, if a measure for short-term memory is not a significant predictor of performance on the problem solving task, how then, does rehearsal training alone produce a significant difference in problem solving performance compared to the double control group? Perhaps the digit forward task, although a good measure for verbal short-term memory, does not really capture the use of articulatory rehearsal, and therefore was not a good predictor of rehearsal use that would improve problem solving performance.

Alternatively, perhaps the rehearsal component of training was not responsible for improvement on the problem solving task in so much as verbal recoding was useful to children. Recall that rehearsal training involved two distinct phases. First, children were familiarized with recoding nonverbal stimuli (pictures) into verbal stimuli. Second, children were trained in using the rehearsal mechanism actively on increasingly longer lists. In the orally presented lists, children were forced to engage the phonological loop instead of relying on visual short term memory, something that children may not have done spontaneously or independently if they started out with a preference for using VSSTM instead of verbal STM. In other words, this speaks to the cognitive shift of nonverbal to predominantly verbal encoding of information as one of the factors responsible for developmental increases in WMC. Perhaps it was not the training of rehearsal that was important in improving performance on the problem solving task, but the enhancement of verbal strategy use itself.

Additionally, the two phases of rehearsal training involved a transition from immediate recall to delayed recall, a component that is not part of the digit forward task. Therefore, the rehearsal training may have been effective in producing a far-transfer effect for several reasons. First training may have been useful in getting children to actively recode visual stimuli into verbal stimuli, and second, facilitate the use of verbal strategies regardless of whether this specifically means rehearsal. Another component of the rehearsal training was to encourage children to hold verbally recoded items for longer delays.

Finally, because of the emphasis in literature on the importance of training effects tapping into core processes, it is also worth mentioning that the problem solving task meets several criteria of a core-based task. It requires the maintenance of information in face of interference. Target stimuli must be remembered and selected in the face of task-irrelevant stimuli. Specifically, children have to remember target items from a current list, and ignore target items from previous lists that are no longer relevant. This risk of mis-identifying previous targets minimizes automatization in that children must reevaluate their planning and strategies in every trial. Children are also required to constantly update items in memory as they collect cards, to make sure they are only committing those to memory that have not been retrieved. The rules of the game, including constraints on the number of tokens and cards allowed in the basket, create a task that has superordinate goals and rules, but also an ambiguous solution. Juggling between these goals and rules creates a task high in cognitive load. The fact that the problem solving game is able to meet some core-based criteria lends further evidence to the idea training effects have impacted core processes. At the very least, there is greater confidence that these far-transfer effects will replicate under conditions of interference and high cognitive load – a necessary prerequisite for transfer to the real world.

Limitations

There are several limitations to this study that must be addressed. First, as mentioned in the power analysis, the one-way ANOVA required a much larger sample size (n=230) given the parameters of 4 groups, a medium effect size (0.25) and 95% confidence. The reported observed power from the test of between-subjects effects was 0.686. Therefore, to have greater confidence that a type I error is not being made, the study would have to be replicated with a much larger sample size. Additionally, a larger sample with increased power would ensure that a Type II error is not being made with respect to the lack of significant differences between the different training conditions.

In retrospect, a more narrow age range could have been selected for this study (e.g., 7-8 year olds). The variability in performance between 6-9 years was perhaps too large, with 6-year olds tending to do much worse, and 9 year olds doing much better than children in the middle of the range. However, the majority of children still fell between the ages of 7-8 and recruitment was adjusted early on to preferentially contact these children within this age range.

Given the time constraints of the Master's thesis, it was not feasible to conduct extended training over multiple sessions. WM studies have typically involved several training sessions (about half an hour long each) over the span of a few weeks. Although extent research has not established a relation the amount and effectiveness of training, it is likely that a more intensive training program would achieve longer-term training effects than a single session lasting just half an hour (i.e., the training component was approximately half an hour and the remaining hour was used for pre-tests and the problem solving post-test). In particular, children who were trained in two strategies may have benefited from more training sessions because of the complexity of mastering and applying two strategies.

Lastly, children's baseline WM could be more precisely established in future studies to more clearly look at how children with low WM respond to training in comparison to children with high WM. Furthermore, it is important that individual differences are not underestimated in any learning situation. Aside from baseline WM, other relevant factors such attention, emotional regulation, conscientiousness, or even curiosity could also be assessed to see how these factors influence individual differences in both responsiveness to training, and in a novel problem solving context.

