8-9-2010

The Combined and Differential Roles of Working Memory Mechanisms in Academic Achievement

Bronwyn Murray
Indiana University of Pennsylvania

THE COMBINED AND DIFFERENTIAL ROLES OF WORKING MEMORY MECHANISMS IN ACADEMIC ACHIEVEMENT

A Dissertation

Submitted to the School of Graduate Studies and Research

in Partial Fulfillment of the

Requirement for the Degree

Doctor of Psychology

Bronwyn Murray

Indiana University of Pennsylvania

August 2010

Indiana University of Pennsylvania

The School of Graduate Studies and Research

Department of Psychology

We hereby approve the dissertation of

Bronwyn Rae Murray

Candidate for the degree of Doctor of Psychology

April 30, 2010 ______ ________________________________
Dasen Luo, Ph.D.
Professor of Psychology, Chair

April 30, 2010 ______ ________________________________
Krzysztof Kaniasty, Ph.D.
Professor of Psychology

April 30, 2010 ______ ________________________________
Donald U. Robertson, Ph.D.
Professor of Psychology

ACCEPTED

________________________________ ____________________
Timothy P. Mack, Ph.D.
Dean
The School of Graduate Studies and Research

Title: The Combined and Differential Roles of Working Memory Mechanisms in Academic Achievement

Author: Bronwyn Murray

Dissertation Chair: Dr. Dasen Luo

Dissertation Committee Members: Dr. Krzysztof Kaniasty

Dr. Donald U. Robertson

Generalized latent variable modeling was used to examine the relationship between working memory and academic achievement. The contributions of two working memory mechanisms that are involved in a wide variety of working memory tasks, namely short-term storage (STS) and generalized attention control (GAC) were examined. The contributions of two working memory mechanisms that are specific to two well-established measures of working memory (Digit Span and Letter-Number Sequencing), namely Backward Ordering (BO) and Mental Sorting (MS), were also examined. The contribution of these working memory mechanisms as a whole was additionally evaluated, and compared to the contribution of traditional measures of intelligence.

Mechanisms that are common across multiple working memory tasks (specifically, STS and GAC) were found to make an important contribution to both intelligence and achievement, while task-specific factors (BO and MS) were not. Furthermore, in this study the combined working memory factors were clearly better predictors of achievement than traditional measures of intelligence. At the same time, results of this study indicate that both of these traditional measures of intelligence make significant and unique contributions to academic achievement above and beyond those of the working memory factors.

The most unique aspect of this study was the examination of the relationship between *independent* latent working memory factors and a latent achievement factor. Unlike previous studies that did not differentiate between the role of storage and the role of attention control, this study was able to provide more precise information about the nature of working memory's contribution. Thus, it was possible to discern that the general factors of GAC and STS *both* made substantial unique contributions and that the contributions of the more specific mechanisms were much lesser. The theoretical and practical implications of these findings are discussed.

ACKNOWLEDGEMENTS

With thanks to Dr. Luo, Randal, and my parents.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

REVIEW OF RELEVANT LITERATURE

Brief Historical Overview of Working Memory

Working memory, “ a temporary storage system under attention control that underpins our capacity for complex thought” (Baddeley, 2007, p 1) has a relatively short history in psychology. William James (1890) was among the first to distinguish between different types of memory in his conception of a short-term “primary memory” that stored consciousness with very little effort, and a “secondary memory” that could store information permanently, but only with deliberate effort.

Around the same time, the first systematic investigations in what would become the field of differential psychology were occurring. Joseph Jacobs is credited with developing the forward digit span task, in which participants are asked to recall a series of numbers in the order of presentation, in order to quantify the limits of memory (Hannary, 1998). Among his findings was a relationship between age and storage capacity, as Jacobs reported that older children were able to retain longer spans of numbers than younger ones. Meanwhile, Francis Galton included a forward digit span task in his battery of tests administered to thousands on people of all ages, in an attempt to map out the mean and range of human mental abilities (Jensen, 2006).

Alfred Binet used a similar method to study memory, but his stimuli were unrelated words and sentences rather than numbers. Peterson (1925) explains Binet’s interest in “immediate memory” thus: “by means of memory tests one can indirectly study the operations and nature of higher mental processes as discrimination, attention, and intelligence” (p. 125).

Digit span and sentence span tasks were both included in Terman's 1916 translation and revision of the Binet-Simon intelligence scale, in which a distinction between forward and backward digit span items was first made. The forward span items were to be given to all children, while the backward items were intended only for those over the age of seven (Berliner, 2006). Terman noted that the backward span test was a superior indicator of intelligence than the forward test as "it is less mechanical and makes a much heavier demand on attention" (p. 208). The role of backward and forward span tasks continues until the present day, with their inclusion in all subsequent revisions of the Stanford-Binet scales as well as in all editions of the Weschler Intelligence Scale for Children, the Weschler Intelligence Scale for Adults and the Woodcock Johnson.

The role of working memory has been investigated in all three of the major traditions of the psychometric study of intelligence (Heitz, Unsworth, & Engle, 2005). The first tradition is associated with Spearman, and emphasizes the concept of *g*, a single intellectual ability underlying performance on a wide variety of tests. The second, the group factors tradition primarily associated with Thurstone, posits a number of specific intelligences or "primary mental abilities", such as reasoning and perceptual speed. The third view falls somewhere in between, and includes Cattell's well-known hierarchical model which places *g* atop two major divisions: fluid and crystallized intelligence.

Spearman (1927) initially stated that intelligence was a function of education and explicitly *not* a function of "bare retention" (p. 285); however, he later revised his thinking based on the very high correlations he found between certain memory tests and a general ability factor, *g*. Meanwhile, in the group factors tradition, Kelley (1928) found that four working memory tests correlated well with a general ability factor (up to r =

.56) but also with a separate working memory factor (up to r = .56), leading to the general conclusion that the two were related. Thurstone also included “associative memory” as one of his seven “primary mental abilities”, although he used tests of word- and picture-recall, rather than span tasks (Sternberg, 2002), while Guildford (1925) included 24 distinct memory factors in his 120-component model of intelligence.

In the third school, Cattell (1943) proposed a distinction between crystallized intelligence (*Gc*) and fluid intelligence (*Gf*), with the former associated with education and acquired experience and the latter associated with biologically-endowed skills and abilities. Horn (1968) investigated the role of memory span in these constructs, and reported no correlation between memory and *Gc* but a substantial correlation (r=.50) with *Gf*.

