Memory or attention? Understanding working memory in children.

by

Adrienne E. Eastwood

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2001

0-612-65235-1

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

This dissertation explores the construct of working memory (WM) in children, defined as the ability to concurrently remember and process information over brief periods of time. The research presented here had several goals with respect to children's working memory: 1) to develop tests of working memory that have adequate psychometric properties; 2) to ascertain whether working memory is distinct from short-term memory; and 3) to investigate the relative contributions of processing speed (PS), controlled attention (CA), and short-term memory (STM) in accounting for individual differences in working memory capacity.

To address these questions, tests thought to measure WM, STM, CA, and PS were administered to 119 normally functioning children between the ages of nine and thirteen. Two working memory tasks were modeled after the work of Daneman and Carpenter (1980), Engle, Carullo and Collins (1991), and Salthouse, Mitchell, Skovronek and Babcock (1989), that involved concurrent storage and semantic/computational processing of orally presented sentences/arithmetic calculations. The new WM measures were shown to have adequate internal consistency but inadequate test-retest reliability. CA was operationalized using the Stroop Colour and Word Test, the Trail-making Test, and commission errors on the Continuous Performance Test. STM was measured using the California Verbal Learning Test and the Semantic Categorization subtest from the Swanson Cognitive Processing Test. PS was assessed using the Visual Matching subtest from the Woodcock-Johnson tests of Cognitive ability, and the Symbol Search subtest from the Wechsler Intelligence Scale for Children, Third Edition.

Structural equation modeling techniques were used to investigate the relations between working memory and other cognitive abilities. The results indicated that WM is distinct from, though strongly correlated with, STM. Path models indicated that this correlation is largely a function of individual differences in controlled attention, which accounts for about half of the variance in the latent WM factor. Tests of PS and CA were found to best fit a one-factor solution. Because PS and CA were very highly correlated (.96, and therefore, indistinguishable) in the present sample, it was not possible to test predictions about how they would interact with each other in the prediction of WM. The implications of this result with respect to understanding individual differences in WM capacity are discussed. Overall, the results of the present study are consistent with Engle, Tuholski, Laughlin and Conway's (1999) model in suggesting that CA is a significant predictor of WM capacity. Indeed, when one accounts for CA, STM appears to add little to the prediction of WM capacity.

Acknowledgments

It is difficult to offer sufficient thanks to my supervisor, Richard Steffy, who has been my attentive and thoughtful mentor, enthusiastic cheerleader, and generous supporter for the last seven years. I never once doubted that he had my best interests at heart. Dick, working with you has been one of the main highlights of my graduate career and I consider myself blessed for having been steered in your direction.

I am grateful to the other members of my committee, G. Ernest MacKinnon and Jonathan Oakman, for going well beyond the call of duty in helping me to complete my dissertation as quickly as I had hoped. Thank you for always being interested in my work, for your thoughtful comments, and for your encouragement.

I am indebted to William Corning, whose ideas about children's learning difficulties have been formative in my intellectual development. I am thankful that he sparked my interest in studying children's cognitive abilities and that he provided me with the opportunity to study children with learning and behaviour disorders.

Thanks are owed to all the children who participated in this study and their parents. I am especially thankful to my research assistants, Lynne Steffy and Chanthalone Vongphrachanh, for their enthusiasm and commitment to this project. I never could have accomplished this without their proficient help. A big thank you goes to the past and present members of the WATPEN research team for their collegiality and interest in my work – Gina Argitis, Karen Baker, Patricia Bowers, Rebecca Cohen, Greta Cramer, Alexandra Gottardo, Galit Ishaik, and Michaela Williams. A special thank you is also owed to Kathy Blom – know that your care for me and support have been greatly appreciated.

I wish to acknowledge the Natural Sciences and Engineering Research Council of Canada, for supporting me financially with a Doctoral Fellowship.

Many friends and fellow graduate students have stuck by me and encouraged me during the completion of this thesis – I am grateful to so many people too numerous to list here. I would like to especially thank Robyn Wiebe, who has been my steadfast friend throughout graduate school. Thank you Robyn for always hanging in there and helping me keep a good perspective on things. I have been blessed by your fun spirit, your sense of adventure, your generosity, and your excellent taste in, well... everything.

I am thankful for the continued support given to me by both my family and John's family. You have been there to provide encouragement and assistance in every way possible, including meals! Thanks for believing in me, for understanding when work has taken too much of my time, and most importantly, for loving me unconditionally.

I am immensely grateful for the gifts of love, time, patience, thoughtfulness, and commitment so generously given to me by my husband, John Eastwood. Your ability to think both creatively and critically is remarkable, and this work is much improved for all the time you spent talking about it with me. I thank you, most of all, for your quiet, unwavering, enduring, selfless love.

Table of Contents

List of Tables

List of Figures

Memory or attention? Understanding working memory in children.

Memory, an ability that is central to all aspects of people's daily functioning, has fascinated researchers for over a century. The study of human memory has been heavily influenced by various versions of the dual-store model, also termed the modal model because of its prominence during the 1950s and 1960s. This model held that memory involves 1) sensory registers that receive information through the senses, 2) a short-term store, and 3) a long-term store (Neath, 1998). In Atkinson and Shiffrin's (1968) model, the information chain begins with sensory registers that receive information from each sensory modality, and that store the physical properties of stimuli for very brief periods of time (under one second). Information is then transferred into the limited-capacity short-term store, where control processes such as rehearsal, coding and retrieval are used to temporarily maintain information. The purpose of the short-term store is to keep information active long enough so that it can be used, or to transfer information to the long-term store for future use (Atkinson & Shiffrin, 1968).

Working memory (WM) is a more recent conception designed to better capture the ability that people have to temporarily store information for immediate, on-line use in learning, reasoning, and comprehension. Earlier dual-store models considered WM to be synonymous with the short-term store (Baddeley, 1990, 67). It was thought that when an individual's short-term store was fully taxed, there would be no further resources available for higher-order processing, because of a trade-off between the amount of information that can be maintained in the short-term store and the resources available to think and problem-solve. Thus, according to dual-store models, the amount of information that can be stored over short periods of time is the limiting factor in the information-processing system (Baddeley, 1990, 59-68).

More recent models of memory, described in greater detail below, have moved away from a unitary conception of the short-term store toward multi-faceted models that include WM as a distinct and more specialized capacity for temporary storage. WM is often measured using complex span tests, first developed by Daneman and Carpenter (1980), that place demands on individuals to concurrently process and remember information. On such tasks, individuals are asked to read and comprehend sentences, while remembering the last word of each sentence. Increasing numbers of sentences are presented, and 'span' is usually defined as the maximum number of words that can be recalled while also processing the sentences correctly (comprehension questions are used to check the accuracy of processing). Complex span tests involving arithmetic operations and spatial tasks have also been developed.

WM has been found to be an important predictor of individual differences in the academic achievement and fluid reasoning of both adults (e.g., Daneman & Carpenter, 1980; Engle, Tuholski, Laughlin & Conway, 1999) and children (e.g., de Jong & van der Leij, 1999; Kail & Hall, 2001; Swanson, 1993; Swanson; 1994). Theories of memory attempt to account for individual differences in WM capacity, and explain how WM and STM are related to each other. However, because of some contradictory findings in the literature on children's memory, the relation between WM and STM, and the subcomponent capacities that underlie WM, require further investigation in children. The present study had two goals with respect to children's WM: 1) to evaluate whether WM and STM could be measured distinctly in children and to determine what tests best measure these constructs, and 2) to investigate the nature of the subcomponent abilities that predict individual differences in WM capacity.

What is working memory?

Baddeley and Hitch's model of WM, first articulated in 1974, has been influential in the development of multi-faceted models of temporary storage. The most recent revision of Baddeley's WM model describes subsidiary slave systems (the phonological loop and visuo-spatial sketchpad) that serve to temporarily store information in an active state via rehearsal and other sustaining processes, and a separate central executive component that controls and regulates the system but does not store information (Baddeley & Logie, 1999). The processing and storage of information is thought to rely on separate resources (Baddeley & Logie, 1999, 39). According to Baddeley's model, WM derives from the collective contribution of the entire operating system, including the processing functions governed by the central executive, and short-term storage is carried out by the subsidiary slave systems.

Baddeley's model focuses primarily on the role of WM in higher cognition. He recently defined WM as "those functional components of cognition that allow humans to comprehend and mentally represent their immediate environment, to retain information about their immediate past experience, to support the acquisition of new knowledge, to solve problems, and to formulate, relate, and act on current goals" (Baddeley & Logie, 1999, 28-9). The phonological loop and visuo-spatial sketchpad components of the working memory model are well-specified by Baddeley and have been researched intensively, investigating for example how information is lost, how much information can be maintained, and the role of rehearsal processes in each of the storage systems (see Baddeley, 1990 and Baddeley & Logie, 1999). However, the nature and functions of Baddeley's central executive were previously less well-specified, making it difficult to formulate and test hypotheses about the central executive. Such criticisms have led Baddeley to move away from a 'ragbag' conceptualization of the central executive, and to specify its functions and relations to the phonological loop and visuo-spatial sketchpad in greater detail. For example, he now suggests that the central executive focuses attention, switches attention, and activates information from long-term memory (Baddeley, 1996).

Engle and his colleagues have also sought to specify an executive component of WM and have come to a precise definition of its function and capacity (e.g., Engle, Kane & Tuholski, 1999). Influenced by the work of Cowan (e.g., 1999), Engle and his colleagues suggest that WM requires the application of controlled attention (CA) to information that is held in short-term memory (STM). Information in STM consists of memory traces (e.g., digits, words, ideas) that have been activated above a certain threshold. This level of activation is thought to be achieved through the use of domain-specific rehearsal, grouping and/or coding strategies. Individual differences in STM capacity reflect domain-specific functions, that is, an individual could be skilled at remembering verbal information but not visual information. Information in STM is thought to be subject to decay and interference, and corresponds to Baddeley's phonological loop and visuo-spatial sketchpad. STM is described as a storage component requiring rehearsal and coding strategies, but not involving CA functions to the same degree that is required in WM. Engle and his colleagues have postulated that when individuals use WM, information is maintained at a higher level of activation by CA, which serves to inhibit distracting information while maintaining task goals. CA is thought to be domain-free, that is, not dependent on the specific features of the task or type of information to be remembered (Engle, Kane & Tuholski, 1999). In this manner, controlled attention is comparable to Baddeley's central executive. Thus, Engle and his colleagues have depicted the working memory system using the equation WM = STM + CA, which is meant to convey that WM consists of information that is held in STM, and that is maintained in the focus of attention, despite interference. Furthermore, Engle and his colleagues postulate that the limiting factor on working memory capacity is determined by CA, noting: "We assume that "working memory capacity" is not really about storage or memory per se, but about *the capacity for controlled, sustained attention in the face of interference or distraction*" (Engle, Kane & Tuholski, 1999, 104; italics in original). Individual differences in WM performance, therefore, are thought to be determined by individual differences in CA, rather than by STM storage capacity.

Are working memory and short-term memory distinct constructs?

Clarifying the relationship between WM and STM has been a focus of recent investigations (Engle, Tuholski, Laughlin & Conway, 1999; de Jonge & de Jong, 1996; Kail & Hall, 2001; Swanson, 1993). Research attempts to distinguish WM and STM have, in a number of ways, benefitted from the application of structural equation modeling (SEM) techniques to correlational data. Among its advantages, SEM methods involve the use of more than one test to operationalize constructs of interest. Assuming that no one test is a perfect measure of any construct, SEM allows investigators to evaluate constructs by sampling abilities in a way that moves beyond the specific demands associated with any one test. When different tests are used as measures of the same latent factor in SEM, the systematic variance that is shared by those tests reflects what is common to them – presumably a relatively more 'pure' version of the latent construct. Moreover, the use of more than one test to measure a construct allows researchers to estimate and control for the effects of random measurement error that would otherwise bias the correlations between variables. This is accomplished because SEM methods also estimate whatever variance is *not* held in common by a set of tests, generally referred to as measurement error, including both random measurement error (due to random factors such as whether a participant was tired that day or not putting in his/her best effort), as well as systematic variance in a test that is not shared with the other tests, often due to method variance (Kline, 1998, 58). The application of SEM techniques also allows models concerning the relation between constructs to be tested and compared to competing models.

A recent study using SEM techniques has provided support for the hypothesis that WM and STM are distinct constructs in adults. Engle, Tuholski, Laughlin and Conway (1999), administered tests of WM, namely, Operation Span, Reading Span, and Counting Span (analogues of the Daneman & Carpenter [1980] Sentence Span Test) to a sample of adults. Tests of STM were also administered, namely, Forward Span of similar (rhyming) words, dissimilar words and Backward Span of dissimilar words. Forward span tests are often referred to as 'simple span' tests, because no concurrent processing is required. All memory tasks were correlated with each other (*r*s ranging from .31 to .59), but WM tests tended to correlate more strongly with each other than with the STM tests, and STM tests tended to correlate more strongly with each other than with the WM tests. The results of confirmatory factor analyses, comparing a one-factor (i.e., WM = STM) to a two-factor (i.e., WM ≠ STM) model, demonstrated that the two-factor solution provided the best fit to their data, although the latent WM and STM factors correlated strongly with each other ($r = .68$; Engle et al., 1999).

The distinctiveness of WM and STM is less well established in children than in adults, and there is less consensus regarding which types of tests would best measure WM and STM in children. Indeed, some research has provided contradictory results regarding whether WM and STM are distinct constructs in younger populations. For example, de Jonge and de Jong (1996) administered a battery of simple span tests (Word and Digit Span), complex span tests (Reading Span, Computation Span), and the locally developed Star Counting Test (thought to measure activation and inhibition processes involved in working memory), to Dutch children in grades four through six (between the ages of approximately nine and twelve). On the Star Counting Test, children were presented with a display that contained a number in the top left corner, followed by stars interspersed with plus and minus signs in the remaining rows and columns. Children were asked to count the stars from left to right and top to bottom starting from the value of the number in the first row, but they were required to change the direction of counting when they encountered a minus sign (thus instructing counting backward) or a plus sign (forward counting). On the second part of the test, the meaning of the plus and minus signs were reversed, such that a minus sign would be indicative of forward counting. These authors hypothesized that simple span tests, presumed to measure STM, would be distinct from the complex span tests and the Star Counting Test, that were presumed to measure WM. However, their results indicated that simple and complex span tests were significantly intercorrelated with each other; indeed, the correlation between simple and complex span tests ($r = .23$ - $.34$) was not noticeably different from the correlation between the Reading Span and Computation Span tests ($r = .28$). Although confirmatory factor analyses found that the best fitting model to these data was a two-factor solution (with the inter-factor correlation moderately strong at .42), simple and complex span measures loaded *together* on the first factor, whereas the two subtests of the Star Counting Test loaded on the second factor (de Jonge & de Jong, 1996). Because the second factor consisted solely of two subtests from the same test, this factor likely reflects something very specific to the procedures used, and therefore it cannot be easily inferred that one factor measures STM and the other WM. Nonetheless, these results leave open the possibility that WM and STM may not be distinct constructs in children, or alternatively, that simple and complex span tests may not be the best measures of WM and STM in children.