Summary of results

In summary, results of this study showed that strategy-based WM training produced a far-transfer effect to a problem solving task. All groups that received training, including strategy training in phonological rehearsal, semantic categorization, or both, performed better on the problem solving task compared to controls. The study also expanded on the findings for near-transfer effects of training by showing that targeted training can improve performance in tasks under conditions of mental fatigue, where performance would otherwise deteriorate. Examining the interaction between training and baseline WM, the study found that children with lower baseline WMC did not benefit more from training compared to children with higher WMC. In fact, results suggest that children benefitted from training regardless of their initial WMC. This finding is encouraging with respect to the broad applications of training, however the current study was limited by its lack of long-term follow up and future research must still be conducted on long-term, distributed training paradigms which take place over time and adjust to children's changing competencies. Furthermore, an unresolved conundrum remains that children with lower WMC are most susceptible to utilization deficiencies, particularly in situations of high cognitive demand. In order for training to be more effective for these populations, future research must focus on developing training methods that addresses this limitation.

A final comment on training studies

I would like to end this paper by reemphasizing that individual differences cannot be underestimated in any learning context, and this certainly applies to cognitive training. Interventions should be sensitive to the needs of the individual, and the best training is ideally customized to those needs. Researchers in this field may be prone to adopting the mentality of being on a quest for the "holy grail" of training. We seek a parsimonious solution to training which shows advantages in a widespread range of outcomes; a permanent, ubiquitous far-transfer effect that can be replicated with every individual (or a solution that comes as close to this as possible), and this simply may not be realistic. Earlier in the paper I gave the analogy that distributed learning is more effective than massed learning and results in better retention of information. It is arguable that to achieve any permanency in training effects, one must undergo ongoing training, or training that adjusts to the increasing proficiency of the individual so that progress does not stagnate.

I would argue that an important requirement of intervention efforts is knowing what final outcomes are expected; in other words what is the ultimate goal of the training? In the current study, strategies were embedded in the problem solving task such that there was a reasonable expectation that training would produce improvements in the problem solving task. However, it is less likely that this training would help children in arithmetic problems or problems related to social situations, even though these domains fall within the broader context of problem solving. However, if the goal of an individual is to improve his or her problem solving skills as it relates to arithmetic or social interaction, then training would be tailored accordingly and less emphasis would be placed on specific working memory strategies as done in this study. The point is that one cannot realistically expect a magical solution to training that results in a ubiquitous far-transfer training effect. Training must be reasonable for a set of specific goals relevant to the individual.

The fervor that continues to surround cognitive training stems from a desire to help individuals reach their full potential in spite of, or regardless of what predetermined limitations are imposed by nature. Gilbert Gottlieb (2007) discussed this interplay of environmental factors and genes in producing specific probabilistic outcomes in his paper on “probabilistic epigenesis”. He distinguished between two phenomena, “norm of reaction” and “reaction range”. To illustrate, Gottlieb discussed two phenotypically different groups of rats: “maze-bright” and “maze-dull”. The rats were reared in deprived, usual, or enriched conditions. What was predicted was a reaction range phenomenon, so that the relative difference in maze performance between groups would be maintained across the three conditions; maze “bright” rats were expected to always outperform their duller counterparts regardless of which condition they were being compared in. However, results did not support this theory, and instead a norm of reaction phenomenon was observed. The difference between maze-bright and maze-dull rats was only apparent in the usual rearing condition. In the deprived condition, both groups of rats did worse, and in the enriched condition, both groups of rats did better, and there was no significant difference between the “bright” and “dull” groups.

The relevance of this example to the current thesis is that it forces us to ponder to what extent children’s cognitive outcomes are truly limited by predetermined factors. The enriched rearing condition for Gottlieb’s rats is akin to cognitive intervention and training for children with the goal of improving their cognitive outcomes. One might counter the idea that there stratified groups of high and low performing children in the domain of WM and problem solving, given that these divisions might dissolve if we find specific conditions or interventions to enrich what outcomes are expected. So long as we remain optimistic and interested in determining the factors that are most useful to producing these optimal outcomes for children, there is no telling what a child can achieve.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131(1),* 30-60.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short Term and Working Memory in Children: Are They Separable?. *Child Development, 77(6),* 1698-1716.