Around the same time, multi-component views of memory were re-surfacing amongst researchers outside the psychometric tradition. In 1949, Donald Hebb proposed a two-component view of memory, with short-term memory associated with temporary electrical activity in the brain, and long-term memory associated with lasting changes to the brain (Baddeley, 2007). Around the same time, information-processing theories were beginning to influence memory research, including George Miller’s famous “7+/- 2” work on short-term capacity. One of the most influential models of memory was the linear conception of Atkinson and Shiffrin (1968), which included a sensory store (for all incoming information), a short-term store (for incoming information that is attended to), and a long-term store (for information in the short-term store that has been rehearsed and therefore retained).

In the early 1970s, Baddley and Hitch developed a three-componential view of short-term memory, which they called "working memory." From a series of experiments in which they paired a standard span task with a concurrent reasoning, learning, or comprehension task, they found that while response time on the latter task increased with span length, the error rate was relatively unchanged (Baddeley, 2007). From this they concluded that multiple systems were in operation, and described a general mechanism called the central executive overseeing a number of separate "attention control" (p.12) processes as well as two content-specific storage systems: the phonological loop for aurally-presented information, and the visuo-spatial sketchpad for visual and spatial information. Most subsequent research into individual differences in working memory has been based on this dynamic, multi-component model.

Predictive Power of Working Memory

Working memory derives its importance in differential psychology from its status as a strong predictor of intelligence. A number of studies in recent years have indicated that working memory has an important role as part of the cognitive basis of intelligence. Research indicates that it is a powerful predictor of psychometric *g* (Ackerman, Beier, & Boyle, 2005; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Luo, Thompson, & Detterman, 2006; Schweizer & Moosbrugger, 2004), reasoning ability or fluid intelligence (Buehner, Krumm & Pick, 2005: Engle, Tuholski, Laughlin, & Conway, 1999: Fry & Hale, 1996; Kane & Engle, 2002; Kyllonen & Christal, 1990; Salthouse, Babcock, Mitchell, Palmon & Skovronek, 1990; Miyake, Friedman, Rettinger, Shah, and Hegarty, 2001; Süβ, Oberauer, Wittman, Wilhelm & Schulze, 2002), verbal ability (Cantor, Engle & Hamilton, 1991; Conway & Engle, 1996; Daneman &

Carpenter, 1980), and math ability (Kyttälä & Lehto, 2008). The magnitude of these correlations led many to conclude that working memory and intelligence are highly related (Conway, Kane, & Engle, 2003) if not near-identical constructs (Kyllonen & Christal). On the other hand, others have argued that the strength of the relationship has been overstated (Ackerman et al., 2005). A brief overview of key studies and their findings follows.

Tillman, Nyberg, and Bohlin (2008) investigated how the working memory components of storage and "executive" attention in both verbal and visuospatial domains relate to fluid intelligence. Their participants were 196 Swedish children aged between 6 and 13, and they partialled out shared variance to distinguish WM from STS tasks. Their predictors were one verbal (word span) and one visuospatial (picture span) storage tasks, and one verbal (Children's Size Ordering Task) and one visuospatial (picture span with a classification component) working memory task, while the sole dependent variable was Raven's Progressive Matrices. Their multiple regression analysis found that all four task types made significant unique contributions, with the verbal STS task explaining the largest proportion of Raven's variance.

Daneman and Carpenter (1980) were among the first differential psychologists to examine correlates to performance on working memory tasks, reporting a strong relationship between performance on a reading span task (in which participants were to recall the last word of a sentence) and on a reading comprehension test. Although their reported correlation is impressive (r=.72), their work has been criticized for the small size of the sample used, and possible inflation due to shared content and method variance

(Ackerman et al., 2005). Indeed, later studies (reviewed in Baddeley, 1986) that used span tasks other than reading span reported lesser correlations.

Daneman and Merikle (1996) conducted a meta-analysis of studies investigating the relationship between short-term memory and working memory indicators and indicators of language comprehension. Their analysis included 77 different samples of vastly divergent age ranges, although those who were classified as poor readers were excluded. Their analysis separated simple storage (i.e. span) measures from working memory measures that additionally required performance of another task, as well as "verbal" versus numerical tasks. They reported correlations with comprehension of .28 and .14 for verbal storage and numerical storage respectively; correlations with comprehension of .14 and .48 were found for verbal working memory and numerical working memory respectively. Thus, their analysis provided evidence that something in addition to simple storage was responsible for the relationship between working memory and measures of achievement.

Deary, Strand, Smith, and Fernandes (2007) conducted an unusually extensive investigation into the relationship between intelligence and educational achievement that included data from more than 70,000 English schoolchildren. In this prospective longitudinal study, intelligence test scores obtained at age 11 were compared with academic achievement scores obtained five years later. A factor analysis of the Cognitive Abilities Test found that the first factor *g* accounted for about 70% of the variance, while a residual orthogonal verbal factor could also be obtained. Subsequently, the authors investigated correlations between the *g* factor scores and the participants' scores on the General Certificate of Secondary Education (GCSE) achievement tests. They found significant, positive correlations between *g* factor scores and all subjects, including 0.67 for English and 0.77 for mathematics, as well as 0.69 for the overall GSCE score.

This was followed by structural equation modeling to investigate the relationship between latent *g* and a latent achievement factor. The authors reported there was a .81 correlation between the two factors, and that the achievement factor accounted for 72% of the variance of the GCSE indicators. This was followed by general linear modeling (ANCOVA) to determine the relative contribution of *g* and the verbal factor. They discovered that g accounted for 49% of achievement variance, while the verbal factor explained an addition 3% above and beyond that.

An important real-world criterion in the United Kingdom is whether students obtain at least five GCSE scores within the A-C range, as this determines eligibility for further education and training. Using logistic regression, the researchers found that of the students who obtained at least a mean g factor score, 58% met this criterion. Furthermore, 91% who were one or more standard deviations above the mean met the criterion, while only 16% of those who were one standard deviation or below did the same. The Receiver Operating Characteristic curve for this analysis covered an area of 0.859, indicating that the *g* factor scores were good predictors. Therefore, these analyses were consistent in showing that cognitive ability tests are strong predictors of later academic achievement.