In contrast, Kail and Hall (2001), using simple and complex span tests, found that WM and STM could be distinguished in a sample of eight to twelve-year-old children. In two studies, WM was measured using Reading Span, Listening Span and Least Number Span (all complex span tests), and STM was measured using Forward Digit Span, Letter Span and Word Span (simple span tests). Although all memory tests were correlated with each other (*r*s ranging from .19 to .47), complex span tests were more strongly related to each other than they were to simple span tests, and vice versa. A confirmatory factor analysis indicated that a two-factor solution provided the best fit to the data in both studies; the inter-factor correlation between WM and STM in study one was .32, in study two, .36

(Kail & Hall, 2001). Not only do the Kail and Hall (2001) results suggest that WM and STM can be measured distinctly in children using simple and complex span tests, but the correlation between the two factors was quite low.

A related issue, introduced above, pertains to which types of tests best measure WM and STM in children. If subsequent research using simple and complex span tests fails to differentiate WM and STM in children, it is still possible that WM and STM might be distinct abilities because different tests may be required to measure these constructs in children from those used in studies of adults. Insofar as children's memory skills are less well-developed than those of adults, it is possible that rehearsal and other memory strategies used to maintain information in STM may be less automatic; consequently, simple span tests may be more attention-demanding for children than for adults (Engle et al., 1999). If this is true, then simple span tests may reflect WM in children. The results of de Jonge and de Jong (1996) could be seen as consistent with such a notion; loadings of simple and complex span tests on a single factor could reflect a shared tendency of those tests to require controlled attention processes.

In testing the distinctiveness of WM and STM in children, the present investigation complements previous research by attempting to expand the repertoire of tests used to measure these constructs. The inference that 'WM' and 'STM' are distinct is viable only as additional tests of these constructs continue to demonstrate convergent and divergent validity. Thus, in the present study, Backward Digit Span (BDS) was used as an indicator of WM, in addition to analogues of Daneman and Carpenter's (1980) complex span test. In BDS, digits are recalled in the reverse order to which they were presented – this additional transformation is thought to operationalize the additional processing demands that are characteristic of WM tests (Daneman & Merikle, 1996). In addition, free recall of sets of semantically related words was used to measure STM in addition to a traditional simple span task (i.e., Forward Digit Span).

How can individual differences in working memory be understood?

A second major question in this thesis addressed the source of individual differences in WM capacity. Based on previous research described below, the present study attempted to examine the predictive validity of two important subcomponents – controlled attention and processing speed, as well as to address the relation between these constructs.

Predicting working memory – the role of controlled attention. According to Engle and his colleagues, individual differences in WM are thought to be related to controlled attention (CA), reflecting the ability to maintain activation of task-relevant information in the face of distraction or interference (see for example, Conway & Engle, 1994; Rosen & Engle, 1998). In order to explore this theory, Engle and his colleagues (1999) sought to test the hypothesis that the CA component of WM would account for the relation between WM and fluid reasoning. Fluid reasoning (gF) refers to the ability to solve novel problems, is thought to be nonverbal, relatively culture-free, and is often measured using matrices and figural analyses (Engle et al., 1999). A strong relationship between WM and gF has been documented (e.g., Kyllonen & Christal, 1990); Engle and his colleagues have hypothesized that individual differences in CA underlie this relation. To investigate this possibility, Engle et al., (1999) applied SEM techniques to model the relations between latent factors representing WM, STM, and gF. Although CA was not measured directly, it was reasoned that if the variance shared between WM and STM tasks was statistically removed from the WM latent factor, then the remaining variance would theoretically correspond to CA, and would be the driving factor behind the relationship between WM and gF.

Confirming this position, Engle et al., (1999) found that WM was a strong predictor of gF (path coefficient = .59, $p < .05$), whereas STM did not predict gF (-.13, n.s.). Constraining the path from STM to gF to zero did not result in a significant loss of model fit, providing confirming support that STM was not a unique predictor of gF. Engle et al., (1999) also attempted to evaluate whether CA would account for the strong association between WM and gF. In this SEM, STM and WM were modeled to account for variance in a "common" factor. According to Engle et al.,'s (1999) theory, this common factor should consist of variance due to STM – because STM is thought to be what is shared by simple and complex span tests. After controlling for this "common" factor, the WM residual was hypothesized to consist solely of CA, and the residual was found to be strongly associated with gF (.49). In contrast, the correlation between the STM residual (thought to represent only error – because all of STM-related variance should logically be included in the "common" factor) and gF was non-significant. Engle and his colleagues concluded that their findings were consistent with the notion that limits on WM are a function of individual differences in CA ability.

The present study sought to expand on this work by attempting to measure CA directly, and then evaluate its relations with WM and STM in a sample of children aged nine to thirteen. In this way, the hypothesis that individual differences in CA predict WM capacity to a greater extent than short-term storage capacity could be directly evaluated. The choice of tests to measure CA in the present study relied upon Engle et al.'s description of situations in which it is required:

> "Controlled attention is necessary: a) when task goals may be lost unless they are actively maintained in working memory; b) where actions competing for responding or response preparation must be scheduled; c) where conflict among actions must be resolved to prevent error, d) where there is value in maintaining some task information in the face of distraction and interference; e) where there is value in suppressing task irrelevant information; f) where error monitoring and correction are controlled and effortful; and g) when controlled, planful search of memory is necessary or useful" (Engle et al., 1999, 312).

The three tests chosen to measure CA included the Stroop Colour and Word Test (Golden, 1978), the Trail-making Test from the Halstead-Reitan Neuropsychological battery (Reitan, 1958), and commission errors from the Continuous Performance Test (Gordon & Mettelman, 1988). All three tasks (described in greater detail in the Method section) require selective attention to target information while inhibiting distracting information, and were thus seen to exemplify the criteria described in the above quote.

Predicting working memory – the role of processing speed. Processing speed (PS), defined as the rate at which an individual can take in information and produce a simple response without performing any problem-solving operation on that information, has been measured using a variety of tasks. As noted by Salthouse (2000), PS has been assessed using measures of psychomotor speed such as finger tapping rates or simple reaction time to target visual stimuli. Along with psychomotor speed tests, PS has been measured with perceptual speed tests, using timed paper-and-pencil tasks requiring the matching of visual stimuli, or the search for target stimuli. Both types of PS tests are related to each other and to more complex cognitive processes (Salthouse, Fristoe, McGuthry & Hambrick, 1998). There is a substantial body of research relating individual differences in processing speed (PS) to memory capacity, showing that more rapid processing is associated with better retention of information. Briefly, it is thought that slow processors have weaker retention because stimuli are more likely to decay from memory when participants take greater amounts of time to process the information (Towse, Hitch & Hutton, 1998).

Although the relation between PS and retention/recall has been widely investigated, particularly in older adults, the relative contribution of PS to WM versus STM, is contentious. Some investigators (e.g., Kail & Hall, 2001) have hypothesized that PS may be an *especially* important predictor of WM, because the information to be retained in WM is subject to interference to a greater extent than information held in STM (italics added). Children with faster PS may be better able to handle concurrent demands for processing and recall. Kail and Hall (2001) found in two studies that PS, as measured by the Visual Matching and Cross-out subtests from the Woodcock-Johnson Tests of Cognitive Ability – Revised, significantly predicted WM as measured by complex span tasks. In one of those studies the associations between PS and STM and between PS and WM were equivalent, but, in the second study, PS was not found to be a significant predictor of STM (Kail & Hall, 2001). Thus, although the relation between WM and PS was significant in both studies, the specificity of this association remained unclear. The findings from another investigation have called into question the predictive power of PS in accounting for individual differences in WM capacity. In a study contrasting poor and skilled readers, de Jong (1998) found that the groups differed with respect to WM capacity, but these group differences could not be accounted for by differences in PS. Therefore, the present study sought to evaluate whether PS would contribute significantly to the prediction of WM, and secondly, whether this contribution would be greater than that of PS to STM.

The relation between controlled attention and processing speed

Because both CA and PS have been hypothesized to be predictors of WM capacity, one question remaining is how these constructs are related to each other in mutually predicting WM capacity. Kail and Hall (2001) postulated that faster processing may facilitate WM by improving an individual's ability to direct attention between the competing demands of the complex span task used to index WM. The model implied by such an hypothesis suggests that PS influences CA, which, in turn, influences WM.

Only a few studies have investigated attention-related and processing speed constructs simultaneously in the prediction of working memory – and these studies offer partially conflicting results. Salthouse and Meinz (1995) were interested in the ability of prepotent response inhibition (arguably a function of CA) to predict age-related differences in adult WM; they were also interested in how PS would affect the relation between inhibition and WM. They measured PS using tests of perceptual and psychomotor speed (including rapid same-different judgements about letters and patterns, and digit symbol-type tasks) and they assessed prepotent response inhibition using a number of procedures including the Stroop Colour and Word Test. Their results indicated that both inhibition and PS measures explained considerable variance in the prediction of age-related WM differences. However, age-related variance in inhibition was shared closely with age-related variance in PS (84.8%), suggesting that these two constructs both measured the same functions.

de Jong and Das-Smaal (1993) conducted an exploratory factor analysis (principal components analysis with varimax [oblique/correlated] rotation) of a set of STM, WM, attention and PS tests (administered to children). The tests included Forward Digit Span, Backward Digit Span, a test of reading speed, the Stroop Colour and Word Test, a Dutch version of the Rey Auditory Verbal Learning Test, a word association test, the Trail-making Test, a pattern cancellation test, and Digit Symbol Substitution. Inconsistent with Salthouse and Meinz's (1995) results, de Jong and Das-Smaal (1993) found that factors that appeared to reflect CA and PS were distinct, although strongly correlated ($r = .61$). More specifically, PS tests (Digit-Symbol Substitution and pattern cancellation) loaded on a *different* factor than did the Stroop Colour and Word Test. However, they found that PS tests loaded on the same factor as the Trail-making Test, another test that involves CA because of its requirement to switch between competing task demands.

The present study

As described thus far, two general questions were addressed in the present study – first, are WM and STM distinct cognitive abilities in children, and second, how can individual differences in WM capacity be understood and predicted? To address these questions, tests thought to measure WM, STM, CA, and PS were administered to a sample of normally functioning children between the ages of nine and thirteen. SEM techniques were used to address the questions raised.

With respect to the first question, if WM and STM are indeed distinct constructs, then confirmatory factor analyses should demonstrate that a two-factor solution provides a better fit to the memory test data than a one-factor solution. Further, it was expected that tests that involved concurrent storage and controlled processing demands would best represent the WM factor, whereas tasks that involved only storage demands would tap the STM factor. To this end, two tests making concurrent demands on remembering and controlled processing were developed, modifying procedures from previous research (Daneman & Carpenter, 1980; Engle, Carullo & Collins, 1991; Salthouse, Mitchell, Skovronek and Babcock, 1989; Turner & Engle, 1989). It was expected that these tasks would demonstrate adequate test-retest reliability and internal consistency, and that they would intercorrelate with each other and with another hypothesized test of working memory (Digits Backward from the WISC-III) to a greater degree than they would correlate with tasks thought to measure STM.

In order to investigate how individual differences in children's WM can be understood and predicted, SEM techniques were also used to investigate how PS, CA, and STM predict performance on tests of WM. According to previous research, individual differences in WM were expected to be better predicted by CA than by STM (Engle et al., 1999). According to the results of Kail and Hall (2001), PS should also make a significant contribution to the prediction of individual differences in WM, although this contribution might possibly operate through CA.

Method

Participants

Children were recruited from a Summer day Camp for children, and from lists of children who had participated in a previous, unrelated study. A total of 123 children (71 boys and 52 girls) participated in approximately three hours of testing, conducted over two testing sessions scheduled three to six weeks apart. Children completed tests measuring WM, STM, CA, and PS, as well as tests of other cognitive abilities and reading comprehension not reported on here. There were 20 nine-year-olds, 32 ten-year-olds, 21 eleven-year-olds, 26 twelve-year-olds, and 22 thirteen year olds. One seven-year-old child and one eight-year-old child, who were the younger siblings of children in the study, also participated. All children were given a $15 honorarium.

Because relations between the cognitive abilities to be tested in the present study might be different in normally functioning samples, compared to samples of children with learning disorders, it was important to ensure that the present sample consisted of children without significant learning difficulties. Children were initially screened for diagnosed learning, emotional and behavioural disorders through a phone contact with parents, who were interviewed regarding their child's adjustment and appropriateness for the study.

While the child completed testing, parents completed the Checklist of Functioning (CoF) scale, on which they rated their child across achievement domains (including oral language, reading, written language, mathematics and attention to assigned tasks) and behaviour domains (including attention to assigned tasks, physical coordination, self-esteem, emotional maturity, and behavioural conduct) (Steffy, personal communication). Parents indicated on a four-point scale whether their child's development in any of these areas caused either: distinct concerns requiring professional intervention (rated 4); mild concerns (rated 3); no concerns about functioning (rated 2); or that their child was stronger than others his/her age (rated 1). The reliabilities of the whole scale (alpha = .88), the academic subscale (alpha = .90) and the behaviour subscale (alpha = .83) of the CoF were found to be satisfactory in a previous investigation (Eastwood & Steffy, submitted for publication). In the present study, the results of reliability analyses found that alpha was .79 for the whole scale, .77 for the academic subscale, and .75 for the behaviour subscale, providing further evidence that the CoF has adequate psychometric properties.

Noteworthy difficulties were identified during and after testing in four cases, and these four children were subsequently excluded from the sample. One of these children received the diagnoses of Attention-Deficit/Hyperactivity Disorder, and Expressive/ Receptive Language Disorder approximately four months after testing had been completed. The other three children were noted by the examiner to have obvious and significant difficulties reading (decoding words) during the testing sessions (on tests that were not used in analyses presented here). Two of these three children also scored below the 25^{th} percentile on a test of reading comprehension, and had been rated on the CoF as having significant difficulties (rated '4') with all three of reading, writing, and attention to assigned tasks. The third child was reported to have had a significant history of reading difficulties (four years behind grade level) during the phone interview, and scored well below the 25^{th} percentile on a test of phonological awareness.

The final sample of 119 participants consisted of children free from significant learning disorders but still heterogeneous with respect to cognitive ability and achievement. For this sample, mean CoF ratings and standard deviations were as follows: Language = 1.48 (.57); Reading = 1.45 (.62); Written Language = 1.79 (.76); Math = 1.59 (.67); Attention to assigned tasks = 1.92 (.74); Physical coordination = 1.77 (.53); Self-esteem = 2.01 (.75); Emotional maturity = 1.82 (.61) and Behavioural conduct = 1.85 (.67). Most children were rated '1' (strong), '2' (no concern), or '3' (mild concern) on all academic subscales. However, one child was rated '4' (distinct concern) on the reading subscale, three children were rated '4' on the writing subscale, and three children were rated '4' on the attention to assigned task subscale. These seven children were not excluded from the sample, because their parents had not reported any diagnosed disorders during the phone interview, and because no difficulties were noted by the examiner on reading tests during testing sessions. All seven children had reading comprehension scores in the average range or better, and none were outliers on any of the measures reported here.