Al-Namlah, A. S., Fernyhough, C., & Meins, E. (2006). Sociocultural influences on the development of verbal mediation: Private speech and phonological recoding in Saudi Arabian and British samples. *Developmental Psychology, 42(1),* 117-131.

Baddeley. A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 2(11),* 417-423.

Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, A., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology, 36(2),* 233-252.

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36(8),* 1383-1390.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83.

Black, M. M., & Rollins, H. A. Jr. (1982). The effects of instructional variables on young children's organization and free recall. *Journal of Experimental Child Psychology, 33(1),* 1-19.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin, 130(4),* 553-573.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273–293.

Carretti, B., Borella, E., & De Beni, R. (2007). Does strategic memory training improve the working memory performance of younger and older adults? *Experimental Psychology, 54(4),* 311-320.

Cohen, G. & Conway, M. A. (Eds.) (2008). *Memory in the real world.* New York: Psychology Press.

Comblain, A. (1994). Working memory in Down syndrome: Training the rehearsal strategy. *Down Syndrome Research and Practice*, *2*(3), 123-126.

Camos, V., & Barrouillet, P. (2011). Developmental change in working memory strategies: From passive maintenance to active refreshing. *Developmental Psychology, 47(3),* 898-904.

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition, 39(2),* 231-244.

Comblain, A. (1994). Working memory in Down syndrome: Training the rehearsal strategy. *Down Syndrome Research and Practice, 2 (3),* 123–126. doi:10.3104/reports.42

Conners, F. A., Rosenquist, C. J., Arnett, L., Moore, M. S., & Hume, L. E. (2008). Improving memory span in children with Down syndrome. *Journal of Intellectual Disability Research, 52(3),* 244-255.

Conrad, R., Baddeley, A. D., & Hull, A. J. (1966). Rate of presentation and the acoustic similarity effect in short-term memory. *Psychonomic Science, 5(6),* 233-234.

Cowan, N. (2005). *Working memory capacity*. New York and Hove, UK: Psychology Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and Brain Sciences, 24(1),* 87–114. doi:10.1017/S0140525X01003922.

Darling, S., Sala, S. D., & Logie, R. H. (2009). Dissociation between appearance and location within visuo-spatial working memory. *The Quarterly Journal of Experimental Psychology*, *62*(3), 417-425.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*, *37*(10), 1189-1199.

Della Sala, S., Baddeley, A. D., Gray, C., & Wilson, L. (1997). *Visual Patterns Test: VPT*. Harcourt.

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology, 60(9),* 1227-1245.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 972-992.

Ford, C. E., Pelham, W. E., & Ross, A. O. (1984). Selective attention and rehearsal in the auditory short-term memory task performance of poor and normal readers. *Journal of Abnormal Child Psychology*, *12(1),* 127-141.

Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37(2),* 283–299. doi:10.2307/1126804

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology: General*, *133*(1), 101-135.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265-281.

Gaultney, J. F., Kipp, K., & Kirk, G. (2005). Utilization deficiency and working memory capacity in adult memory performance: Not just for children anymore. *Cognitive Development*, *20*(2), 205-213.

Gerjuoy, I. R., & Spitz, H. H. (1966). Associative clustering in free recall: Intellectual and developmental variables. *American Journal of Mental Deficiency*, *70(6),* 1966, 918-927.

Gilhooly, K. J., Logie, R. H., & Wetherick, N. E. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition, 21*, 115-124.

Henry, L. (2011). The development of working memory in children. London: Sage.

Henry, L. A., Messer, D., Luger-Klein, S., & Crane, L. (2012). Phonological, visual, and semantic coding strategies and children's short-term picture memory span. *The Quarterly Journal of Experimental Psychology, 65(10*), 2033-2053.

Henry, L. A., Messer, D. J., & Nash, G. (2012). Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, *53*(1), 37-45.

Henry, L. A., Turner, J. E., Smith, P. T. & Leather, C. (2000). Modality effects and development of the word length effect in children. *Memory, 8,* 1-17.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General*, *130*(2), 184-198.

Holmes, J., Gathercole, S. E., & Dunninng, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science 12(4),* F9-F15

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108(25*), 10081–10086. doi:10.1073/pnas.1103228108

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634-646.