In two studies, Luo, Thompson, and Detterman (2003) compared the strength of working memory as a predictor of academic achievement to the strength of indices of fluid and crystallized intelligence. Their first study analyzed the Woodcock-Johnson III (WJ III) normative data using the Total Achievement scores as the criterion, with Comprehension Knowledge, Fluid Reasoning and Working Memory clusters among the independent variables. In multiple regression analyses, they found that the Fluid Reasoning cluster alone provided no significant unique contribution. In a series of structural equation models, their Fluid Intelligence factor provided the weakest explanatory power, while their working memory factor explained about as much of the variance of Achievement as did their Crystallized Intelligence factor.

In their second study, Luo et al. (2003) used elementary school pupils of a range of ability levels. The participants included four working memory tasks (digit span as well as other storage and processing tasks), along with the Weschsler Intelligence Scale for Children Third Edition (WISC-III) Verbal, Performance and Full IQ indices. Their criterion this time was a battery of language, math, and reading tests. Using a similar series of multiple regression and structural equation modeling analyses, they found that all predictors made substantial contributions, with processing speed notably weaker than the others. The authors conclude that fluid intelligence may be somewhat redundant when used with other predictors (such as working memory), and of insufficient power when used on its own.

Rohde and Thompson (2005) also addressed the relative importance of traditional intelligence measures (in this case, Raven's Progressive Matrices and the Mill Vocabulary Scales) versus specialized (working memory, processing speed, and spatial ability) cognitive skills in predicting academic achievement. Using separate multiple regression analyses to look at three different criterion variables, (WRAT III, college GPA, and SAT scores), they found no significant unique contribution from working memory to any of the criterions. However, some weaknesses in the study should be noted. First, only a single measure of working memory (operation span) was included in the study; therefore, task-specific variance may be diluting the strength of the correlation. Second, as the authors acknowledge, their participants were drawn exclusively from a "gifted" college population; therefore, more modest correlations are to be expected than if a wider range of ability levels had been included. This is due not only to restriction of range, but to the established phenomenon that correlations between all kinds of ability tests are highest at the low end of the ability continuum (Detterman & Daniel, 1989; Jensen, 2003).

Mechanisms of Working Memory

Based on Baddeley's multi-componential model of working memory, research has also sought to determine *which* of the components of working memory are most important in explaining the relationship with cognitive abilities. Many studies addressed this issue by distinguishing between tasks that require storage and those that require an additional processing component in addition to storage. Short-term storage (STS) tasks are generally simple span tasks, in which items from a stimulus list are individually presented for later recall. More complex span tasks also involve the serial presentation of a list, but have an additional processing component; for example, operation span requires the participant to solve simple arithmetical problems while remembering the list (Conway et al., 2003). Such tasks that require the participant to shift attention between the stimulus list and a processing task are thus said to be WM tasks, in distinction to STS tasks that do not have any such attention-shifting requirement (Engle et al., 1999).

By selecting sets of tasks in this manner, Engle et al. (1999) tried to address the question of whether working memory is more predictive of fluid intelligence that short-term memory. The authors' basic methodology was to compare the fits of one-factor

(STS) and two-factor (STS and working memory) models, on the assumption that a superior fit for the two-factor model would reflect that attention control is the mechanism responsible for the observed relationship between working memory and fluid intelligence. The authors did indeed find that the two-factors models they tested provided a superior fit, but their results are inconclusive, as their models do not differentiate between different components of working memory. Their models include either correlated STS and working memory exogenous factors, or STS and working memory factors defined by orthogonal indicator sets. Although the authors hypothesize that "working memory capacity = STM capacity + central executive or controlled attention + the error of measurement" (p. 313), this relationship is not directly represented in any of their models because no model actually breaks down working memory into these three factors. This absence of independent factors means that the contributions of STS and attention control as two additive predictors cannot be directly determined. Similar methodology was later used to address the same question by Cantor et al. (1991) and Conway, Cowan, Bunting, Therriault, and Minkoff (2002), and consequently suffers from the same drawbacks.

Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000) used a taxonomic approach to studying working memory in their classification of 23 working memory tasks based on three stimulus dimensions (verbal, numerical, and spatial-figurative) and four"functions" (simultaneous storage and transformation, supervision, and coordination). These tests were administered to 128 participants along with a battery of 45 intellectual ability tests that were divided into reasoning, spatial, and numerical composites. They derived three factors from the working memory tests: a

Verbal/Numerical factor that included the functions of simultaneous storage and transformation and coordination, a Spatial-Figural factor that included the same functions, and a final factor that included those supervisory functions that required speed. In terms of correlations with the intellectual ability tests, they found the strongest correlation (.61) between the third working memory factor and a composite of the speed test. Relatively strong correlations were also found between the second factor and all three intelligence composites, and between the first factor and the numerical and reasoning composites. These results suggested that there was not a simple relationship between working memory and intelligence, but rather differential relationships between the various mechanisms and content areas.

A later study by the same authors (Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) used the same working memory model and a model of intelligence that encompassed three content areas (verbal, figural, number) and three "operations" (speed, memory, creativity, and reasoning). Additionally, they attempted to control for personality factors, subjective stress, and computer skills by checking for significant correlations between relevant measures for these possible mediators and all working memory and ability measures administered; none were obtained. Consistent with the previous study and Kyllonen and Christal (1990), the authors found that working memory generally was related to intelligence and particularly reasoning. As in the previous study, they reported that the storage and processing plus co-ordination factor was the strongest predictor of intelligence.

Based on the Oberauer model, Krumm, Schmidt-Atzert, Buehner, Ziegler, Michalczyk, and Arrow (2009) investigated the relationships among short-term storage, working memory (involving storage and processing of different stimuli), sustained attention (involving storage and processing of the same stimuli), executive function (including updating, inhibition, and shifting tasks), processing speed, and reasoning -- categories that clearly involve a high degree of overlap. While they were able to partial out processing speed and storage factors, other sources of shared variance were not identified using this paradigm. Thus, their conclusion that it is only storage (and not other factors associated with working memory) that predicts reasoning does not reflect the possible contribution of other generalized working memory mechanisms. Additionally, the fact that their sample includes only participants from a selective college suggests that the relationships found in this study may be weaker than if a more varied sample had been used.