Measures

Working memory (WM). Children were administered three tests of working memory, involving either words or digits as stimuli, that were thought to encompass both the processing and storage demands involved in working memory (Daneman & Merikle, 1996). 1) The **Backward Digit Span** (BDS) test from the Wechsler Intelligence Scale for Children, Third Edition (WISC-III; Wechsler, 1991), in which the child is asked to repeat back strings of digits in the reverse order in which they were presented. Two trials are presented for each digit string length, and one point is given for every correct trial prior to the point of discontinuation. 2) A **Sentence Span Test** (SST) was developed for use in this study, based upon the work of Daneman and Carpenter (1980) and adapted for use with children, following the specifications outlined by Engle, Carullo and Collins (1991). In this task, children were required to listen to a series of two to five sentences, with the goal of recalling, in order, the final word (a concrete noun within the children's expected vocabulary) of each sentence. Sentences were presented from an audio cassette at a comfortable pace. After the children recalled the final words, a comprehension question relating to one (randomly selected) sentence in each set was administered as a check to confirm that participants were actively processing the meaning of the sentences. Children only received credit for recall if the comprehension question was answered correctly. Five trials at each of four difficulty levels – two, three, four, or five sentences were used with a discontinuation rule. Children received one point for each level on which 3, 4, or 5 trials were correct, an extra .5 point if two trials were correct, and .25 point if one trial was correct at a given level. 3) A **Computation Span Test** (CST) was developed on the basis of a task used by Salthouse, Mitchell, Skovronek, and Babcock to measure working memory in adults (1989). The present task was also similar to the Operation Digit task used by Turner and Engle (1989). In the CST, children listened to a series of simple arithmetic problems (involving addition or subtraction), solved each problem out loud immediately after it was presented, and subsequently recalled the answers in the order in which the problems had been presented. The arithmetic problems were read at a steady and deliberate rate (one utterance per second), with a pause between each problem long enough for the child to respond. Although the children were prevented from using their fingers to keep track of the answers to be recalled, they were permitted to use their fingers to solve the problems, if they wished. As in the sentence span test, there were five trials at each of four levels (with two, three, four, or five arithmetic problems in each item), and children received one point for each level on which 3, 4, or 5 trials were correct, an extra .5 point if two trials were correct, and .25 point if one trial was correct at a given level.

Short term memory (STM). Children were administered two tests involving either words or digits as stimuli, that were thought to encompass only the storage demands involved in short- term memory (Daneman & Merikle, 1996). 1) The **Forward Digit Span** (FDS) test from the WISC-III (Wechsler, 1991), in which the child was asked to repeat back strings of digits in the same order in which they were presented. Two trials were presented for each digit string length, and one point was given for every correct trial prior to the point of discontinuation. 2) The **California Verbal Learning Test, Children's Version** (CVLT; Delis, Kramer, Kaplan, Ober & Fridlund, 1994), in which 15 'shopping list' words were presented to the child, and the child was asked to recall as many as possible in any order. Words were semantically related according to several categories (e.g., fruits, clothing). Scores were obtained from the total number of words recalled across each of five trials (list A), and from the number of words recalled from a second shopping list (list B) of fifteen words. The CVLT is thought to be a measure of verbal short-term memory that places fewer demands upon attentional processes than do working memory tests (Groth-Marnat, 1997, 577).

Other memory tests. Two memory tests from the Swanson Cognitive Processing Test (Swanson, 1996) were administered. 1) On the **Semantic Categorization** (SC) subtest, the child was presented with a list of words that are organized such that a category name was always followed by exemplars of the category (e.g., animal, donkey, cat, colour, blue, yellow). The child was asked to recall the words by stating first a category, then all the exemplars of that category, before recalling the next category and associated words. The child could recall the words in the same order in which they were initially presented, but also received full credit if the categories and exemplars were given back in a different order, as long as a category name was given first, followed by all the appropriate exemplars (e.g., *colour*, yellow, blue, *animal*, donkey, cat). After completing recall, the child was asked a recognition question (for example, "Which word was presented: brown or yellow?") that had to be answered correctly in order to receive credit for his/her recall on that item. 2) On the **Semantic Association** (SA) subtest, the child was presented with a list of words in a mixed-up order, and then was asked to recall them back in groups according to which category they belonged to. The categories were *not* initially given to the child, but needed to be inferred by the child. After completing recall, the child was asked a recognition question (of the same form as was used in SC) that needed to be answered correctly in order to receive credit for recall. Swanson (1996) indicates that these tests measure 'dynamic working memory', however, some features of these tasks, particularly SC, suggested that they might not be good measures of WM. First, the post-recall recognition question, which Swanson (1996) reported would provide an extra processing demand, did not seem as though it would interfere with recall. Second, on the SC subtest, children were permitted to recall words in the same order in which they were heard, thereby decreasing demands placed on the child to actively control his/her attention in order to alter the order of words during recall. Further, the provision of category names may have reduced the controlled processing necessary to recall the words. Because of these task characteristics, no advance predictions were made about which memory factor SC and SA would measure.

Controlled attention (CA). Three tests were used to evaluate CA capacity that were thought to encompass the CA characteristics described by Engle et al., (1999). Although these tests were purposely chosen to involve varied task demands, they shared a number of common characteristics. For example, in each task, children had to actively maintain task goals, had to inhibit task irrelevant information, had to resolve conflict between competing actions, and had to actively monitor their responses in order to prevent errors. 1) In the **Stroop Colour and Word Test** (Stroop; Golden, 1978), the child was presented with a card of 100 colour words (red, green, blue) that were printed in an incongruent colour. The child was asked to name the ink colour each word was printed in, while inhibiting the name of the colour that the word actually spelled. The child was given 45 seconds to name the ink colour of as many words as possible. In order to control for the speed at which children could name colour patches, children were first asked to name as many colour patches as possible in 45 seconds, and the number of patches named was subtracted from the number of colours named on the interference trial. The Stroop requires the resistance of interference to a habitual response, and has been reported to be useful in screening for attention difficulties (Groth-Marnat, 1997, 539). 2) Parts A and B of the **Trail-making Test** (Reitan, 1958) were administered; in part B, participants were presented with a page containing numbers from 1 to 13, and letters from A to L. The child was asked to draw lines between increasing numbers alternating with letters (i.e. 1-A-2-B-3-C... etcetera), and their score was the time to completion. Errors, which rarely occurred without spontaneous self-correction, were pointed out by the examiner, and were to be corrected before the child could continue. Thus errors (self-corrected or not) served to lengthen the time to completion. Trails B is reported to measure a variety of skills related to attention, and low scores are thought to reflect weak executive functions "related to initiating, inhibiting, sequencing, and monitoring" behaviour, as well as "difficulty dealing with more than one stimulus at a time and maintaining a flexible mental orientation" (Groth-Marnat, 1997, 574). Individuals with weaknesses on Trails B might also be expected to have difficulty performing tasks that require divided attention (Groth-Marnat, 1997, 570-4). In order to control for the speed with which individuals can complete simple visual search, the Trails A score (time to connect 13 numbers in order) was subtracted from the Trails B score to derive the measure used in the present study. 3) One computerized test of attention, a version of Gordon's AX model **Continuous Performance Test** (CPT; Gordon & Mettelman, 1988) was used. This test measures response latencies to target and distractor stimuli, assessing sustained attention, impulsivity and distractibility. Visual displays of digits were presented one at a time at the rate of one digit per second, and the participant was required to respond with a button press to targets (a '9' following a '1') while inhibiting responses to distractors (such as '2-9' or '1-4'). Erroneous responses, particularly commission errors (responding to distractors) on the CPT are indicative of difficulties inhibiting a competing response and actively maintaining the task demands. Thus, the number of commission errors was used as the measure of interest in the present study.

Processing speed (PS). Two tasks thought to measure processing speed, involving the speeded search and processing of visual information, were administered. 1) On the

Visual Matching (vismatch) subtest from the Woodcock Johnson Tests of Cognitive ability, Revised (Woodcock & Johnson, 1989), the child repeatedly circled two identical numbers in a row of 5 numbers as rapidly as possible. The score was the number of correct items completed in a three minute time limit. 2) On the **Symbol Search** (symsrch) subtest from the WISC-III (Wechsler, 1991), the child was repeatedly asked to identify whether or not one of two target symbols was present among a set of five alternatives. The child's score was the number of correctly completed items in two minutes, with errors reducing the child's score.

Procedure

Tests were individually administered to children by examiners trained in test administration and skilled in establishing rapport with children. Tests were given in quiet rooms free from distractions, over two testing sessions, each 90 minutes in length and scheduled three to six weeks apart. Children were given breaks as necessary, but most children did not require a substantial break during each 90 minute testing session. The SST and CST were administered on both testing occasions, in order to facilitate the estimation of test-retest reliability. A number of additional tests were given that were not reported on in the present study. The order of administration for the first testing session was as follows: SST, the Test of Nonverbal Intelligence (TONI) Form A, CST, Children's Test of Non-word Repetition, TONI Form B, FDS, BDS, Trails A & B, CPT, Digit Naming Speed, and CVLT. The order of administration for the second testing session was as follows: SST, Finger Tapping, Semantic Categorization, Symbol Search, CST, Visual Matching, Semantic Association, Stroop, the Auditory Analysis Test (AAT), and Passage Comprehension from the Woodcock Johnson tests of Achievement, Revised (Woodcock and Johnson, 1989). In planning the order of administration, an attempt was made to minimize the likelihood that children would become fatigued during testing. More specifically, memory tests were interspersed with other kinds of tests, and longer and/or more attention-demanding tests tended to be given earlier in the testing session when children might be expected to have the most energy and motivation.

It was not expected that the order of administration would significantly affect the results of factor analyses. de Jong and Das-Smaal (1993) found that the factor structure of a set of tests similar to the ones given in the present study was invariant across two different orders of administration. Other, similar investigations have also used a fixed order of administration (de Jonge & de Jong, 1996; Kail & Hall, 2001).

Results

Data preparation and screening

Missing data. Data were complete for most variables. However, data were missing for two cases on Semantic Categorization, one case on Trails B-A, three cases on the Stroop Colour and Word Test, one case on the Continuous Performance Test, and 29 cases on the second shopping list ('list B') of the CVLT. The high rate of missing data for CVLT list B occurred because this was the last test given during the first testing session, and these 29 children did not have sufficient time to complete all the tasks. Although children often vary in the time it takes to complete testing, the 29 children with missing data were compared to the other 90 children to ensure that they were not different in any systematic way. Although the children with missing CVLT list B data tended, on average, to be about eight months younger than the children without missing data (t (117) = 2.12, $p < .05$), they did not differ with respect to gender or nonverbal intelligence. Children with missing data on the CVLT list B were also no different than those without missing data with respect to their ability to remember words from list A of the CVLT (standardized scores; $t = 1.02$, p = n.s.). Thus, it is not likely that the missing data on the CVLT list B were systematically related to the children's short-term memory ability.

Missing data were not replaced in the present sample, and cases with missing data were retained. Instead, missing data were handled using a modification of standard maximum likelihood estimation that involves estimating latent variable means and intercepts in the AMOS SEM analyses (Arbuckle, 1997). This method of handling missing data results in more stable parameter estimates than listwise or pairwise deletion (see Kline, 1998, 303), and requires less stringent assumptions to be made about the randomness of the missing data. SEM methods make the assumption that data are *missing at random* (where missing data are unrelated to participants' true scores on the variable in question), rather than the more stringent assumption (made in data replacement, and in listwise or pairwise deletion) that data are *missing completely at random* (where missing data are unrelated to participants' status on all variables under investigation) (Arbuckle, 1997, 500). In the present study, the assumption that missing data on the CVLT list B were unrelated to children's 'true' or potential performance on this task was thought to be a reasonable assumption to make, especially given the comparisons reported above.[1]

---

[1] In order to be more certain that the missing data on CVLT list B did not have an impact on the results of the present study, key analyses were re-calculated after excluding

Relations of variables with age. Because the age range of the present sample was quite broad, the correlations of each variable with age were examined. These correlations, as well as raw score means, standard deviations, minimum and maximum values for each variable, are presented in Table 1. Means and standard deviations for each variable, broken down by age, are presented in Table 2.

Insert Tables 1 and 2 about here

Chronological age was mildly related or unrelated to most variables; however, the correlation with age was moderately strong for tests of PS and for two tests of CA. Because correlations between variables in the present study could be magnified by variance shared with age, it was considered important to statistically control for variance due to age. In this way, relations between latent constructs would be less likely to be confounded by the tendency for all cognitive abilities to improve with age. Therefore, each variable was entered into a regression in which chronological age was used as the predictor. Unstandardized residuals (reflecting the amount of variance in each variable that was not related to chronological age) were saved as new variables, and were used in all SEM analyses.

A second age-related issue pertains to implicit assumptions that are made when testing factor models using children of a wide age range grouped together. By looking at the

this variable (in particular, the results depicted in Figures 5 and 6). Removal of CVLT list B from these analyses did not alter the results reported here.

relations between constructs within the entire sample, an assumption was made that the tested factor structures are invariant across age groups. Ideally, such an assumption should be formally tested by comparing the measurement and path models from age group to age group. Unfortunately, the relatively small sample size (for SEM analyses) in the present study could make such age group comparisons insufficiently stable. A less ideal, but perhaps satisfactory, alternative strategy to evaluate this assumption, is to examine other research for evidence regarding the invariance of similar factor structures across age groups in childhood. Kail and Hall (2001) found that factor analyses of simple and complex span tests consistently produced the same results (i.e. a two-factor solution) in both younger (8-10-year-olds) and older (11-13-year-olds) children. Similarly, de Jonge and de Jong (1996) found that a five-factor confirmatory factor analysis of tests of attention, memory span, reasoning, reading comprehension and reading speed yielded acceptable model fit within each of Grades four, five and six. Further, de Jonge and de Jong (1996) tested a model in which it was specified that the parameters of the model (correlations between the five factors) were equivalent across age groups (quite a stringent assumption) – this model provided an acceptable fit to their data (Chi-square $p = .01$, NNFI = .96, CFI = .96). Although the specification of equivalence of parameter estimates across groups did result in a significant decrease in model fit (Chi-square difference test was significant at $p = .03$), the authors noted that the fit of the model in which parameters were free to vary across age groups was not appreciably better with respect to fit indices (Chi-square $p = .07$, NNFI = .97. CFI = .98). Thus, de Jonge and de Jong (1996) concluded that the relations between the investigated factors were similar across Grades four, five and six. Given that the present study investigated similar constructs in children the same ages as those investigated by de Jonge and de Jong (1996) and Kail and Hall (2001), it seems reasonable to make the assumption that the factor relations are similar across the age groups studied here.