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),* Jan 2005, 66-71. doi: 10.1037/0033-2909.131.1.66

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, *14*(1), 1-10.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139(1),* 39–50. doi:10.1016/j.neuroscience. 2005.07.002

Lange, G., Guttentag, R. E., & Nida, R. E. (1990). Relationships between study organization, retrieval organization, and general and strategy-specific memory knowledge in young children. *Journal of Experimental Child Psychology, 49(1),* 126-146.

Lehmann, M. & Hasselhorn, M. (2007). Variable memory strategy use in children adapted into a task learning behavior: Developmental changes and working memory influences in free recall. *Child Development, 78(4),* 1068-1082.

Logie, R., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*(3), 241-257.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, *29*(1), 10-17.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49(2),* 270-291. doi: 10.1037/a0028228

Melby-Lervåg, M., Lyster, S. A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin, 138(2*), 322.

Melkman, R., Tversky, B., & Baratz, D. (1981). Developmental trends in the use of perceptual and conceptual attributes in grouping, clustering, and retrieval. *Journal of Experimental Child Psychology, 31,* 470–486.

Miller, G. (1956). The magical number seven plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63(2),* 81–97.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science, 21(1)*, 8-14.

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.

Müller, U., Carpendale, J. I. M., Budwig, N., & Sokol, B. W. (Eds.) (2008). *Social life and social knowledge: Toward a process account of development*. New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.

Naus, M. J., & Ornstein, P. A. (1983). Development of memory strategies: Analysis, questions, and issues. *Trends in Memory Development Research*, *9*, 1-30.

Nevo, E., & Breznitz, Z. (2011). Assessment of working memory components at 6 years of age as predictors of reading achievements a year later. *Journal of Experimental Child Psychology, 109(1)*, 73-90.

Nutley, S. B., Söderqvist, S., Bryde, S., Humphreys, K., & Klingberg, T. (2009). Measuring working memory capacity with greater precision in the lower capacity ranges. *Developmental Neuropsychology*, *35*(1), 81-95.

Northwestern University (1996). Do your data violate one-way ANOVA assumptions? *Research Computing Services "Stat Guide"* Retrieved from: http://www.basic.northwestern.edu/statguidefiles/oneway_anova_ass_viol.html.

Rabinowitz, M. (1984). The use of categorical organization: Not an all-or-none situation. *Journal of Experimental Child Psychology*, *38*(2), 338-351.

Rawson, K. A, & Van Overschelde, J. P. (2008). How does knowledge promote memory? The distinctiveness theory of skilled memory. *Journal of Memory and Language, 58(3),* 646-668.

Romano, J., Wallace, T. L., Helmick, I. J., Carey, L. M., & Adkins, L. (2005). Study procrastination, achievement, and academic motivation in web-based and blended distance learning. *The Internet and Higher Education, 8(4),* 299-305.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, *126*(3), 211-227.

Salamé, P., & Baddeley, A. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24(4),* 263-265.

Sanefuji, M., Takada, Y., Kimura, N., Torisu, H., Kira, R., Ishizaki, Y., & Hara, T. (2011). Strategy in short-term memory for pictures in childhood: a near-infrared spectroscopy study. *Neuroimage*, *54*(3), 2394-2400.

Schelble, J. L., Therriault, D. J., & Miller, M. D. (2012). Classifying retrieval strategies as a function of working memory. *Memory & Cognition*, *40*(2), 218-230.

Senkova, O., & Otani, H. (2012). Category clustering calculator for free recall. *Advances in Cognitive Psychology*, *8*(4), 292-295.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2010). Does working memory training generalize? *Psychologica Belgica, 50,* 245–276.

Schleepen, T. M., & Jonkman, L. M. (2012). Children's use of semantic organizational strategies is mediated by working memory capacity. *Cognitive Development, 27(3),* 255-269.

Schmeichel, B. J., Volokhov, R. N., & Demaree, H. A. (2008). Working memory capacity and the self-regulation of emotional expression and experience. *Journal of Personality and Social Psychology*, *95*(6), 1526.

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high-and low-aptitude children. *Journal of Educational Psychology*, *81*(3), 306-312.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179-193.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: *Human Learning and Memory, 6(2),* 174.