Kane et al. (2004) sought to investigate the generalizability of working memory mechanisms through a design that separated verbal and visual-spatial content and included working memory tasks and their "short-term memory" analogues. Their short-term memory latent factor was defined by tasks that did not require the participant to shift attention from a stimulus list to be remembered, while the WM tasks required the participant to move between the stimulus list and a distracting processing task, such as arithmetic. Using structural equation modeling, they found that a one-factor (ie. generalized) working memory model fit their data (collected from adults of varied ability levels) better than a two-factor model with separate verbal and visualspatial factors. In contrast, they found the reverse was true for short-term memory. Additionally, they found that working memory was a much stronger predictor of "reasoning" (as defined by a range of verbal, visual-spatial, and inductive reasoning tasks) than was shot-term memory. the authors claim their study provides support for Baddeley's working memory model, with its central executive as a generalized mechanism, and two content-specific storage mechanisms (the phonological loop and the visuospatial sketchpad). In interpreting their results, the authors suggest that the short-term memory factor and the working memory factor (which are highly correlated) share variance due to attention control as well as storage. More specifically, they posit that working memory tasks primarily capture domain-general attention and domain-specific storage secondarily, while short-term memory tasks capture primarily domain-specific storage and secondarily domain-general attention. The implication that the variance shared by working memory and short-term memory tasks indicates that there is no pure (ie. domain-specific) working memory task is not substantiated by their study, which does not distinguish the content variability from the storage variability in their storage factor. Therefore, another possible explanation for the correlation between working memory and storage tasks is that all working memory tasks share short-term memory.

Within the field of differential psychology, the vast majority of recent investigations into the role of WM in intelligence have used latent variable analyses. This methodology includes structural equation modeling (SEM), in which causal paths between latent variables are modeled, and confirmatory factor analysis (CFA), which does not include causal paths. Latent variable analysis involves the administration of multiple measures for each latent construct hypothesized to a large number of participants. Latent factors can then be derived from the covariance among tasks said to tap the same construct, while at the same time task-specific variance is removed. Statistically, the aim of a latent variable model is to account for all of the correlations found among the measures included. A reproduced correlation matrix is generated based on the relations among the measures specified in the model, and this is then compared to the observed correlation matrix (Bollen, 1989). If the two are similar, the model is said to be a good fit.

Kyllonen and Christal (1990) were among the first to investigate the relationship between working memory and cognitive ability at the latent level. They conducted four different studies with military recruits, and administered working memory batteries (including digit span, and mental ordering tasks) and reasoning batteries, including subtests from the Armed Services Vocational Aptitude Battery (ASVAB), arithmetic and grammatical tasks. The authors acknowledge a slighter theoretical base on which to chose working memory tasks than for reasoning tasks, and "concede to a certain degree of arbitrariness" (p. 392) in their design of working memory tasks. Indeed, while a "grammatical reasoning" task is classified as a working memory task in one study, it is classed as a reasoning task in another. Both construct and content validity may be compromised as their latent factor may include a great deal of task-specific variance and fail to include relevant aspects of working memory. Additionally, the fact that their design presents working memory as an undifferentiated construct makes it impossible to identify the underlying mechanisms that account for the impressive correlations estimated (r=.80-.88) between working memory and reasoning. Although the authors' have also been criticized for prematurely concluding that working memory and reasoning and nearly identical constructs (Ackerman et al., 2005), this study was the first to attempt a "purified" view of the constructs, and included an unusually wide range of ability levels among their participants.

For a number of reasons, task selection is a particular problem when studying working memory. Beyond span tasks there is no much agreement on what constitutes a working memory task, it is difficult to find or create the range of tasks needed for a structural equation model, and most tasks require multiple mechanisms that are difficult to disentangle. One possible option is to construct something similar to a multi-trait, multi-method (MTMM) matrix in which a range of general and specialized mechanisms are included. However, this may be impossible to achieve in practice because-- with the possible exception of simple span tasks—very few working memory tasks involve only a single mechanism. For example, it is difficult to imagine a task that requires the mechanism of attention control and the specialized mechanism of backward ordering, but not the additional mechanism of STS. Many mechanisms are intrinsically intertwined in tasks in this manner. In addition, Kenny and Kashy (1992) have cautioned against the use of the MTMM matrix with confirmatory factor analyses. Matrices that use a similar paradigm, such as those that include multiple mechanisms and multiple executive operations, will experience the same problems with model estimation that they describe.

Almost all previous work into the relationship between working memory and reasoning or intelligence has been at the task level. There are many studies in which the relationship is evaluated at the level of observed variables (Conway et al., 2003; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004); as previously noted, questions relating to reliability and error are unavoidable in this approach. Furthermore, the information provided by these kinds of studies is limited. Most working memory tasks must be considered an amalgam of domain- or task- specific (e.g., verbal or arithmetic) processes and more general processes (e.g., storage). In fact, Oberauer, Schulze, Wilhelm, and Süβ (2005) write that there are four sources of variance in any working memory task; variance "specific to the task paradigm" (e.g., storage), variance of the working memory construct as a whole, content-related variance (e.g., spatial skills), and method variance (e.g., computer administration). While the number of sources of variance is arguable, failure to distinguish between them severely limits our understanding of working memory.

The findings with regard to the relationship between working memory and the various omnibus measures of intelligence have undoubtedly been inconsistent. In fact, when Ackerman, Beier and Boyle (2005) conducted a meta-analysis of the relevant studies, they found an average correlation of only.479, with a range of .21 to .55. This range is much lower than the very strong correlations reported in some studies, such as seminal work by Kyllonen and Christal (1990), in which they reported correlations of between .8 and .9 between working memory and reasoning based on four investigations. One general pattern is that studies of observed measures produce much lower correlations than do latent variable analyses, and this is generally explained in terms of the reliability of the observed measures (reflecting common variance) and error variance (random variation) which are corrected for in latent variable analyses (Bryant & Yarnold, 1995).

At the same time, the legitimacy of the correction made in the process of latent variable analyses may reasonably be called into question by the magnitude of the discrepancy. These corrections are based on assumptions of multivariate normality and linearity (Klem, 2000) that may be violated when dealing with studies of individual difference in cognitive ability. As previously noted, there is some evidence that correlations among cognitive tests of various kinds are appreciably stronger for low-ability participants than for participants of higher ability levels (Detterman & Daniel, 1989; Facon, 2004; Jensen, 2003), suggesting that a strict linear relationship between working memory and intelligence may not apply.

Generalized Latent Variable Modeling to Study Working Memory

The term "generalized latent variable modeling" (Skrondal & Rabe-Hesketh, 2004) encompasses a wide variety of latent variable modeling, including various combinations of discrete latent variables (indicating latent population heterogeneity) and continuous latent variables, as well as discrete (nominal or ordinally scaled) and continuous observed variables. As such, it includes finite mixture modeling, latent class modeling, and item response models. Luo, Chen, Zen & Murray (2010) pioneered the use of generalized latent variable modeling to explore the multiple cognitive underpinnings of working memory tasks, and explore the relationships between these mechanisms and traditional intelligence measures. As the current study is an extension of this work, this earlier study is described in some detail.