Distribution characteristics. Before proceeding with statistical analyses, the data were plotted, and distributions were examined for outliers, skew and kurtosis. Outliers were defined as values that fell three standard deviations from the sample mean. Each outlying score was adjusted to the value corresponding to 3z (or -3z). After these adjustments, skew and kurtosis were calculated for each variable. Table 3 presents the number of outliers transformed for each variable, as well as the estimates of skew and kurtosis for each variable. There were only six outlying scores that needed to be adjusted. Most variables did not exhibit significant skew or kurtosis (using the criterion of skew/kurtosis estimates greater than one), indicating that their distributions did not deviate significantly from a normal distribution. However, the distribution of scores on the Sentence Span Test was considerably more normal at the second administration. The distribution of the Computation Span Test was equally normal from the first to second administration. Therefore, scores from the second administration of the SST and CST were used in all SEM analyses. The distribution of CPT commission errors was positively skewed (1.34), and exhibited a positive kurtosis (1.65). The normality of this distribution was improved by applying a logarithmic transformation (log # errors + 1), and the new skew and kurtosis estimates are given in Table 3.

Insert Table 3 about here

Other data transformations. Scores on Stroop, Trails B-A and log CPT commissions were transformed (multiplied by -1) so that higher values would be indicative of stronger performance. In this way, interpretation of path values in the SEM would be more straightforward (that is, a positive path would always be indicative of stronger performance on one construct being related to stronger performance on the other construct).

Psychometric properties of, and performance on the SST and CST

Estimates of internal consistency (coefficient alpha), split-half reliability and test-retest reliability were calculated for both the SST and CST. To calculate internal consistency for each test, each of the 20 items were scored either correct ('1') or incorrect ('0'), and these item scores were used to estimate inter-item correlations and coefficient alpha. To calculate split-half reliability, each 20 item test was divided into two halves by alternating items. The total score for each half was calculated by adding the number of correctly completed items in each half. The two halves were then correlated with each other and corrected for the full length of the test using the Spearman-Brown formula. To calculate test-retest reliability, total test scores over the two administrations were correlated with each other. Reliability estimates are given in Table 4 and mean scores for each age group are given in Table 2.

Insert Table 4 about here

Both the CST and the SST behaved similarly across age groups. On average, nine-year-old children could manage the dual processing and storage demands for two sentences or operations, and showed a beginning ability to manage three sentences/operations. Performance on the CST and SST improved from age nine to age ten – on average, ten-year-old children could manage the dual processing and storage demands for three sentences/operations. There did not appear to be a further improvement in performance from age ten to age thirteen on either test.

Reliability of the Sentence Span Test. Estimates of alpha and split-half reliability were similar within each administration, and indicated that the SST had an acceptable degree of internal consistency. Estimates of internal consistency were slightly higher at the second administration compared to the first administration – this is likely related to the relatively non-normal distribution that was evident in first administration SST scores (as noted earlier). The distribution of first administration SST scores was positively skewed, and 50% of cases fell within a narrow range of performance – between -.4z to +.3z. At the second administration, 50% of cases fell within a broader range of performance – between -.8z to +.9z. It seemed likely that one reason for this discrepancy is that the SST was the first test administered during the first testing session. Children may have been slightly anxious or unsure about what testing would be like, or they may not have fully understood what was expected from them on this task. Although children had been given one practice item, it is recommended that subsequent research use at least three practice items with complex span tests in order to ensure that children fully understand what is expected of them.

The test-retest reliability of the SST, although significant, was found to be low (r = .47), further suggesting that performance on the first administration of the SST was likely influenced by factors other than working memory ability (e.g., anxiety and/or misunderstanding). The correlation between administration times was corrected for attenuation due to lack of internal consistency (alpha), and although this correction yielded a noteworthy improvement in the estimate of test-retest reliability ($r' = .57$), that estimate is still low and inadequate.

Reliability of the Computation Span Test. Estimates of alpha and split-half reliability were similar within and across administrations, and indicated that the CST has an acceptable degree of internal consistency. The test-retest reliability of the CST, although significant, was found to be inadequate ($r = .60$). The correlation between administration times was corrected for attenuation due to lack of internal consistency (alpha), and although this correction yielded a noteworthy improvement in the estimate of test-retest reliability ($r' = .73$), that estimate is still low and inadequate.

Structural equation modeling (SEM) of the relations between cognitive abilities

These results will be discussed in two sections: first, the relation between STM and WM, and then, the prediction of individual differences in WM ability. To address both these questions, the AMOS software program was used to conduct SEM analyses (Arbuckle, 1997). Using SEM, investigators can specify measurement models and structural models, and then test whether the observed pattern of correlations between variables is consistent with the specified models. Measurement models are used to evaluate whether correlations between variables are consistent with theories about the psychological constructs thought to underlie those variables. For example, in the present study it was hypothesized that indicators (measured variables) of WM would intercorrelate more strongly with each other than with indicators of STM. Structural models are used to evaluate whether correlations between variables are consistent with theories about the predictive relations between psychological constructs. Such questions are analogous to questions about predictions addressed using regression analyses, and the path coefficients obtained in SEM are interpreted in the same way as regression weights. For example, following Kail and Hall (2001), it was expected that PS might predict a greater amount of variance in WM than it would in STM.

There are a number of important notations used in SEM that should be defined. Indicators (measured variables) are represented in measurement and structural models by rectangles; latent variables (constructs that are tapped by indicators) are represented by ellipses. Curved paths between latent variables represent correlations/covariances, whereas straight arrows between latent variables represent predictive paths. It is important to remember that when a path or correlation is *not* specified between latent variables, then that parameter is hypothesized to be equal to zero. Thus, a straight arrow from CA to WM reflects the prediction that CA contributes/predicts individual differences in WM, but the converse is not true (that is, WM is hypothesized to have *no* unique effect on CA). Similarly, if there is no path between two latent variables, then they are hypothesized to be unrelated.

In the present study, maximum likelihood estimation was used to estimate the fit of the obtained covariance matrix to various measurement and structural models. When a particular model is said to 'fit well', this means (depending on the fit index) that either: a) the covariances implied by the fixed and free parameters specified in the model are consistent with the observed covariances used to estimate the free parameters in the model, or, b) that the covariances implied by the fixed and free parameters specified in the model are more consistent with the observed covariances than are the parameters associated with a "null" model that specifies no covariance among variables (Hoyle & Panter, 1995, 165). The assessment of how good the fit is for a particular model is not entirely straightforward. As a result, numerous fit indices are provided in AMOS output. The recommendations presented by Hoyle and Panter (1995) were followed in choosing the three following fit indices. 1) Chi-square (X^2); which is a statistical test of the lack of fit resulting from the path specifications. The chi-square is associated with a p (exact) value, which, when non-significant ($p > .05$), indicates that the lack of fit resulting from the specified paths in the tested model is not significant. 2) The Tucker-Lewis index (TLI); which compares the lack of fit of the tested model to that of the null model. Values greater than .90 are indicative of significantly better fit in the tested model compared to the null model. 3) The comparative fit index (CFI); which is interpreted in the same way as the TLI (Hoyle & Panter, 1995, 166-7). Two additional fit indices were also chosen because they have been used in previous research (Engle et al., 1999) relevant to the current investigation. 1) The root-mean-square error of approximation (RMSEA); for which good fit is indicated by values less than .05. 2) p (close) for the RMSEA; which tests for an acceptably close model fit, indicated by a non-significant result ($p > .05$). The goodness of fit index (GFI) and adjusted goodness of fit index (AGFI), commonly reported in SEM analyses, could not be calculated in the present study because data were missing on a number of variables as described earlier.

Table 5 presents the correlation matrix and standard deviations for all variables (residuals controlling for age).

Insert Table 5 about here

Are WM and STM distinct constructs? The first prediction was that WM, as measured by the SST, CST and BDS, would be distinct from, but correlated with STM, as measured by FDS, CVLT-A and CVLT-B. Support for this prediction would be indicated by a two-factor model that provided a better fit to the data than a one-factor model (a one-factor model hypothesizes that all memory tests measure the same construct). The results of the two-factor solution are shown in Figure 1. The two-factor model did not provide a good fit to the data, but neither did the one-factor model (for the one-factor model, Chi-square [9 *df*] = 17.95, $p = .036$; RMSEA = .092, *p* close = .123; TLI = .753; CFI = .894, and for the two-factor model, Chi-square [8 *df*] = 17.30 $p = .027$; RMSEA = .099, *p* close = .096; TLI = .711; CFI = .890). Unacceptable values were obtained for all fit indices, and the estimate of the correlation between WM and STM was not admissible (>1). The lack of fit evident in both two- and one-factor models suggested that inferences about the relation between WM and STM could not be made based on either model.

Insert Figure 1 about here

Next, an attempt was made to improve the fit of the measurement model by including a correlated error between CVLT-A and CVLT-B. It was reasoned that because these tasks came from the same test, and involved similar demands (recalling as many shopping list items as possible), they might share method variance that could be uncorrelated with Forward Digit Span (FDS). By accounting for the method variance shared by CVLT-A and CVLT-B, FDS might correlate more strongly with the remaining variance in those verbal memory tasks. Therefore, a correlated error (between e5 and e6 in Figure 1) was added, and the model re-tested. Although both the one-factor and the two-factor solutions yielded acceptable fit indices, the two-factor solution was inadmissible, because the correlation between factors was found to be greater than one (1.23). The one-factor solution provided a good fit to the data, as is shown in Figure 2 (for the one-factor model, Chi-square [8 *df*] = 7.90, $p = .443$; RMSEA = .000, p close = .643; TLI = 1.003; CFI = 1.000, and for the two-factor model, Chi-square [7 *df*] = 6.68 $p = .463$; RMSEA = .000, p close = .648; TLI = 1.011; CFI = 1.000).

Insert Figure 2 about here

One interpretation of these results is that the six key memory tests investigated in the present study underlie one single memory ability. Such an interpretation would be consistent with the results of de Jonge and de Jong (1996), who found that simple and complex span tests were highly intercorrelated and loaded on the same factor in a sample of elementary school-age children.

However, a number of features of the analysis presented in Figure 2, and of the correlation matrix presented in Table 5, suggest that the interpretation of a one-factor solution may be premature. First, not all tests appear to be equally good measures of the single construct evident in Figure 2. Indeed, the factor loadings of the CVLT-A and CVLT-B are relatively low, suggesting that these two tests may measure a memory ability that is not part of the single construct evident in Figure 2. Further, the correlated error between CVLT-A and CVLT-B is moderately strong, and may possibly reflect systematic variance that is not shared with the other memory tests, besides just method variance. Second, inspection of the correlation matrix (Table 5) indicated that one variable in particular was not behaving as anticipated. Notably, FDS, thought to be a test of STM, correlated more highly with all hypothesized tests of WM than with other STM variables (i.e., CVLT list A and B). However, if FDS were deleted, then a two-factor structure would seem consistent with the observed correlations between the key memory tests in the matrix. That is, when FDS is excluded from the matrix, intercorrelations within tests of WM and STM, are, for the most part, larger than the intercorrelations between tests of WM and STM. Thus, it seemed likely that the pattern of results described above was a result of the unexpected pattern of correlations between FDS and the other memory measures.

It is difficult to understand why, in the present sample, FDS did not appear to be a good measure of STM. Although one possible interpretation of this result is that FDS is in fact a test of WM, there did not seem to be sufficient evidence in the literature to be consistent with such an interpretation. In a recent review and investigation, Ramsay and Reynolds (1995) explored whether Forward Digit Span and Backward Digit Span measured the same ability, or whether they ought to be considered separately. Based on their review, they concluded that Forward Digit Span is a test of verbal sequential STM, and is associated with left-hemisphere brain activation. Backward Digit Span was also reported to involve verbal sequential memory, and correlated strongly with FDS. However, the added transformation requirement of BDS leads it to also correlate with tests of spatial memory. Further, BDS tends to be associated with nonverbal, fluid reasoning to a greater extent than FDS, and deficits in BDS performance have also been associated with right-hemisphere brain damage. Based on their review and on a set of factor analyses of the subtests of the Test of Memory and Learning (TOMAL), the authors speculated that BDS assesses an individual's ability to shift back and forth between verbal and visual processing (Ramsay & Reynolds, 1995). Thus, the evidence to date suggests that BDS and FDS tap separate cognitive abilities.

Although a clear explanation for the relation between FDS and the WM tests could not be easily understood from the present data, the obtained results were interpreted to simply mean that FDS is not a good measure of STM in the present sample. Therefore, a third factor model was tested, in which FDS was excluded.

The correlation between the two factors in the solution in which FDS was deleted was strong at .58, but the two-factor solution provided a good fit to the data, as shown in Figure 3 (Chi-square [4 *df*] = 5.35, $p = .253$; RMSEA = .054, *p* close = .392; TLI = .895; CFI = .972). In contrast, the alternative one-factor model did not provide a good fit to the data (Chi-square [5 *df*] = 11.78, $p = .038$; RMSEA = .107, *p* close = .102; TLI = .580; CFI = .860), and resulted in a significant decline in fit relative to the two-factor model, as indicated by the chi-square test of the difference between models[2] ($X^2(1) = 6.43$, $p = .01$).

---

[2] The Chi-square difference test is used to compare the fit of hierarchical models (i.e., where one model is a subset of the other model), and evaluates the significance of the decrease in fit as

---

Insert Figure 3 about here

These latter results suggest that the memory data, with the notable exception of FDS, are consistent with a model specifying two distinct constructs – the first of which involves concurrent storage and processing (or transformation) of words and digits (best designated as WM), and the second of which involves the ability to store as many words as possible from lists of fifteen 'shopping list' words (presumably STM).

Other memory measures. Having established a well-fitting two factor solution (once FDS was excluded from the model), the correlations of the two Swanson Cognitive Processing Tests (SC and SA) with the WM and STM latent factors were explored. As discussed earlier, no predictions were made about which factor these two tests would load on, and therefore, a measurement model was tested in which SC and SA were allowed to load freely on both STM and WM factors. The results of this model are presented in Figure 4.

Insert Figure 4 about here

This model provided a good fit to the data, as indicated by all fit indices (Chi-square [11 *df*]

---

paths are eliminated (e.g., a covariance that is set to '1'). The difference between the Chi-squares of the two models is tested using the difference of the degrees of freedom of the two models. A significant result indicates that the path specification led to a significant decrease in fit, and suggests that the model has been simplified too much (Kline, 1998, 133).

= 11.05, p = .439; RMSEA = .006, p close = .672; TLI = .998; CFI = .999). Note that neither SA nor SC were found to be good measures of WM, because both paths were small and non-significant (.23 and .08, respectively). However, SC exhibited a moderate and just-significant loading (.40; critical ratio = 1.97, p = .05) on the STM factor. In contrast, the loading of SA on STM was only .11 and non-significant. Finding SC loading on the STM factor is understandable because on this test, children are not required to manage additional task demands that could interfere with storage of the words to be recalled. On both SC and SA, words must be recalled in order by category – thus imposing a semantic framework within which words are to be recalled. Although some mental manipulation is required in SA to determine which words belong together, it is possible that organizing words by category might reduce the processing demands by facilitating recall. Further, on SC, the words are initially presented to children in order by category; therefore, children can choose to recall words exactly as they are heard, thereby limiting any controlled processing associated with having to re-order the words.