Sullivan, L.E. (Ed.). (2009). The SAGE Dictionary of Social Research Methods. London, England: SAGE Publications, Ltd. doi: http://dx.doi.org.ezproxy.library.uvic.ca/10.4135/9781412972024.n1997

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *The Quarterly Journal of Experimental Psychology, 48(1)*, 108-124.

Tulving, E. (1962). The effect of alphabetical subjective organization on memorizing unrelated words. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 16(3),* 185-191.

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

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2013). Working memory capacity and retrieval from long-term memory: the role of controlled search. *Memory & Cognition*, *41*(2), 242-254.

Wass, S. V., Scerif, G., Johnson, M. H. (2012). Training attentional control and working memory – Is younger better? *Developmental Review, 32(4),* 360-387.

Wager, T., & Smith, E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience, 3(4),* 255–274. doi:10.3758/CABN.3.4.255

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of educational psychology*, *102*(1), 43-53.

Zelazo, P. D., & Müller, U. (2011). Executive function in typical and atypical development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp 576-603). Oxford, England: Wiley-Blackwell.

Appendices

Appendix A: Task instructions

Matrices instructions

Instructions:

“I’m going to show you some pictures of squares. Some squares are coloured red and some are white. Try your best to remember where all the red squares are. You need to look closely because you’ll only see the picture for a short time. After, you’re going to wait for me to give you the blank boxes and then you’ll point to where you think all the red squares were.”

Free recall instructions

Instructions:

Cards are displayed in four rows of five cards each, such that no two items form the same category are horizontally or vertically contiguous (see below before example). As each card is presented, the child names the item. In cases where the child does not name the item, the examiner will name it, and the child will repeat the name. During the study period, the child is shown that the cards can be moved, but no specific instructions are given.

“I am going to show you some pictures and you name them out loud as I put them down.” **(if child does not know an item’s label, experimenter label and ask child to repeat. Ask again after all items have been presented.)**

(after presenting all the cards) “Good, now I would like you to study them and try to remember as many of the items as you can. You can remember them in any order you like and you can also move these cards around however you like (demonstrate) if it helps you study. Take as long as you need to study and let me know when you are ready. When you are ready I will take the cards away and you will tell me all the ones you remember, okay?”

(remove cards, give child 3 minutes to answer. Encourage additional attempts with prompt, "Can you remember anymore?")

Rehearsal training phase A instructions

Instructions:

"So, [name of child], you might have noticed that I've been getting you to remember a lot of things. One way of helping people to remember things is to repeat them over and over again in a list either out loud or inside their head. I'll show you what I mean here (refer to the pictures on Powerpoint). So for example, if I were to say "car-hand-bee" over and over again, if I took away the pictures (change to blank screen), you'd probably still remember them, is that right? What were the items? Good. This will also help you remember the items if I ask you for them much later. I'd like you to use this way of remembering – by repeating them over and over again either out loud or inside your head, whichever you like. When you think you are ready, say, "ok!" and I will take the pictures away and ask you to tell me all the ones you remember."

Rehearsal training phase B instructions

Instructions:

"Good job [name of child]! So now that you've had some practice and you're very good at using this way of remembering, I'm going to make it a bit harder. This time I am not going to show you any pictures. I'm going to read you a list of items. Listen carefully and wait for me to finish the list. I'm going to read it twice to you. After, you're going to repeat those items out loud or inside your head, and keep doing that until you see this "Go" appear on the screen and then you'll tell me all the ones you remember, okay? (a Go is triggered on the screen after a 10-15 second delay.

Problem solving task instructions

Instructions:

[Refer to Figure 6]

(at bear's house) "Here is my friend Mr. Bear. Bear has a card for each of these pictures (point to items on list), but his cards have been stolen by two mean trolls named Grumpy and Grouchy. So [child's name], it's going to be your job to try and get back **all the cards on this list** for Bear. Can you do that? Let me show you where the trolls live."

[Refer to Figure 7]

(in Grumpy's room) "This is Grumpy and he has stolen some of Bear's cards and mixed them up with all of his own cards and put them inside his box. Now before you can open this box you need to pay the troll one coin. (point to the coin bank) Always pay the coin first. This is going to be true for any of the trolls you visit in any of the rooms, you always need to pay them one coin before you open the box, okay? After you pay the coin, then you can open the box (demonstrate). Look through the cards carefully, pick the ones you need, and put the rest of the cards back into the box, and close the box."