By analyzing the factor structure of digit span (DS) and letter-number sequencing (LNS) items from the Chinese version of the Wechsler Intelligence Scale for Children-Revised (WISC-R), they identified four independent mechanisms or traits, which they called Short-Term Storage (STS), Generalized Attention Control (GAC), Mental Sorting (MS), and Backward Ordering (BO). The authors posited that all working memory tasks draw on STS capacities to some extent, while GAC is involved in all working memory tasks where the individual is required to perform a mental operation in addition to maintaining items in memory. Hence, completion of the backward digit span items (BDS) requires both STS and GAC mechanisms, while forward digit span items (FDS) requires only STS. As these mechanisms are expected to be instrumental in a wide variety of tasks, Luo et al. described them as generalized traits. In addition, they identified the more specialized traits of backward ordering (BO) from the BDS items, and mental sorting (MS) from the LNS items. The authors proposed that almost all working memory tasks require some combination of generalized and specialized traits. In other words, the variance found in performance on working memory tasks can typically broken down into general (i.e., that which is shared with a variety of working memory tasks) and task-specific sources.

In obtaining optimally-fitting models for these tasks, Luo et al. (2010) first compared the fit of models with different numbers of continuous latent traits, then compared with models that included discrete (ordinal) latent traits. Further refinement of models was undertaken by varying the number of levels of discrete latent variables specified, and in all cases models that included only discrete latent variables were found to provide superior fit to those with continuous latent variables.

Luo et al. (2010) also addressed both the combined and the differential roles of these working memory mechanisms in traditional intelligence measures. Using the WISC-R summary scores for the Verbal and Performance indices along with academic achievement (Chinese and Mathematics) scores as indicators, he derived a general intelligence (*g*) discrete latent factor. The working memory traits were specified to be both independent (uncorrelated) and exogenous such that the unique contribution of each trait to general intelligence could be assessed.

Three methods of evaluation were used to evaluate the contribution of each working memory factor. First, the regression weights relating the working memory traits to the g trait were evaluated. Next, a series of nested models in which the contribution of one or multiple exogenous traits were constrained to zero were compared in order to determine the impact of the zero constraints in terms of model fit. Finally, correlations between the posterior trait scores for the working memory factors and the *g*-factor were assessed. The actual process involved in calculating these posterior trait scores as well as their theoretical significance is described in detail later in this chapter.

Based on the above methods of evaluation, Luo et al. (2010) found that that the role of the generalized traits of STS and GAC were clear. That is, the parameter estimates of each path to *g* were significant in all models tested, constraining these paths to zero resulted in appreciably worse fits, and that Pearson's r correlations between the factor mean scores for each of the generalized traits and *g* were significant at the 0.01 level.

However, the role of the specialized traits of BO and MS remained somewhat inconclusive. Parameter estimates of the paths from these factors to the *g*-factor indicated that the paths were significant at 0.01, and significant correlations were found between factor mean scores of the *g*-factor and the specialized working memory traits. Subsequent hierarchical multiple regression also suggested that BO but not MS contributes significantly to the variance of *g*. Additionally, while constraining the BO path to zero in a nested model resulted in a worse fit, constraining the MS path had no effect. Furthermore, constraining both paths simultaneously actually result in a better model fit index and in more reliable STS and GAC factors. The authors hypothesized that the inconsistent findings regarding the role of MS and BO may reflect their relatively low factor reliabilities. The factor reliabilities for MS and BO were 0.34 and 0.64 respectively, while those of the other factors were all in the 0.77-0.87 range.

Luo et al. (2010) concluded that the predictive power of working memory task performance is predominantly due to the involvement of STS and GAC, although he speculated that BO may potentially play an important role in certain specific areas of achievement.

Additionally, the authors found apparent discontinuities in the distribution of ability on the GAC trait. The posterior classification of cases revealed that only a single participant belonged in the third-lowest ability level; hence, there was a gap between the two lowest and the three highest levels. This finding suggests that there may be some important qualitative difference between those in the lower levels and those in the upper levels.

Like others, these authors reported higher correlations between working memory and intelligence measures when latent factor scores were examined that when observed indicators were used. Specifically, raw correlations were in the 0.4-0.6 range, while correlations between the working memory and g factor mean scores were in the 0.6-0.8 range. It is worth emphasizing that the latter are not based on model parameter estimates (factorial correlations or structural relations) whose substantive virtue is difficult to validate. Rather, they are tangible scores derived from the model-based posterior classification process, and whether they effectively represent meaningful constructs can be substantiated by evaluating their strength to discriminate certain cognitively exceptional individuals (e.g., those with mild mental retardation or those who are gifted).

General latent variable model has several characteristics that may allow for the surmounting of problems that have previously limited our understanding of working memory. These characteristics include the ability to include both discrete and continuous traits, the principle of local independence, the estimation of posterior probability, and the methods of trait identification. A discussion of the relevance of each of these characteristics follows.

Generalized latent variable models can include discrete as well as continuous traits. Discrete traits are present when a population is composed of subgroups formed on the basis of an unobservable quality or trait, such as certain genetic variations (Muthén, 2001). Each of these subgroups (or categories) has its own mean and variance, which are obscured when the population is treated as a whole (i.e., when the trait is regarded as continuous). When these subgroups are unordered clusters in the population, the relevant latent trait is said to be nominal. When the subgroups can be ordered along some continuum, the trait is said to be ordinal. Moreover, ordinal traits with a sufficient number of levels or categories can be used to approximate continuous traits (Aiken, 1999; Vermunt, 2002)

There are some important differences when working with continuous latent traits as compared to discrete latent traits. When both observed and latent variables are continuous, a structural equation modeling approach is used, and there is no need to estimate the person parameters (factor scores) in the process of estimating the model parameters. Instead, the likelihood function for the entire sample can be derived from the normality-based covariance structure, and the maximum likelihood estimation process can maximize the function through iterations without drawing on any person-parameter values. Although factor scores may be obtained after the model parameters are estimated, they are not an intrinsic part of the estimation process.

In contrast, when some of the variables (either latent or observed) are discrete, the estimation method necessitates a likelihood function for every response pattern. Therefore, estimates of class membership (for discrete factors) and factor scores (for continuous factors) are required for model parameter estimation. In the often preferred expectation-maximization (EM) estimation, these person-parameter values are substituted into the model in each cycle of the iteration and then updated in the next cycle to estimate the response-pattern likelihood until the convergence criterion is met.