One final measurement model, representing the addition of SC to the model given in Figure 3, was then tested. The results of this model are presented in Figure 5.

Insert Figure 5 about here

The model depicted in Figure 5 offers the same interpretation as the model given in Figure 3. The correlation between the two factors was strong at .62, but the two-factor solution provided the best fit to the data (Chi-square [8 *df*] = 9.11, p = .333; RMSEA = .034, p close

= .539; TLI = .953; CFI = .982). The alternative one-factor model did not provide a good fit to the data (Chi-square [9 *df*] = 14.96, p = .092; RMSEA = .075, p close = .238; TLI = .779; CFI = .905), and was associated with a significant decline in fit relative to the two-factor model, as indicated by the chi-square test of the difference between models (X^2[1] = 5.85, p = .02). Note that the path values of the two CVLT tests are strong, and quite similar from Figure 3 to Figure 5, providing further support for the interpretation that their low factor loadings in Figures 1 and 2 indicated they measured something distinct from the other memory tests. Further, subsequent analyses showed that when a correlated error was added between CVLT-A and CVLT-B to the model given in Figure 5, the value of that error was .08, and non-significant. This suggests that the moderately strong correlated error observed in Figure 2 consisted not only of method variance, but also of other systematic variance that presumably reflects STM.

The role of controlled attention in the prediction of working memory. Recall that according to Engle and his colleagues (1999), individual differences in WM are thought to be a function of one's ability to control attention in the face of interference or distraction, rather than a function of storage capacity. From this view, the ability to control one's attention was predicted to be a strong predictor of WM capacity, whereas the ability to store information for short periods (i.e., STM) was not expected to contribute to the prediction of WM once CA was taken into account. This prediction was tested by modeling direct paths *from* latent STM and CA variables *to* a latent WM variable. Direct paths from STM and CA reflect the unique relation of each construct with WM, when the other construct is held constant. According to Engle et al.'s (1999) model, to the extent that individual differences in WM capacity are a function of CA, the path from CA to WM should be strong, and the path from STM to WM (when CA is accounted for) should be weak. The results of the CFA used to test these hypotheses, presented in Figure 6, indicate that this model provides a good fit to the data.

Insert Figure 6 about here

The results are consistent with Engle's theory in a number of ways – first, all fit indices met criteria for acceptable fit (Chi-square [24 *df*] = 25.42, $p = .383$; RMSEA = .022, p close = .725; TLI = .974; CFI = .986). Second, the path from CA to WM, representing the unique variance in CA (controlling for STM), is strong (.67) and significant (critical ratio = 2.14, $p < .03$), indicating that CA is an important predictor of individual differences in WM capacity. In contrast, when CA is held constant, there is *no* relation between STM and WM. Clearly, in this investigation of children's memory, short-term storage capacity does not make a unique contribution to the prediction of WM once CA is taken into account.

Perhaps surprising, however, was the finding that CA and STM were highly correlated (.69; critical ratio = 4.81, $p < .01$), indicating that in a sample of children, tests thought to measure STM are nonetheless associated with individual differences in CA. Based on Engle et al.,'s (1999) model, one might have expected to see a lower correlation between STM and CA, because information held in STM is thought to be maintained using rehearsal and coding strategies that are viewed as less attention-demanding (particularly for adults). Further, according to Engle et al.'s model, CA should distinguish WM from STM (especially when WM and STM tests share common task demands [e.g., the use of words, digits] and therefore rehearsal/coding strategies). However, it is important to note that a strong Pearson *r* association between STM and CA is consistent with the notion that children's performance on STM tests may require CA to some extent. As Engle et al., (1999) have proposed, the rehearsal and coding strategies used to maintain information in STM may be more attention-demanding for children than for adults. Such an interpretation suggests that STM and WM tests may fall on a continuum of increasing CA demands, and the CA requirement could vary with the population being investigated.

Of the results presented in Figure 6, it is also important to note that the path coefficient for the CPT is quite low (.36), suggesting that this test may have weak reliability and/or may have unique systematic variance that is not shared with the other tests of CA. In the present study, test-retest reliability for the CPT could not be calculated, and estimates of the CPT's internal consistency were not considered appropriate given the relative infrequency of errors, and given the timed format of the test. Nonetheless, the reliability of the CPT was estimated in two ways. First, a regression was conducted in which CPT scores were predicted with all variables used in the present study. The value of *R* squared, which can be thought of as an estimate of roughly all the systematic variance in the CPT, and therefore a reliability estimate, was .28. Second, estimates of the test-retest reliability of CPT commission errors were gathered from the literature, and were found to vary considerably. In a sample of normal seven to eleven year old boys, Halperin, Sharma, Greenblatt and Schwartz (1991) found that the test-retest reliability of false alarms in an A-X CPT task, over a time period of about five months, was .50. A 95 % confidence interval around this estimate would be approximately ± .17. Gordon and Mettelman (1988) reported a test-retest reliability (over 2-22 days) of .84 for a sample of mixed hyperactive and learning disabled children. A 95 % confidence interval around this estimate would be approximately ± .18. Thus, although there is one good reliability estimate from the literature (i.e., Gordon & Mettelman, 1988), the estimates from the present study (i.e., the path coefficient in the AMOS analysis and the *R* squared from the regression), as well as another estimate from the literature (i.e., Halperin et al., 1991), suggest that the reliability of CPT commission errors is likely inadequate. It may also be the case that in the present study, the CPT consists of considerable systematic variance that is not shared with any other variables examined here. In light of the evidence of the weak reliability of CPT commission scores, it is suggested that investigators using this variable in the future obtain an estimate of its reliability, or alternatively, consider employing other tests of CA instead.

<u>The role of processing speed in the prediction of working memory</u>. Given the work of Kail and Hall (2001) and Salthouse and Meinz (1995), it was expected that processing speed (PS) might also be a predictor of WM. It has been hypothesized that when individuals can process information more rapidly, they can, as a result, better control their attention and complete more mental operations on information before it decays from memory. This would suggest that PS may affect WM through CA. However, previous factor analyses of PS and CA tests have given varied results (de Jong & Das-Smaal, 1993; Salthouse & Meinz, 1995). Thus, before attempting to model the relation between PS and WM, the hypothesis that the tests used to measure PS and CA in the present study were indeed consistent with a two-factor solution was tested. It was predicted that PS, (as measured by the perceptual speed

tests of Visual Matching and Symbol Search), would be distinct from, but still correlated with CA, (as measured by the Stroop, Trails B minus A, and CPT commission errors). Support for this prediction would be indicated by a two-factor model that provided a better fit to the data than a one-factor model, hypothesizing that all attention and processing speed tests measure the same construct. The results of the two-factor solution are shown in Figure 7.

Insert Figure 7 about here

Both the one-factor and the two-factor model provided a good fit to the data, as indicated by the acceptability of all fit indices (for the one-factor solution, Chi-square [5 *df*] = 1.54, p = .909; RMSEA = .000, p close = .949; TLI = 1.088; CFI = 1.000. For the two-factor solution, Chi-square [4 *df*] = 1.41, p = .842; RMSEA = .000, p close = .900; TLI = 1.083; CFI = 1.000). However, the correlation between PS and CA was very high (.96), and, a test of the fit difference between the two models indicated that there was no significant decrease in fit when the correlation between the two factors was set to one ($X^2(1) = .12$, $p = .73$). Thus, on the basis of parsimony, the one-factor model is preferred over the two-factor model. These results suggest that in this sample of children, in which CA and PS have been measured using the present set of tests, these two constructs are not distinct, but rather, reflect a similar cognitive ability. Because CA and PS (as they were operationalized in this study) were not found to be distinct, questions about how they interact to predict WM capacity are senseless.

Discussion

The first issue examined in the present study concerned the distinctiveness of WM and STM in children. The results reported here suggested that both factors are necessary to account for children's performance on various verbal memory tests. However, there were some surprises regarding which tests would best serve as indicators of these factors. Specifically, Forward Digit Span (FDS), generally understood to be a test of STM, correlated more strongly with tests of WM than with tests of STM. This result was surprising because FDS was not expected to require concurrent processing and storage of information; participants are simply asked to recall strings of digits in the same order in which they were heard. Although it is possible that FDS shared domain-specific variance with other digit-related tasks (i.e., BDS, CST, visual matching), this similarity does not account for its relatively strong correlation with the SST (a WM test). Future investigations of the distinctiveness of WM and STM in children may benefit from using greater numbers of memory tests, in which digit and word stimuli are more equally represented among hypothesized WM and STM factors. In this way, variance that is not stimulus-related (i.e., digits versus words) could be more easily disentangled. Future research on children's memory could also benefit from attempts to better understand the relations between FDS and other memory tests (including tests thought to measure both WM and STM).

The present finding that STM and WM can be measured distinctly in children is inconsistent with the findings of de Jonge and de Jong (1996), who found that only one factor is necessary to account for children's performance on a set of verbal memory tests. The results reported here are consistent with the findings of Kail and Hall (2001), who found that two factors best describe children's performance on various verbal memory tests. However, the current results differ from those of Kail and Hall (2001), who found that FDS loaded more strongly on a STM factor than on a WM factor. In the present study, FDS was not found to be a good measure of STM because it was highly correlated with WM tests. Thus, the present conclusion that STM and WM can be measured distinctly in children should be accepted with caution, because this conclusion was supported only when FDS was excluded from the WM and STM measurement model, and not when FDS was retained in the measurement model.

As was mentioned above, it is very important to employ tests that use a variety of stimuli (e.g., digits and words), in order to be confident that a set of tests are clustering together for reasons other than shared method variance. One drawback of the present study was that STM was defined solely by the ability to recall semantically related words. Thus, the conclusions of the present study rest on the assumption that these tests provide an adequate measurement of STM. This seems to be a reasonable assumption to make, because the task demands of the CVLT and SC conform to what is commonly thought to constitute a STM task. More specifically, both tests involved the storage of information for brief periods of time, without placing extra processing demands on children to think about or to deliberately re-order that information. However, the information that was stored was quite specific in nature – all the words were semantically related. The conclusions made in the present study would have been stronger if STM had been measured using more varied tasks, including short-term storage of semantically unrelated words, nonwords, letters, and digits. It would also have been helpful if more varied task formats had been used. Measurement of

STM should ideally include not only a variety of stimuli, but also a variety of task formats.

The second issue examined in the present study concerned the prediction of individual differences in WM capacity using tests of STM, CA and PS. Although WM and STM were strongly correlated (r = .62 in Figure 5), when individual differences in the capacity to control attention were considered, STM contributed virtually no further variance to the prediction of WM (path coefficient = .07 in Figure 6). Consequently, the results were consistent with the model of WM proposed by Engle and his colleagues (1999), which states that individual differences in WM capacity are a function of the ability to sustain and control attention in the face of distraction or interference. However, the results also found that CA and STM were strongly related to each other in this sample of children, and individual differences in CA were not specific to WM.

One explanation for this result arises from the possibility that tests of STM may be more attention-demanding in children than in adults. Such a possibility was suggested by Engle et al., (1999). Future research could test this possibility by investigating the relations between STM, WM and CA in samples of young children, older children, adolescents, and adults. If tests of STM are more attention-demanding in children than in adults, then the correlation between STM and CA would be expected to be lower in adults than in children. In contrast, the correlation between WM and CA might be expected to remain strong in all age groups.

Such an explanation would imply that there are no 'pure' measures of WM or STM. Rather, various memory tests may be best conceptualized as falling along a continuum of increasing CA demands. The need for participants to use controlled attention functions in memory tasks could arise for a number of reasons – including the management of dual-task demands (as is seen in complex span tests), demands for altering the order of information prior to recall (as is seen in BDS), and strategy-related demands (rehearsal or grouping of information). Strategy-related demands may be more or less attention-demanding dependent on several factors, for example, how spontaneously (or automatically) a child is able to employ a task strategy, how much mental effort is required to implement a strategy, and/or how familiar or meaningful are the to-be-remembered stimulus items. If such a 'continuum' view of memory is pursued, then it might make less sense to talk about separate 'WM' and 'STM' abilities. Instead, the concepts of 'memory' and 'controlled attention' might satisfactorily account for children's performance on a range of memory tests.

Processing speed has also been hypothesized to be an important predictor of WM, but in the present study, PS was not independent of tests thought to measure CA. Tests of both CA and PS also tended to relate highly to age, further suggesting that they measure a single cognitive ability that develops together over the age range studied here. In speculating about the reason for the strong relation obtained between CA and PS, it is notable that these tests are all speeded tests. Children were given a time limit within which to work on Visual Matching, Symbol Search, and on the Stroop Colour and Word Test. On the Trail-making Test, children were asked to complete the task as quickly as possible, and on the Continuous Performance Test, stimuli were briefly presented on the screen one after the other, requiring children to make rapid decisions about whether a pair of stimuli constituted a target. Even though an attempt was made to control for the speed with which children could complete simple visual search (i.e., Trails A) and simple naming speed (i.e., naming colour patches), it was clear that the tests of CA nonetheless required children to rapidly process information. Similarly, tests of PS, although chosen because of their relatively 'simple' task demands that were not thought to impose specific demands on children to inhibit irrelevant or distracting information, may have required children to sustain and control their attention in deliberate ways. Even though the necessity of inhibiting a prepotent response was not specifically 'built-in' to PS tests, the speeded nature of these tests may have resulted in children having to control their attention. For example, in tests where attention must be rapidly shifted from one stimulus item to the next, it becomes difficult for participants to identify whether a particular stimulus is the target or not. If a non-target stimulus shares some features with a target stimulus (e.g., similar shape or similar orientation of parts within the shape), the non-target stimulus may capture the child's attention, requiring him/her to inhibit responding to irrelevant stimulus features. Thus, the tests of PS and CA used in the present study appear to share a number of task demands, and can be thought of as assessing a single cognitive ability, namely, the ability to control one's attention selectively to task-relevant features, especially under conditions of speeded visual search or processing.

The finding that PS and CA shared a considerable amount of common variance is consistent with the results of Salthouse and Meinz (1995), who found that most of the age-related variance in the Stroop task was shared with age-related variance in paper-and-pencil and in reaction time-based processing speed tasks. Salthouse and Meinz (1995) interpreted their results to mean that the Stroop measure of inhibition is in fact another test of processing speed. However, an alternative interpretation, introduced above, is that the measures of processing speed are alternative tests for inhibition. From the present data, these possibilities cannot be teased apart.