The same instructions and demonstration would be given to the child for the other troll in the green room before returning back to Bear's house.

(back at bear's house) "There are a few more rules I have to tell you about this game. You see this basket here? What does that say? (7) Good, this means that you are only allowed to have 7 cards in your basket at any time. You can have less, you can have 5, 6, and you can even have 7, but you can't have more. So [child's name] when you are collecting those cards from the trolls, make sure you count how many you plan to take away because you can only have up to 7.

If there are cards you still want, you can put the ones in your basket down first and get the new cards on another trip.

I also have to tell you that you only have 6 coins. You have a lot of cards to get back but only these coins! So what this means is that you need to **think and plan carefully** about the best way to get these cards back.

After you bring the cards back to bear's room, you're going to put them on his two shelves (point to 2 marked out rows). They need to go in the same order as the list (point to items on list in order) but here's the important thing: when you are collecting the cards from the two rooms, you can collect them in **any order you like.** (point to items on list in random) Whichever way you think is best. It's only when they go back on the shelves that they have to follow the list. If you accidentally get a card that is not on bear's list, that's going to be a mistake and that will go here (point to x box).

Now to win this game, you'll have to try and get back all the right cards on bear's list if you can, and try not to make mistakes. And you're going to do this while trying to use as few of the coins as you can, alright? You can use all the coins if you need to, but try to use as few of them as you can."

At this point the child is asked a few questions to confirm they understand the rules of the game:

"How many cards can you have in your basket at one time? (7)

Where do you put the pictures when you bring them back to bear's house? (on the line)

Where do the mistake cards go? (x box)

What do you have to give the trolls every time **before** you can open the box? (give a token)"

“I will be waiting for you here to see which cards you bring back. Before you start you might want to have a good look at the list, and then you can start whenever you are ready.”

Appendix B: Pre-test scoring forms and stimuli

Digit span forward

Item	Trial	Response	Score
1	2-9		0 1
	4-6		0 1
2	3-8-6		0 1
	6-1-2		0 1
3	3-4-1-7		0 1
	6-1-5-8		0 1
4	8-4-2-3-9		0 1
	5-2-1-8-6		0 1
5	3-8-9-1-7-3		0 1
	7-9-6-4-8-3		0 1
6	5-1-7-4-2-3-8		0 1
	9-8-5-2-1-6-3		0 1
7	1-8-4-5-9-7-6-3		0 1
	2-9-7-6-3-1-5-4		0 1
8	5-3-8-7-1-2-4-6-9		0 1
	4-2-6-9-1-7-8-3-5		0 1

Total = LDSF =

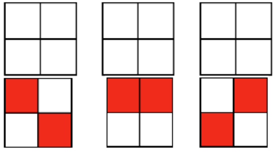
Digit span backward

Item	Trial	Response	Score
S	8-2		
	5-6		
1	2-1		0 1
	1-3		0 1
2	3-5		0 1
	6-4		0 1
3	5-7-4		0 1
	2-5-9		0 1
4	7-2-9-6		0 1
	8-4-9-3		0 1
5	4-1-3-5-7		0 1
	9-7-8-5-2		0 1
6	1-6-5-2-9-8		0 1
	3-6-7-1-9-4		0 1
7	8-5-9-2-3-4-6		0 1
	4-5-7-9-2-8-1		0 1
8	6-9-1-7-3-2-5-8		0 1
	3-1-7-9-5-4-8-2		0 1

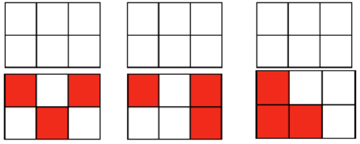
Total = LDSB =

Visual short term memory (Matrices) form.