There are also some important differences in the approach to response-pattern likelihood for continuous and discrete factors in the estimation process. For continuous factors, the probability density at each factor level is predetermined in that it is based on the normal distribution; however, obtaining the marginal probability (by integrating the probability densities) requires a great deal of computational power. For discrete factors, on the other hand, the probability density of a given class or category is empirically estimated, and the related marginal probability is obtained by summing over the class/category densities without need to integrate. When the latent distribution can be properly approximated by a discrete factor with a few levels (e.g., <10), the computational demand for a less sparse contingency table may be considerably reduced from that for a continuous latent factor whose approximate integration typically requires 10 or more intervals (nodes) for an adequate accuracy. Therefore, discrete factors can be considered to be preferable to continuous factors in this practical sense.

In the present study, a logistic link function relates the discrete observed indicators to the underlying latent traits. In the item-level modeling of the working memory tasks, the indicators are on an ordinal scale reflecting the number of prompts correctly recalled. As they relate to the underlying latent variable in a non-linear manner, the use of linear models such as employed in structural equation modeling will result in distorted estimations. Generalized latent variable modeling provides for a more appropriate method of regressing the ordinal indicators with latent predicators through a set of logistic link functions.

To illustrate, let Y_i^t be an ordinal response, where i is the participant, t is the item, and the potential responses may be coded as 0, 1…M^t. For example, in a working memory task in which full recall is scored as 4, M^t is 4. There are therefore five possible response probabilities for the item (i.e. 0, 1, 2, 3, 4) and a constraint on the first item category (0) must be imposed in order to be identified. The logistic link function that then relates the observed response to the underlying trait or traits is as follows:

$$Logit^t_m = In\left[\frac{P(Y_i^t = m \mid X_1 = x_1, X_2 = x_{2\ldots})}{P(Y_i^t = m-1 \mid X_1 = x_1, X_2 = x_{2\ldots})}\right] \qquad (1)$$

$$= \beta^t_{0m} + \beta^t_1 x_1 + \beta^t_2 x_2\ldots$$

In equation (1), x_1 and x_2... are scores of the latent traits X_1 and X_2, and β^t_{0m}, β^t_1, β^t_2...are regression weights relating the latent traits to the log-transformed odds ratio between two adjacent item-categories. An adjacent-category logit is used for which each item t with $M^t + 1$ categories results in M^t unconstrained equations and one special equation for the constrained 0 category. Only the model intercept (ie, β^t_{0m}) varies across item categories; the slopes (ie, β^t_1, β^t_2) do *not* vary. The slopes β^t_1, β^t_2 indicate how change in the latent trait X influences the logit between the categories m and $m-1$.

To provide a concrete example, imagine the logits for the item with $M^t + 1 = 5$ categories have the following intercepts: $\beta^t_{0m=1} = 0.011$, $\beta^t_{0m=2} = 0.223$, $\beta^t_{0m=3}$=0.0564, and $\beta^t_{0m=4}$=-0.123. The slope is β^t_1= 2.299, and there is a single latent trait. In order to identify parameters, the intercept for the category of 0 is set to zero, such that $\beta^{t}_{0m} = 0$. If the participant has the trait score $X_1 = 0.33$, the five logits have the following predicted values:

$logit_1 = (0.011 - 0) + 2.299\ (0.33) = 0.76967$

$logit_2 = (0.223 - 0.011) + 2.299\ (0.33) = 0.97067$

$logit_3 = (0.0564 - 0.223) + 2.299\ (0.33) = 0.59207$

$logit_4 = (-0.123 - 0.0564) + 2.299\ (0.33) = 0.57927$

$logit_0 = (0) + 2.299\ (0.33) = 0.75867$

Each of the unconstrained logits shows the degree of change in probability from one item-category to the next (the odds ratio). For example, $logit_1 = 0.76967$ means that for those with the trait score $X_1 = 0.33$, the probability of the response 1 is equal to the probability of the response 0 multiplied by 2.15905 (because $e^{logit\ 1} = e^{0.76967} = 2.15905$). Similarly, the value of $logit_2$ (0.97069) reflects the odds ratio between the response probabilities of the categories 2 and 1. As $e^{logit\ 2} = e^{0.97069} = 2.5398$, the probability of the response 2 is 2.5398 times the probability of the response 1.

In addition, odds ratios can be calculated for non-adjacent categories. For example, to find the probability of change from $m=2$ to $m=0$, calculate $e^{logit1} * e^{logit2}$ (or $e^{logit1\ +\ logit\ 2}$). Since $e^{logit1\ +\ logit\ 2} = e^{0.76967 + 0.97069} = 5.6994$, the probability of the response 2 equals the probability of the response 0 multiplied by 5.6994.

Furthermore, the odds ratios that relate to the constrained category (m=0) are used together in the back-transformation of the logits to item-category response probabilities for each category. Continuing the same example, to calculate the conditional probability for the response $m=2$ when $X_1 = 0.33$:

$P(Y^{t}_{i} = 2 \mid X_1 = 0.33)$

$$= \frac{e^{logit0+logit1+logit2}}{e^{logit0} + e^{logit0+logit1} + e^{logit0+logit1+logit2} + \ldots + e^{logit0+logit1+logit2+logit3+logit4}}$$

$$= \frac{e^{0.75867+ 0.76967+0.97069}}{e^{0.75867} + e^{0.75867+ 0.76967} + \ldots + e^{0.75867+ 0.76967+0.97069+0.59207+0.57927}}$$

$$= \frac{(2.1354)(5.6994)}{2.13543+(2.13543)(2.15905)+(2.13543)(5.6994)+(2.13543)(10.30274)+(2.13543)(18.38766)}$$

$$= \frac{5.6994}{1+ 2.15905+5.6994+10.30274+18.38766}$$

$$= \frac{5.6994}{37.5489}$$

$$= 0.15179$$

The adjacent-category logistic link function above gives rise to a partial-credit item response model when the latent trait X_1 is continuous. This link function can also be applied to models which include latent endogenous variables, such as when an ordinal latent variable is treated as a dependent variable for other explanatory latent variables. Furthermore, the same adjacent-category function relates both the latent and the observed exogenous measures to the logits of the endogenous variables.