To better explore the relation between CA and PS, it may be helpful to attempt to measure CA in ways that do not require speeded processing or response. Measuring controlled attention in non-speeded situations may allow researchers to better investigate the relative contributions of speeded processing, versus controlled attention processes (such as inhibition), in the prediction of working memory capacity. It hardly seems possible to measure processing *speed* in conditions where speed is not necessary, but it might be possible to measure controlled attention under conditions in which speed of performance is not important. However, such a measurement may be difficult to accomplish, since requirements for controlled attention do not seem as though they would be as challenging under conditions where participants have all the time that they need to cope with a set of task demands. Another potentially helpful approach was suggested in a recent study by Salthouse, Fristoe, McGuthry & Hambrick (1998). They attempted to measure the ability to switch attention (arguably a function of controlled attention) independently of PS. These authors looked at the relations among task switching, processing speed (reaction time), short-term memory, fluid reasoning (gF) and age in young and older adults, wondering whether speed or task switching would account for more variance in the prediction of gF and STM. Task switching was operationalized by evaluating the cost, in terms of the detriment to performance (error rates and loss of speed), that resulted when participants were asked to switch their mode of responding on a speeded task. For example, on one task, participants were presented with repeated displays of two digits that the participant was asked to add. When cued to switch, participants were required to subtract subsequent digits. PS was operationalized as the baseline reaction time to complete addition or subtraction; task switching was operationalized as the increase in reaction time or errors (compared to the average of three preceding trials) resulting from the cue to switch task operation. Their results indicated that PS accounted for considerable variance in the prediction of gF and memory. However, task switching did not account for additional variance in gF and memory, once speed was taken into consideration. These results suggest that speed is the more fundamental construct, mediating the relationship between task switching and higher-order cognitive abilities. However, one could argue that such an operationalization of task switching does not fully capture all aspects of selective or controlled attention. Further research needs to be conducted with children, using methods that capture other aspects of attention, to more fully understand the relations between speed, attention, and memory.

Future research investigations of cognitive processes in children may continue to benefit from SEM methods, however, the results of the present study also highlight challenges that can complicate such investigations. When using SEM, it is of the utmost importance to carefully consider the tests used to measure the various cognitive abilities. As was seen in the present study, the results of measurement models can vary substantially depending upon the tests used as indicators of each construct. When tests do not behave as expected, it can become difficult to sort through whether the model is wrong, or whether the model has not been tested appropriately. Accordingly, the results of SEMs are best interpreted in the context of a solid theoretical understanding of the constructs of interest.

SEM will likely continue to enjoy popularity in the investigation of children's cognitive abilities – one significant advantage of SEM is the ability to account for random measurement error, and the ability to assess constructs independently of the specific methods used in any one given test. Given that assessment and measurement are important contributions that psychologists can make toward the goal of understanding children's cognitive functioning, SEM methods have the potential to facilitate this contribution.

References

Arbuckle, J. L. (1997). Amos Users' Guide, Version 3.6. Chicago, IL: SmallWaters Corp.

Atkinson, R. C. and Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence, K. W. and Spence, J. T. (Eds.) The Psychology of Learning and Motivation, Volume 2. New York: Academic Press, Inc., 89-195.

Baddeley, A. (1990). Human Memory: Theory and Practice. London: Lawrence Erlbaum Associates Ltd., Publishers.

Baddeley, A. (1996). Exploring the central executive. The Quarterly Journal of Experimental Psychology, 49A(1), 5-28.

Baddeley, A. D. and Logie, R. H. (1999). Working memory: The multiple-component model. In Miyake, A. and Shah, P. (Eds.) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge: Cambridge University Press, 28-61.

Conway, Andrew R. A., and Engle, Randall, W. (1994). Working memory and retrieval: A resource-dependent inhibition model. Journal of Experimental Psychology: General, 123(4), 354-373.

Cowan, N. (1999). An embedded-processes model of working memory. In Miyake, A. and Shah, P. (Eds.) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge: Cambridge University Press, 62-101.

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

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin and Review, 3(4), 422-433.

Delis, D. C., Kramer, J. H., Kaplan, E., Ober, B. A., and Fridlund, A. J. (1994). The California Verbal Learning Test, Children's Version. San Antonio, TX: The Psychological Corp.

Eastwood, A. E., and Steffy, R. A., (submitted for publication). The role of EEG in the diagnosis of Learning Disorder and Attention-Deficit/Hyperactivity Disorder.

Engle, R. W., Carullo, J. J. & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. Journal of Educational Research, 84(5), 253-62.

Engle, R. W., Kane, M. J., and Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In Miyake, A. and Shah, P. (Eds.) Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge: Cambridge University Press, 102-134.

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

Golden, C. J. (1978). The Stroop Color and Word Test. IL: Stoelting Co.

Gordon, M., & Mettelman, B. B. (1988). The assessment of attention: I. Standardization and reliability of a behavior-based measure. Journal of Clinical Psychology, 44, 682-690.

Groth-Marnat, Gary. (1997). Handbook of Psychological Assessment, Third Edition. New York: John Wiley and Sons, Inc.

Halperin, J. M., Sharma, V., Greenblatt, E., and Schwartz, S. T. (1991). Assessment of the Continuous Performance Test: Reliability and validity in a nonreferred sample. Psychological Assessment, 3(4), 603-608.

Hoyle, R. H. and Panter, A. T. (1995). Writing about structural equation models. In Hoyle, R. H. (Ed.), Structural Equation Modeling: Concepts, Issues and Applications. CA: Sage Publications, Inc.

de Jong, P. F. (1998). Working memory deficits of reading disabled children. Journal of Experimental Child Psychology, 70, 75-96.

de Jong, P. F. and Das-Smaal, E. A. (1993). Factor structure of standard attention tests for children: A distinction between perceptual speed and working memory. European Journal of Psychological Assessment, 9(2), 94-106.

de Jong, P. F. and van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. Journal of Educational Psychology, 91(3), 450-76.

de Jonge, P. and de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. Personality and Individual Differences, 21(6), 1007-20.

Kail, R., and Hall, L. K. (2001). Distinguishing short-term memory from working memory. Memory and Cognition, 29(1), 1-9.

Kline, R. B. (1998). Principles and Practice of Structural Equation Modeling. New York: The Guilford Press.

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

Neath, I. (1998). Human Memory: An Introduction to Research, Data, and Theory. CA: Brooks/Cole Publishing Company.

Ramsay, M. C. and Reynolds, C. R. (1995). Separate digits tests: A brief history, a literature review, and a reexamination of the factor structure of the Test of Memory and Learning (TOMAL). Neuropsychology Review, 5(3), 151-171.

Reitan, R. M. (1958). The validity of the trail making test as an indicator of organic brain damage. Perceptual and Motor Skills, 8, 271-6.

Rosen, Virginia M., and Engle, Randall W. (1998). Working memory capacity and suppression. Journal of Memory & Language, 39(3), 418-436.

Salthouse, T. A. (2000). Aging and measures of processing speed. Biological Psychology, 54, 35-54.

Salthouse, T. A., Fristoe, N., McGuthry, K. E., and Hambrick, D. Z. (1998). Relation of task switching to speed, age, and fluid intelligence. Psychology and Aging, 13(3), 445-61.

Salthouse, T. A., and Meinz, E. J. (1995). Aging, inhibition, working memory, and speed. Journal of Gerontology: PSYCHOLOGICAL SCIENCES, 50B(6), P297-P306.

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. Journal of Experimental Psychology: Learning, Memory and Cognition, 15(3), 507-16.

Swanson, H. Lee. (1993). Individual differences in working memory: a model testing and subgroup analysis of learning-disabled and skilled readers. Intelligence, 17, 285-232.

Swanson, H. Lee. (1994). Short-term memory and working memory: do both contribute to our understanding of academic achievement in children and adults with learning disabilities? Journal of Learning Disabilities, 27(1), 34-50.

Swanson, H. L. (1996). The Swanson Cognitive Processing Test. Austin, TX: Pro-Ed Inc.

Towse, J. N., Hitch, G. J., and Hutton, U. (1998). A reevaluation of working memory capacity in children. Journal of Memory and Language, 39, 195-217.

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

Wechsler, D. (1991). The Wechsler Intelligence Scale for Children, Third Edition. San Antonio, TX: The Psychological Corporation.

Woodcock, R.W., & Johnson, M. B. (1989). Woodcock-Johnson Psycho-Educational Battery - Revised. Allen, TX: DLM Teaching Resources.

Woody, Erik Z. (1998). Psychology 800 – Psychometric Theory and Methods: Course notes. Department of Psychology, University of Waterloo, 69.

Table 1. Raw score means, standard deviations, minimum values, maximum values, counts, and correlations with age.

Variable	Mean	s.d.	min	max	N	r_{age}
Digits Backward score (max 14)	5.74	1.69	2	10	119	.11
Sentence Span Test (max 4)	2.17	1.05	.25	4	119	.22*
Computation Span Test (max 4)	2.28	.84	.50	4	119	.23*
Digits Forward score (max 16)	9.48	2.02	4	14	119	.17
CVLT total list A score (max 75)	53.78	6.59	30	69	119	.23*
CVLT list B score (max 15)	7	2.06	3	14	90	.19
Semantic Categorization (max 8)	2.26	.96	0	4	117	.00
Semantic Association (max 8)	2.39	1.55	0	6	119	.16
Stroop Colour Word Test*	-27	7.71	-48	-5	116	-.12
Trails B-A (time in seconds to completion)	47.52	19.7	16	93	118	-.39**
CPT (number of commission errors)	3.47	3.94	0	19	118	-.27**
Symbol Search (# completed in 2 min.)	30.5	5.24	20	44	119	.45**
Visual Matching (# completed in 3 min.)	43.7	6.62	26	60	119	.59**

* Stroop score is # colours named on interference trial (max 45) - # colours named on colour patch trial (max 45). Scores closer to zero reflect decreasing interference relative to simple colour naming speed.

Table 2. Raw score means (standard deviations) for each age. One 7- and one 8-year-old are included with the 9-year-old age group.

Variable	9	10	11	12	13
Digits Backward score (max 14)	5 (2)	6 (2)	6 (2)	6 (2)	6 (2)
Sentence Span Test (max 4)	1.6 (.8)	2.2 (1.1)	2.6 (.9)	2.1 (1.0)	2.4 (1.2)
Computation Span Test (max 4)	1.7 (.6)	2.5 (.9)	2.4 (.8)	2.4 (.9)	2.4 (.7)
Digits Forward score (max 16)	9 (2)	10 (2)	9 (2)	10 (2)	10 (2)
CVLT total list A score (max 75)	50 (6)	53 (6)	57 (7)	53 (7)	56 (5)
CVLT list B score (max 15)	6 (2)	7 (2)	7 (2)	8 (2)	7 (2)
Semantic Categorization (max 8)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
Semantic Association (max 9)	2 (1)	2 (2)	3 (2)	2 (1)	3 (1)
Stroop Colour Word Test*	-27 (7)	-26 (7)	-26 (8)	-28 (9)	-29 (7)
Trails B-A (time in seconds to completion)	65 (19)	48 (21)	41 (17)	42 (12)	41 (17)
CPT (number of commission errors)	5.7 (4.6)	3.6 (4.2)	1.9 (2.1)	3.0 (2.7)	2.8 (3.5)
Symbol Search (# completed in 2 min.)	26 (3)	30 (5)	32 (5)	32 (4)	33 (6)
Visual Matching (# completed in 3 min.)	37 (6)	42 (6)	45 (4)	47 (6)	49 (5)

* Stroop score is # colours named on interference trial (in 45 sec) - # colours named on colour patch trial (in 45 sec). Scores closer to zero reflect decreasing interference relative to simple colour naming.

Table 3. # of adjusted outliers, subsequently calculated skew and kurtosis for each variable.

Variable	# outliers	skew	kurtosis
Digits Backward score	0	.07	-.51
Sentence Span Test*	0	.01	-.83
Computation Span Test*	0	.26	-.37
Digits Forward score	0	.05	-.23
CVLT total list A score	1 (-3z)	-.15	.12
CVLT list B score	1 (3z)	.50	.06
Semantic Categorization	0	-.72	-.31
Semantic Association	0	.12	-.97
Stroop Colour and Word Test	0	.28	.75
Trails B-A	1 (3z)	.54	.03
Log CPT commission errors**	2 (3z)	.01	-.93
Symbol Search	1 (-3z)	.25	.15
Visual Matching	0	-.40	-.05

* The second administration of these tests was used in all analyses.

** It was necessary to apply a logarithmic transformation to CPT commissions errors (log # errors + 1), in order to obtain an approximately normal distribution.

Table 4. Reliability estimates for the SST and CST.

	Alpha	Split-half	Test-retest	*Test-Retest′
Sentence Span Test				
1st administration	.79	.82	.47**	.57
2nd administration	.85	.88		
Computation Span Test				
1st administration	.82	.83	.60**	.73
2nd administration	.83	.82		

*Estimate of test-retest stability corrected for unreliability as estimated by coefficient alpha, using the following formula: $r'_{xy} = r_{xy}$ / [square root of $(r_{xx} \times r_{yy})$].

** correlation is significant at $p < .001$

Table 5. Correlation matrix of all variables used in the present study. Variables are unstandardized residuals after controlling for chronological age. Significant correlations ($p < .05$) are in boldface type.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Digits Backward													
2. Sentence Span Test	.157												
3. Computation Span Test	**.381**	**.279**											
4. Digits Forward	**.329**	**.380**	**.446**										
5. CVLT total list A	**.197**	**.280**	**.239**	**.240**									
6. CVLT list B	.084	**.280**	.168	.148	**.411**								
7. Semantic Categorization	.064	**.295**	.161	.124	**.317**	**.279**							
8. Semantic Association	.180	**.211**	.110	**.252**	.179	.201	.128						
9. Stroop Colour Word Test	.084	.061	**.270**	.089	**.305**	.209	**.188**	.061					
10. Trails B-A	.101	**.235**	**.312**	.177	**.292**	**.245**	.167	**.254**	**.280**				
11. log CPT commission	.133	**.204**	**.267**	.073	.129	-.140	.094	-.015	**.212**	**.185**			
12. Symbol Search	.147	**.264**	**.247**	.167	**.265**	.112	.079	**.223**	**.364**	**.384**	**.280**		
13. Visual Matching	**.326**	**.278**	**.434**	**.333**	**.444**	.195	.112	**.197**	**.379**	**.502**	**.294**	**.639**	
Standard deviation	1.684	1.028	.817	1.986	6.256	1.997	.957	1.533	7.662	18.155	.349	4.677	5.338

Figure 1. Confirmatory factor analysis of working memory (WM) and short-term memory (STM).