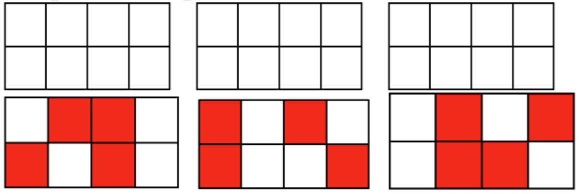
4-squares (2-span)



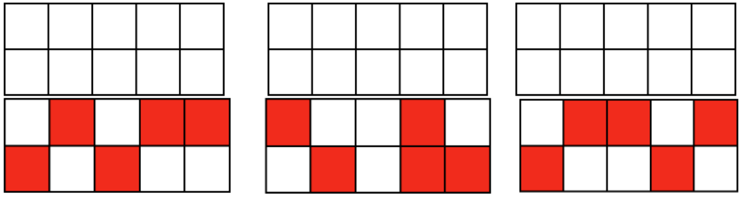
6-squares (3-span)



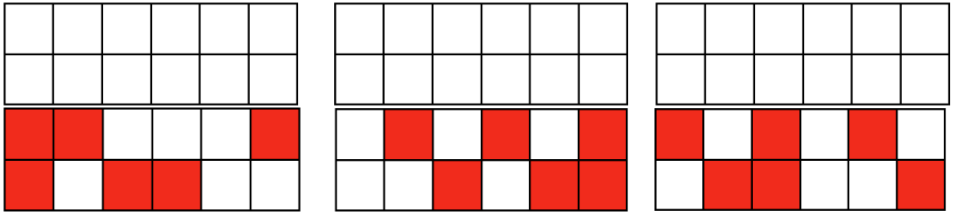
8-squares (4-span)



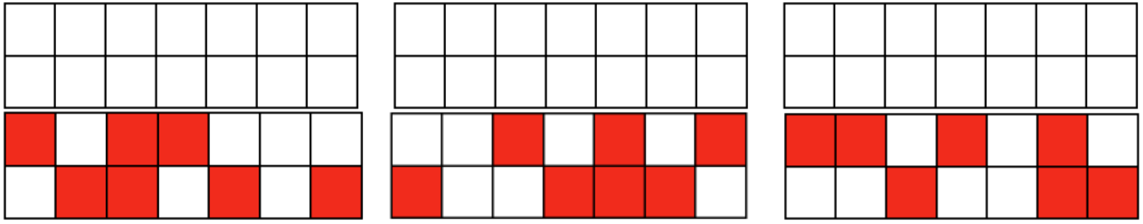
10-squares (5-span)



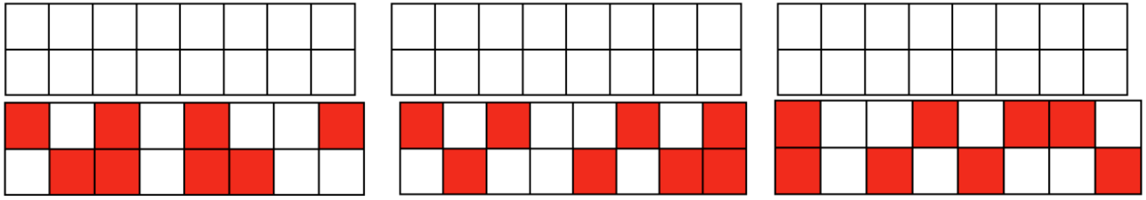
12-squares (6-span)



14-squares (7-span)



16-squares (8-span)



Free recall Pre-test

Correct = 1, Incorrect = 0
Categories: Fruit = 1, Insects = 2, Vehicles = 3, Furniture = 4
Participant Recall output (record in order of output):

	Item	Correct	Category		Item	Correct	Category
1	Banana	1	1	11	Car	1	3
2	Grapes	1	1	12	Plane	1	3
3	Orange	1	1	13	Bus	1	3
4	Cherry	1	1	14	Boat	1	3
5	Apple	1	1	15	Bike	1	3
6	Spider	1	2	16	Table	1	4
7	Ant	1	2	17	Chair	1	4
8	Bee	1	2	18	Lamp	1	4
9	Butterfly	1	2	19	Bed	1	4
10	Beetle	1	2	20	Sofa	1	4

Appendix C: Training tasks stimuli

S1. CATEGORY INCLUSION DECISIONS

Banana **(insects)**	Ant **(body parts)**	Fork **(vehicles)**
Bike **(furniture)**	Leg **(kitchen utensils)**	Knife **(hold liquid)**
Ant **(fly)**	Clock **(have legs)**	Boat **(wheels/land)**

Note: a group of images is displayed, correct answer is first word, category of other items is the second word in bold.