The principle of local independence is another important characteristic. Responses to separate items on a test are not independent in that knowing the response to one item on a test provides some information about likely responses to other items on the test. In terms of probability, the joint conditional probability for a given set of responses is generally not equal to the product of the unconditional probabilities for each separate item-response. An assumption of generalized latent variable modeling is that the latent traits that underlie responses entirely account for this interdependence. Accordingly, the conditional probabilities of responses would become independent if the latent traits were statistically controlled. This axiom of local independence means that the joint probability of a particular response pattern is equal to the product of the conditional probabilities of each individual item-response. To illustrate, imagine the conditional probability for a person with trait level X_1=0.33 to respond 1 on item 1 is 0.5, and to respond 1 on item 2 is 0.01. The axiom of local independence means that the joint probability of this response pattern, also known as the likelihood function, is 0.005 (the product of 0.5 and 0.01), as the latent X_1 is assumed to account for any observed association between responses to the two items. Unlike structural equation modeling, Latent Gold accounts for associations between observed variables that are non-linear in nature, as well as standard correlations and covariances

One of the major advantages of generalized latent variable modeling over more established methods of latent class modeling is that it provides classification results for each individual as well as tangible individual trait scores. These scores can then be used to validate the theoretical relationships derived from the model. The classification of individuals into subgroups or trait levels occurs through the estimation of posterior probabilities. The calculation of posterior probabilities is based on Bayes' Theorem, and the following formula is used:

$$P\,(X_1=x_1 \mid Y^{t=1}{}_i = m^{t=1},\ Y^{t=2}{}_i = m^{t=2},\ \ldots)$$

$$= \frac{P\,(X_1=x_1)\ P(Y^{t=1} = m^{t=1},\ Y^{t=2}{}_i = m^{t=2}, \ldots \mid X_1=x_1)}{P(Y^{t=1} = m^{t=1},\ Y^{t=2}{}_i = m^{t=2}, \ldots)}$$

$$= \frac{P\,(X_1=x_1, Y^{t=1} = m^{t=1},\ Y^{t=2}{}_i = m^{t=2}, \ldots)}{P(Y^{t=1} = m^{t=1},\ Y^{t=2}{}_i = m^{t=2}, \ldots)} \tag{2}$$

This formula takes into account the following elements: the marginal probability of specific trait levels, the marginal probability of a specific response level, and the joint conditional probability of a specific response pattern and trait level. These elements will be discussed separately, along with how they combine in the formula, and how the overall classification process occurs.

The numerator of equation 2 includes the joint conditional probability for a given response pattern. As described in the previous section the joint probability for a person with trait level 0.33 to have the response pattern 11 is 0.005, which can be written as:

$P(Y^{t=1}_i = 1, Y^{t=2}_i=1 \mid X_1 = 0.33) = 0.005$

Further, assume that there are four levels of the latent trait X_1: $X_1 = 0$, $X_2 = 0.33$, $X_3 = 0.66$, and $X_4 = 1$. The marginal probability of each of these levels is obtained through model parameter estimation, but imagine they are as follows:

P $(X_1 = 0.0) = 0.22$

P $(X_2 = 0.33) = 0.43$

P$(X_3 = 0.66) = 0.19$

P$(X_4 = 1.0) = 0.16$

The numerator of the equation combines these two elements. To continue the same example, for the response pattern $Y^{t=1}_i = 1$, $Y^{t=2}_i=1$ and the trait level $X_1= 0.33$, the numerator is as follows:

$$P(X_1 = 0.33)^* P(Y^{t=1}_i = 1, Y^{t=2}_i=1 \mid X_1 = 0.33) = 0.43^*0.005 = 0.00215. \quad (3)$$

The denominator of the Bayes' formula is the marginal probability of the given response pattern. It is calculated by summing the joint probabilities of the response pattern in question and all possible trait levels. For example, the marginal probability $P(Y^{t=1}_i = 1, Y^{t=2}_i=1)$ is equal to the total probability of the response pattern $Y^{t=1}_i = 1$, $Y^{t=2}_i=1$ occurring at all levels of X_1 ($X_1 = 0$, $X_2 = 0.33$, $X_3 = 0.66$, and $X_4 = 1$):

$P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1)$

$= P(X_1 = 0)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 0)$

$+ P(X_1 = 0.33)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 0.33)$

$+ P(X_1 = 0.66)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 0.66)$

$+ P(X_1 =1)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 1)$

One of the elements on the right side of the equation is the joint probability for the response pattern in question and the trait level X_1= 0.33, which has been shown above to be 0.0021. Assume that the remaining elements of this side of the equation are as follows:

$P(X_1 = 0)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 0) = 0.6500$

$P(X_1 = 0.66)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 0.66) = 0.0200$

$P(X_1 =1)\, P(Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1 \mid X_1 = 1) = 0.0030$

Thus, the marginal probability of the response pattern (ie, the denominator of Equation 2) is as follows:

$$P\,(Y^{t=1}{}_i =1, Y^{t=2}{}_i = 1) = 0.6500 + 0.0012 + 0.0200 + 0.0030 = 0.6742 \qquad (4)$$

Finally, the values of equations 4 and 3 can be substituted into Bayes' formula in equation 2 in order to obtain the estimated posterior probability for participant *i*, who has the response pattern $Y^{t=1}{}_i = 1, Y^{t=2}{}_i=1$, to have the latent trait level X_1=0.33, as follows:

$P(X_1 = 0.33 \mid Y^{t=1}{}_i =1, Y^{t=2}{}_i = 1) = 0.00215/0.6742 = 0.0032.$

This value can be compared to the posterior probabilities for the other trait levels, which are as follows:

$P(X_1 = 0 \mid Y^{t=1}{}_i =1, Y^{t=2}{}_i = 1) = 0.6500/0.6742 = 0.9641$

$P(X_1 = 0.66 \mid Y^{t=1}{}_i =1, Y^{t=2}{}_i = 1) = 0.0200/0.6742 = 0.0297$

$P(X_1 = 1 \mid Y^{t=1}{}_i =1, Y^{t=2}{}_i = 1) = 0.0030/0.6742 = 0.0044$

Clearly, the highest of the posterior probabilities is 0.9641; therefore, a participant with this response pattern would be classified as having a trait level of 0. Latent Gold uses the same procedure of matching participants to trait levels based on posterior probabilities in its classification process.