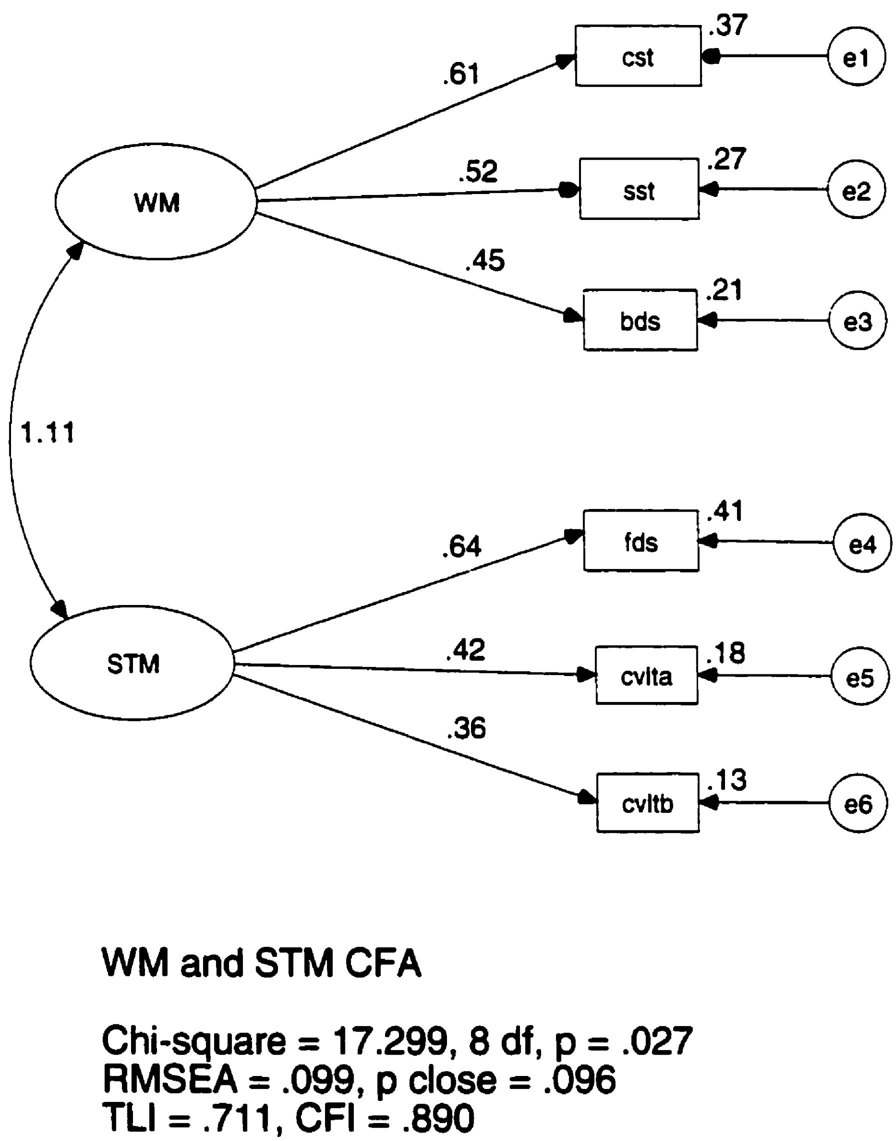
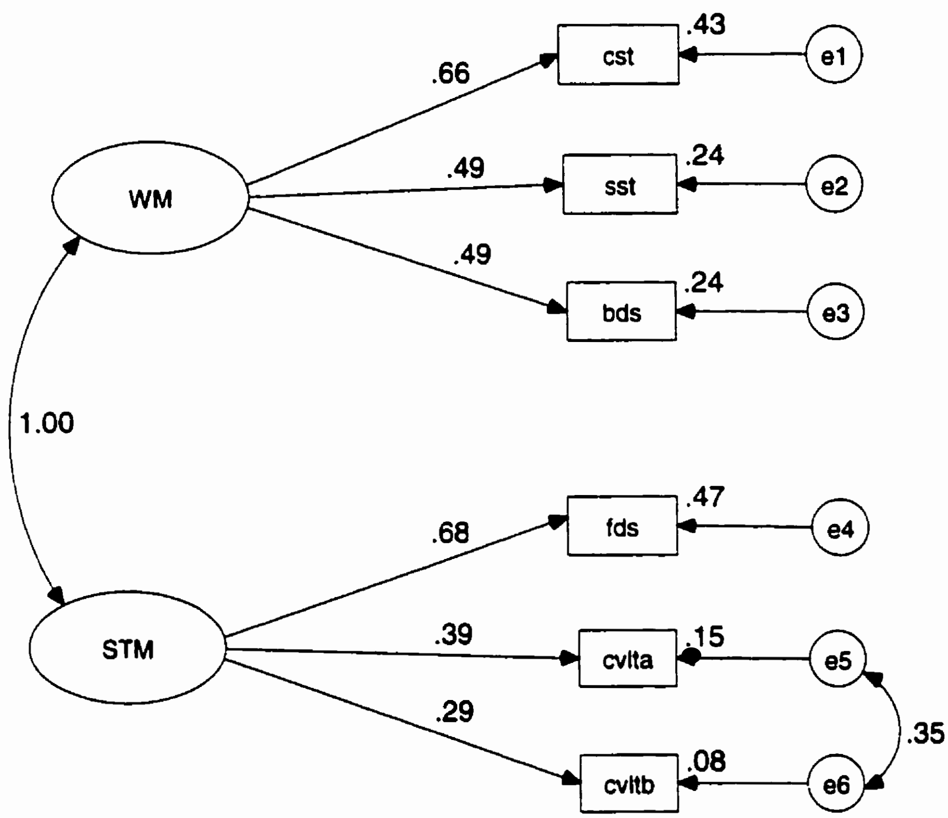


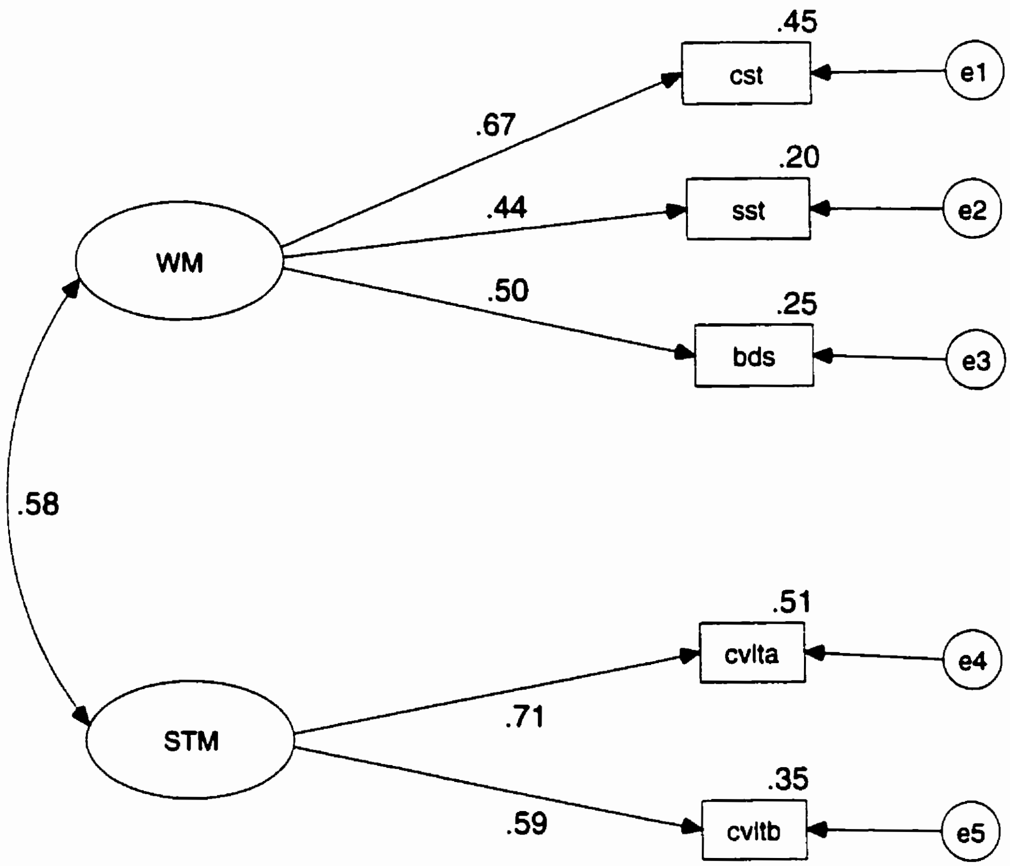
Figure 2. Confirmatory factor analysis of WM and STM, with a correlated error added between the two subtests of the California Verbal Learning Test.



WM and STM CFA

Chi-square = 7.904, 8 df, p = .443
RMSEA = .000, p close = .643
TLI = 1.003, CFI = 1.000

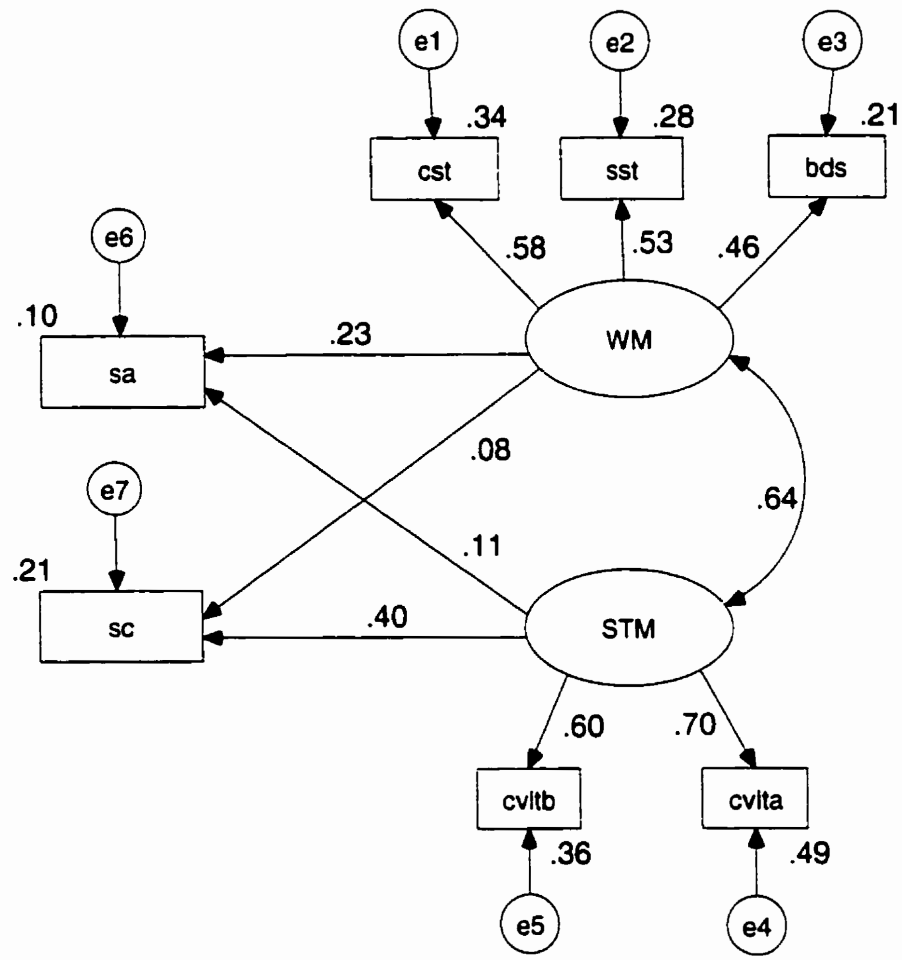
Figure 3. Confirmatory factor analysis of WM and STM, excluding Forward Digit Span (FDS) from the STM factor.



WM and STM CFA

Chi-square = 5.353, 4 df, p = .253
RMSEA = .054, p close = .392
TLI = .895, CFI = .972

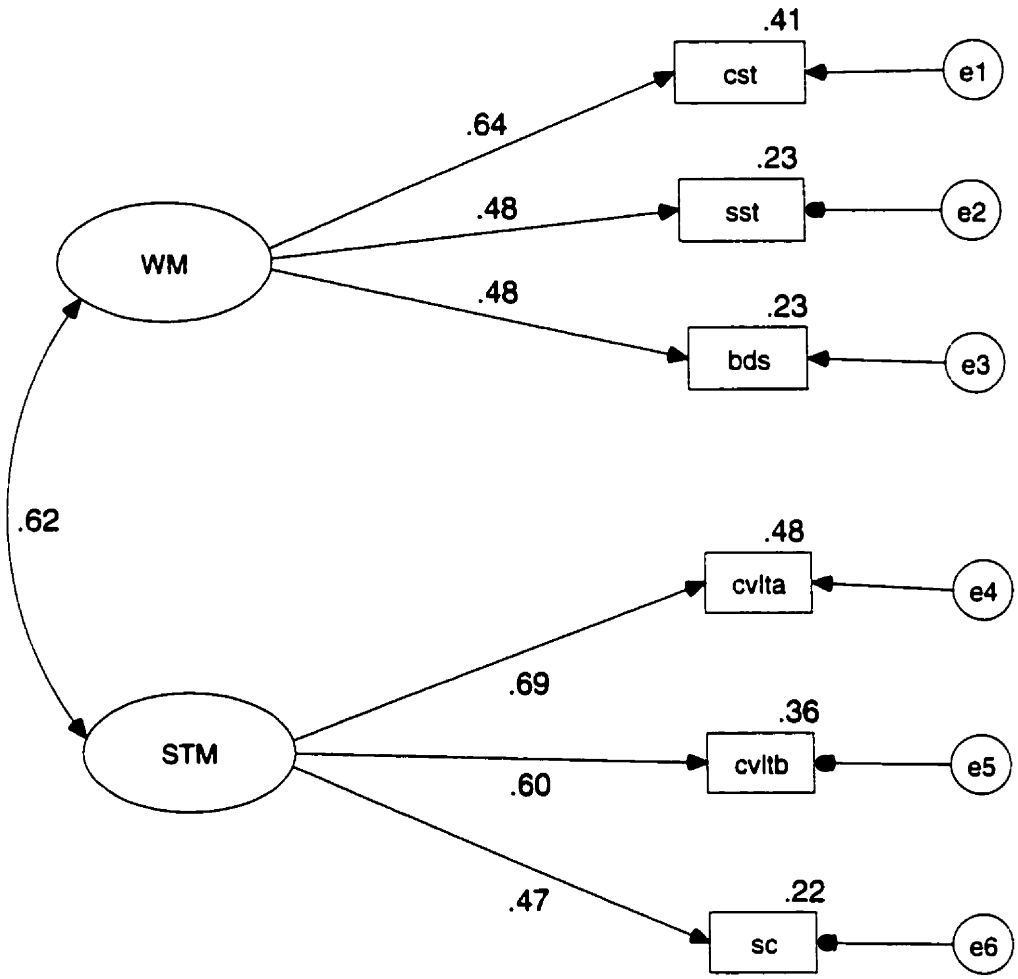
Figure 4. Factor analysis exploring the relations of Semantic Categorization (SC) and Semantic Association (SA) with the WM and STM factors.