S2. TRAINED FREE RECALL (same as free recall pre-test)

REHEARSAL TRAINING LISTS

Visual + Verbal
Ant-eye-car Lamp-bike-spoon-grape Eye-boat-knife-bird-clock-drum Fly-shoe-bear-hat-train-house-ball
Verbal presentation only
Chair-lamp-apple Clock-cup-cherry-bus Fork-cherry-spider-drum-bee Knife-fly-train-clock-bed-lamp Train-bed-house-car-beetle-spoon

R1. READING from either Robert Munsch (younger kids) or A Series of Unfortunate Events (older kids)

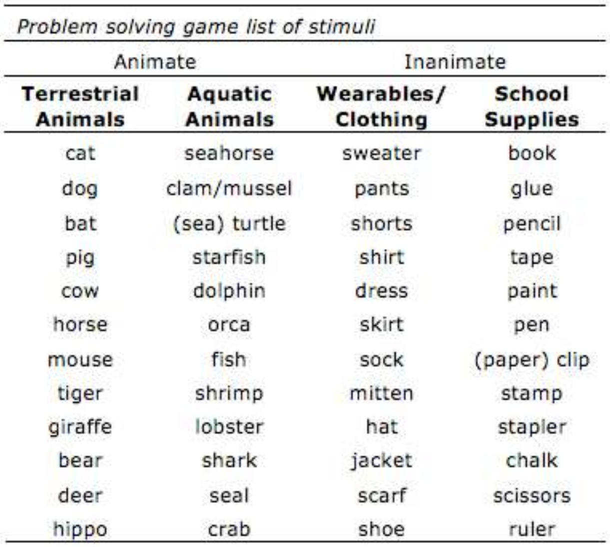
Appendix D: Post-test scoring forms and stimuli

PROBLEM SOLVING TASK
[**EXAMPLE**] TRIAL 1. (3 trials administered)

	Item	Correct	T	Return		Item	Correct	T	Return
1	cat	1	1		15	pen	1	4	
2	cow	1			16	scissors	0		
3	mouse	1			17	tape	1		
4	dolphin	1		X	18	stamp	1		X
5	seal	1	2		19	pig	1	5	
6	lobster	1			20	horse	1		X
7	starfish	1			21	shirt	1	6	
8	clam	1			22	scarf	1		X
9	seahorse	1		X	23				
10	skirt	1	3		24				
11	pencil	1			25				
12	dress	1			26				
13	book	1			27				
14	stapler	0		X	28				

Total /24 (correct) = 20/24
Total errors = 2
Tokens used: 6
Other notes: correct span per token = [4, 5, 4, 3, 2, 2]

Notes
Item: item depicted on retrieved card
Correct: 1 = yes, 0 = no
T: indicate the token number (/6) each time a new one is used
Return: mark x each time the child returns to bear's house



Problem solving game list of stimuli

Animate		Inanimate	
Terrestrial Animals	**Aquatic Animals**	**Wearables/ Clothing**	**School Supplies**
cat	seahorse	sweater	book
dog	clam/mussel	pants	glue
bat	(sea) turtle	shorts	pencil
pig	starfish	shirt	tape
cow	dolphin	dress	paint
horse	orca	skirt	pen
mouse	fish	sock	(paper) clip
tiger	shrimp	mitten	stamp
giraffe	lobster	hat	stapler
bear	shark	jacket	chalk
deer	seal	scarf	scissors
hippo	crab	shoe	ruler

Free Recall POST

Correct = 1, Incorrect = 0
Categories: Body parts = 1, Nature = 2, Instruments = 3, Kitchen utensils = 4
Participant Recall output (record in order of output):

	Item	Correct	Category		Item	Correct	Category
1	Leg	1	1	11	Guitar	1	3
2	Arm	1	1	12	Piano	1	3
3	Hand	1	1	13	Flute	1	3
4	Eye	1	1	14	Drum	1	3
5	Nose	1	1	15	Violin	1	3
6	Rock	1	2	16	Fork	1	4
7	Tree	1	2	17	Spoon	1	4
8	Leaf	1	2	18	Knife	1	4
9	Flower	1	2	19	Pan	1	4
10	Stick	1	2	20	Cup	1	4

Bear's lists

T1

T2

T3