Relations of SA and SC to WM and STM

Chi-square = 11.052, 11 df, p = .439
RMSEA = .006, p close = .672
TLI = .998, CFI = .999

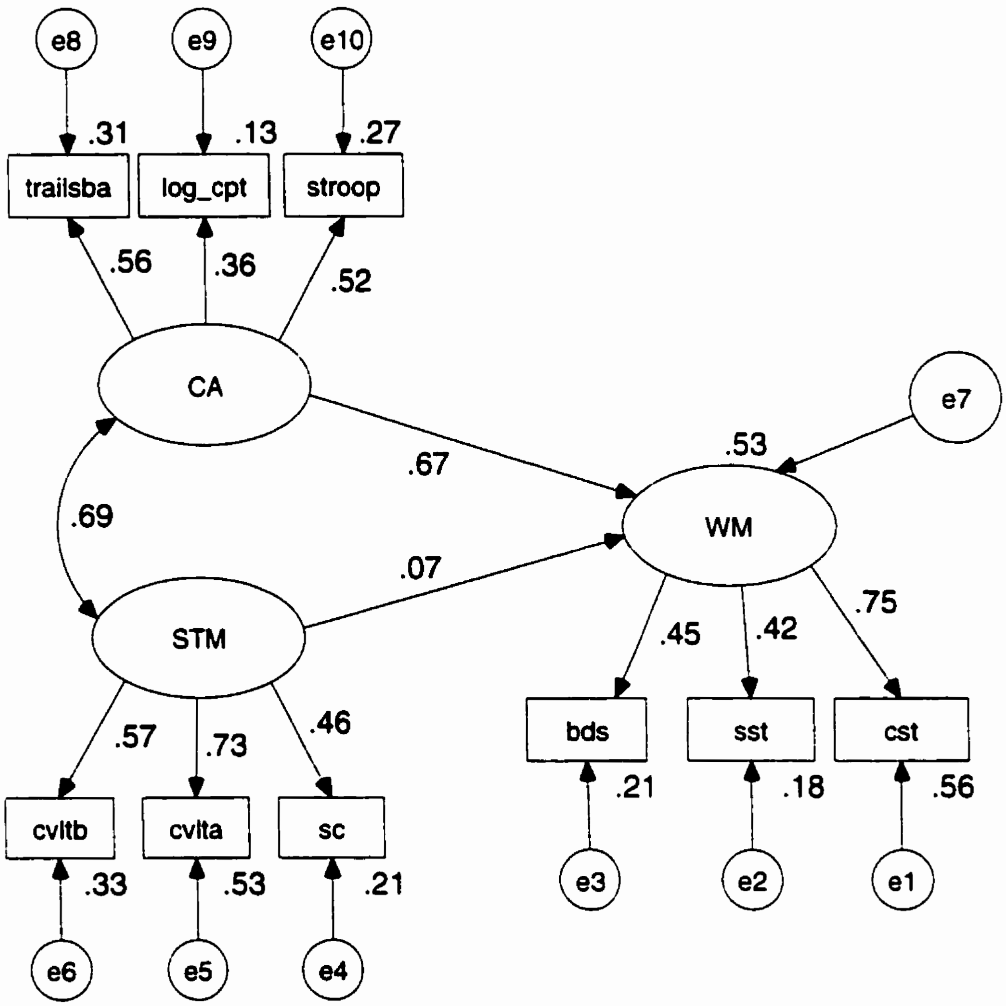
Figure 5. Confirmatory factor analysis of WM and STM, with Semantic Categorization (SC) loading on the STM factor.



WM and STM CFA

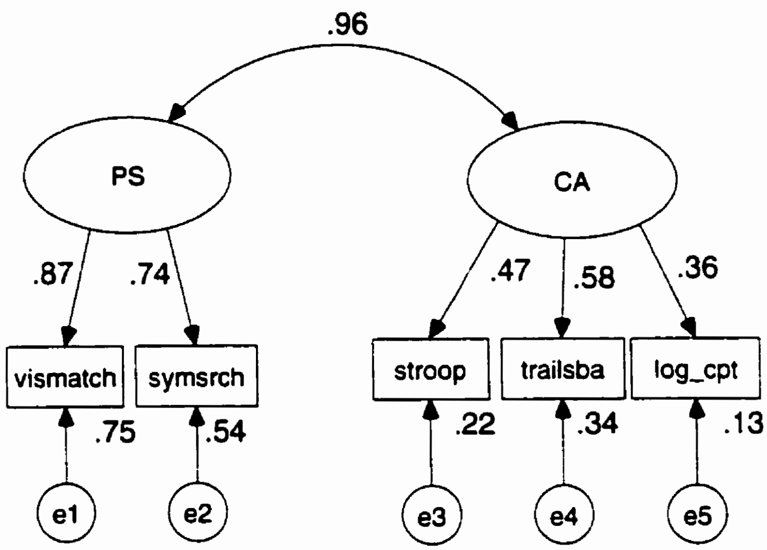
Chi-square = 9.113, 8 df, p = .333
RMSEA = .034, p close = .539
TLI = .953, CFI = .982

Figure 6. The role of controlled attention (CA) and STM in the prediction of WM.



Chi-square = 25.423, 24 df, p = .383
RMSEA = .022, p close = .725
TLI = .974, CFI = .986

Figure 7. Confirmatory factor analysis of controlled attention (CA) and processing speed (PS).



Chi-square = 1.412, 4 df, p = .842
RMSEA = .000, p close = .900
TLI = 1.083, CFI = 1.000

Appendix A

The Sentence Span Test

Instructions:

For this next activity, I'm going to ask you to try and remember words. You are going to listen to some sentences, and your job is to try and remember the last word of every sentence in order. Then I will ask you a question about one of the sentences. Let's try one for practice.

Jill loves to sail her boat.

Fred lives in a large city.

What were the last words to those sentences?

Okay, now answer this question. Who loves to sail?

Good! (if child answers correctly; if failed, praise child's effort and attempt practice trial again)

Are you ready to do the real ones? Okay

You'll hear more and more sentences as we go along. Be sure to pay close attention because I can't rewind the tape and you will only hear the sentences once. When I stop the tape, it will be time for you to remember the last words to those sentences in order. Also don't forget to pay attention to the sentences, because I'll ask you a question about one of them to see how closely you were listening.

Ready?

Two word Span:
Trial 1:
Women are now able to join the army.
Jack's job is to clean the floor.
Comp: Who's job is cleaning?

Trial 2:
Kate enjoys working with young children.
Jill wanted to put the picture in a nice frame.
Comp: What kind of children were worked with?

Trial 3:
While walking home, John was caught in the rain.
Tom and Sue met at the annual fall square dance.
Comp: Where was John going?

Trial 4:
The singer threw his sweaty shirt into the crowd.
Every morning, mom gets up and reads the paper.
Comp: What was thrown?

Trial 5:
The rain came through the leaky roof.
After the game, John felt pain in his shoulder.
Comp: When did the pain start?

Three word Span:
Trial 1:
She quietly crept up to the window.
The silly kitten got stuck up in the tree.
The angry tribe was ready to do battle.
Comp: How was the tribe feeling?

Trial 2:
In the summer I love to go to the beach.
On Tuesdays my dad sells produce at the market.
The teacher had the kids sit in a circle.
Comp: When was produce sold?

Trial 3:
The cat's purr sounded like a motor.
Sarah waited and waited for the important letter.
Her favorite flower was the lily of the valley.
Comp: What was the name of the girl who was waiting?

Trial 4:
When growing vegetables, you need good soil.
Rick was careful when crossing the ancient bridge.
Fred quickly noticed the pretty girl.
Comp: What was being grown?

Trial 5:
The volunteers tried hard to put out the fire.
The beautiful ring was made of gold.
She loved to write for the newspaper column.
Comp: Who tried hard?

Four word Span:
Trial 1:
All kinds of animals live on a farm.
They danced under a harvest moon.
She welcomed the cool northern wind.
John wondered what was in the wrapped box.
Comp: What kinds of animals?

Trial 2:
By eating right you can look after your heart.
Sue preferred to spend time with her dog.
The children did not get along with their cousin.
Long ago cave men invented the wheel.
Comp: Who did not get along?

Trial 3:
Sue asked her son to go to the store.
She forgot where her family kept the spare key.
After the shipwreck, they got to a deserted island.
My grandfather helped build the Canadian National Railroad.
Comp: Who helped build?

Trial 4:
The world is full of interesting people.
They ended the meeting on a positive note.
The explorer lived to investigate uncharted land.
George thought that he was a curious fellow.
Comp: Who investigated?

Trial 5:
Peter wanted to see the west coast.
Pepperoni pizza is her favorite food.

The sheep dogs were skilled at herding cattle.
After dessert, they asked for the bill.
Comp: What kind of pizza?

Five word span:
Trial 1:
On Sundays they usually had pancakes for breakfast.
Every morning Sue spends hours on her hair.
Miss Andrews was Becky's favorite teacher.
They paddled quietly down the beautiful river.
On rainy days, I like to curl up with a novel.
Comp: What day did they have pancakes?

Trial 2:
Mary was tired after a day at the office.
The strongest wall was built of rock.
For the prom, Jane made her own dress.
When the bombs dropped, they ran for cover.
Finally, she no longer owed anyone a cent.
Comp: What was dropped?

Trial 3:
He slipped and fell and broke his arm.
In the evenings Karen liked to listen to music.
Nothing smells sweeter than a rose.
After the long journey they found a hotel.
John hated having meatloaf for dinner.
Comp: When did Karen listen?

Trial 4:
Who knows what fairies live in the forest.
The cats chased the mouse in a corner.
The newlywed couple asked someone to take their picture.
Because he was noble, they made him king.
Kelly dove gracefully into the pool.
Comp: What kind of couple were they?

Trial 5:
The salesman believed in his product.
He could not look her in the eye.
Pam was excited when he asked her on a date.
John asked Karen to give him a hand.
Millions of bugs live under the ground.
Comp: How many bugs?

Appendix B

Computation Span Test

Instructions:

Now I am going to ask you to solve some math problems. When you solve each problem, I want you to say the answers out loud. I will read a series of math problems and you have tell me the answers for each one out loud and remember them in your head. When I finish giving you the questions, you will have to remember all the answers in the order you solved them in. Let's try some for practice.

1st Practice:	2nd Practice:
2 + 3 =	8 - 2 =
6 - 5 =	5 + 3 =
okay	4 + 2 =
	okay

Are you ready for the real ones?

Be sure to pay close attention, because I can only read them once. I'll read each problem to you and you solve each of them and say the answers out loud as we go. When I say "okay" at the end, this means that it's time for you to remember all the answers in the order you solved them in.

***HANDS ON THE TABLE (to prevent using fingers to keep track of answers)

Computation Span Test

Trial one:	Trial two:	Trial three:	Trial four:	Trial five:
1 + 3 =	5 + 4 =	4 + 5 =	5 - 2 =	4 + 3 =
6 - 2 =	3 - 2 =	9 - 7 =	3 - 1 =	4 + 6 =
(4,4)_____	(9,1)_____	(9,2)_____	(3,2)_____	(7,10)_____

Trial one:	Trial two:	Trial three:	Trial four:	Trial five:
5 - 2 =	8 - 6 =	5 + 3 =	6 + 3 =	8 - 3 =
3 + 5 =	7 - 4 =	2 + 5 =	8 - 7 =	5 - 4 =
8 - 5 =	6 + 2 =	3 + 2 =	8 - 6 =	2 + 5 =
(3,8,3)	(2,3,8)	(8,7,5)	(9,1,2)	(5,1,7)
_______	_______	_______	_______	_______

Trial one:	Trial two:	Trial three:	Trial four:	Trial five:
5 - 4 =	5 + 3 =	8 + 1 =	8 - 6 =	7 - 4 =
3 + 5 =	3 + 6 =	7 - 5 =	2 + 5 =	4 + 6 =
6 + 4 =	2 + 3 =	1 + 3 =	4 + 1 =	5 + 2 =
4 - 3 =	6 - 3 =	8 - 5 =	4 + 6 =	2 + 3 =
(1,8,10,1)	(8,9,5,3)	(9,2,4,3)	(2,7,5,10)	(3,10,7,5)
_______	_______	_______	_______	_______

Trial one:	Trial two:	Trial three:	Trial four:	Trial five:
4 + 2 =	7 - 6 =	4 + 3 =	8 - 5 =	6 + 3 =
5 + 4 =	8 - 5 =	3 + 2 =	7 - 4 =	5 + 2 =
2 + 5 =	8 + 1 =	3 + 3 =	2 + 2 =	1 + 1 =
9 - 8 =	8 - 7 =	7 - 5 =	6 - 4 =	8 - 3 =
7 - 4 =	5 - 3 =	8 - 4 =	9 - 6 =	7 - 5 =
(6,9,7,1,3)	(1,3,9,1,2)	(7,5,6,2,4)	(3,3,4,2,3)	(9,7,2,5,2)
_______	_______	_______	_______	_______

Appendix C

Example of the type of items used in the Digit Span Test (from the WISC-III, Wechsler, 1991).

Digits Forward:

Digits are presented orally at the rate of one per second. After digits are presented, the child is asked to recall them orally in the same order in which they had been given. Digit strings of increasing length are presented, using two trials at each digit length, and the test is discontinued once the child provides an incorrect response for both trials of any digit length.

Example: 5 - 9 - 4

Digits Backward:

This subtest uses the same format as Digits Forward, however, the child is asked to recall the digits in the reverse order in which they had been given.

Appendix D

Example of the type of items used in the California Verbal Learning Test, Children's Version (Delis, Kramer, Kaplan, Ober, & Fridlund, 1994).

The child is presented with a list of fifteen items and is asked to recall as many items as possible. In part A of the test, the child is presented with the same list five times, and is asked to recall as many words as possible on each trial. In part B of the test, the child is presented with a new list of fifteen words (similar to list A) and is asked to recall as many words as possible.

Example: apples, shirt, game, coat, grapefruit, lego, orange, pants, markers, pears, dolls, cap, raspberries, tie, ball

Appendix E

Example of the type of items used in the Semantic Categorization subtest of the Swanson Cognitive Processing Test, Swanson, 1996).

Words are presented orally – the list contains both category names and exemplars of the categories. The child is asked to recall the words orally by naming a category name first, followed by all exemplars of that category, before going on to the next category. Categories can be recalled in any order, but all exemplars must be recalled with each category to be correct. The eight trials consist of the following numbers of words and categories:

Trial	# categories	total number of words presented in the trial
1	1	3
2	2	6
3	2	8
4	3	12
5	4	16
6	5	20
7	7	28
8	8	32

Example: cats - persian - siamese - himalayan - meats - ham - turkey - beef
Which word was presented: chicken or himalayan?

The child's score is the number of the last trial in which all words were remembered correctly.

Appendix F

Example of the type of items used in the Semantic Association subtest of the Swanson Cognitive Processing Test, Swanson, 1996).

Words are presented orally – the words can be grouped into two or more categories, but the category names are not given. The child is asked to recall the words in groups according to category, but categories can be recalled in any order. The eight trials consist of the following numbers of words and categories:

Trial	# categories	total number of words presented in the trial
1	2	4
2	2	6
3	2	6
4	3	6
5	3	9
6	3	12
7	3	14
8	4	16

Example: baseball - robin - football - bluejay - basketball - cardinal
Which word, "bluejay", or "canary", is in the list of words I presented to you?

The child's score is the number of the last trial in which all words were remembered correctly.