Latent Gold also provides a weighted mean sum, called the factor mean score, for each participant based on the estimation of posterior probabilities. The factor mean score for participant *i* can be obtained by multiplying the posterior probability of each trait level by its value (0, 0.33, 0.66, or 1) and then summing the products. Using the information already obtained, the factor mean score is therefore:

$$0.9641*0 + 0.0032*0.33 + 0.0297*0.66 + 0.0044*1 = 0.025058$$

Current Study

The primary objective of the current study was to test the hypothesis that each of the working memory factors already defined by Luo et al. (i.e., STS, GAC, MS, and BO; 2010) contributes not only to intelligence but to academic achievement. This was the first such study to investigate the relationship between orthogonally-defined working memory components and academic achievement, and was significant in its ability to isolate the particular working memory components responsible for the relationship. Previous investigations have relied on correlated working memory factors, and have therefore been unable to address this issue. The ability to specify which components of working memory are instrumental in academic achievement has practical as well as theoretical significance, as there is evidence that improving performance on some working memory tasks can lead to improvement in school performance. Therefore, the results of this study will potentially facilitate the design of more effective working memory training tasks in the future.

At the same time, the study addressed the issue of the role of working memory in intelligence. While this question is less novel, the methodology utilized allows for some clarification of a controversial issue. As previously stated, generalized latent variable modeling has the benefits of traditional structural equation modeling (i.e., a purified, error-free measure), without some of the major drawbacks, such as the possibly faulty normality and homogeneity assumptions and the difficulty to substantively validate the theoretically derived correlations and structural relations. Therefore, it is hoped that the results of this study will help resolve what has been an ongoing and highly-contentious issue.

A related but secondary objective is to investigate whether working memory components – individually or in some combination – provide more explanatory power with regard to academic achievement than do either or both of the traditional measures of intelligence. As previously discussed, work by Luo et al. (2003) suggested that the Performance index may play a minimal role in academic achievement and that other measures (such as processing speed tasks) provide more explanatory power. Thus, it may well be the case that working memory tasks provide one such superior measure. By providing clarification of the role of fluid intelligence in academic achievement relative to other important predictors (e.g., working memory and crystallized intelligence), the validity of this construct may be enhanced. Specifically, the study will help clarify the relationships between fluid intelligence, working memory, and academic achievement.

CHAPTER TWO

METHODS

Participants

This study used archival data collected from 1197 elementary school children in Shanghai and Yanchen, China. These children were recruited for a project investigating the cognitive etiology of mild mental retardation (MMR; Luo et al., 2010). All of the children were in grade three or four,and age was treated as a covariate in analyses to control for its possible impact on performance. Of these 1197 children, 140 had previously been diagnosed with mild mental retardation absent co-existing behavioral or health problems, while the rest were in the normal cognitive range. A total of 207 participants had missing values on one or more measures and listwise deletion was used, leaving a total of 990 cases (123 in the MMR group). Because the size of the MMR group was disproportionate to that than found in the population (approximately 2%), these cases were weighted down to be comparable (Luo et al.)

Measures

A large battery of tests was administered to the children in the sample. The battery included elementary cognitive tasks, cognitive ability tests, and academic achievement tests; however, only data from the latter two types of tests was included in this study. The cognitive ability tests included subtests from the Chinese translation of the WISC-R (WISC-RC). In the construction of the WISC-RC, most items and instructions were directly translated from the American version; however, a few items on some of the verbal subtests (such as Vocabulary) were altered to better fit the Chinese context (Dan, Jin, Vadenber, Yuemei, & Caihong, 1990; Dai & Lynn, 2001).

Digit Span (DS): DS is one such subtest from the WISC-RC, and a working memory task with a storage demand. It contains two parts: forward digit span (FDS), and backward digit span (BDS). In FDS, the examinee must repeat an orally-presented list of digits varying from two to nine digits in length; each digit length has two trials. In (BDS), the examinee must repeat this list of digits in backwards order, thereby adding a processing demand. The digit lengths vary from two to eight, and each length has two trials. The split-half reliability of DS is .85, and the test-retest reliability is .77 (Dai et al., 1990).

Letter-Number Sequencing (LNS): LNS is another well-established working memory task, and is included in the current English-Language version of the Wechsler Intelligence Scale for Children (WISC-IV). In this task, the participant must sort orally-presented lists consisting of letters and digits mixed together into numerical and alphabetical order. Therefore this task requires both storage and a processing demand. The lists range in length from two to eight digits and letters, and there are three trials of each digit length. The split-half reliability of this subtest is .69, and the test-retest reliability is .67 (Dai et al., 1990).

Verbal and performance summary scores: The four Verbal subtest scores from the WISC-RC (Comprehension, Information, Similarities, and Vocabulary) were summed to form a Verbal summary score. Similarly, the four Performance subtest scores (Block Design, Object Assembly, Picture Arrangement, and Picture Completion) were summed to form a Performance score.

Math Test: This test consists of 30 multiple-choice questions about mathematical concepts, the application of formulas, and applied problems.

Chinese Test: This test also consists of 30 multiple-choice questions relating to vocabulary, sentence structure, and reading comprehension.

Analysis

The methods of generalized latent variable modeling were used to conduct all model testing and comparison in this study. This general term encompasses a variety of latent variable models including finite mixture modeling and latent class modeling (Skrondal & Rabe-Hesketh, 2004). Finite mixture modeling is used to model discrete latent variables – indicating latent population heterogeneity—that underlie continuous observed variables, while latent class modeling is used with discrete latent traits and discrete latent variables. Similar methods can also be applied to continuous latent traits for discrete indicators (as in traditional item response theory) and various combinations of continuous and discrete latent traits (Vermunt, 2001). In the present study, responses to all working memory items were analyzed at the item level. The observed variables were on an ordinal scale (reflecting the number of trials successfully completed), while a logistic link function related these indicators to the underlying latent traits, as described in the previous chapter. Verbal, Performance, mathematics and Chinese responses were analyzed at the task level, with discrete latent variables underlying continuous observed variables.

The initial part of the study was a replication of the modeling of working memory traits reported by Luo et al. (2010). As in the prior study, model selection followed two standard procedures. First, models with varying numbering numbers of continuous traits were tested, as working memory and intellectual abilities are conventionally treated as continuous traits. The working memory traits were specified as orthogonal to each other and therefore represent independent sources of variance. By the same token, the latent traits underlying the Verbal and Performance scales are conventionally treated as overlapping with working memory traits; therefore both of these traits were allowed to covary with each of the working memory traits. Then, following selection of the best-fitting continuous-trait model, the possibility of a discrete-trait model being a better fit was examined. For example, if a one-continuous trait model was found to fit the data better than models with multiple continuous traits, then models with one discrete trait and varying numbers of levels (i.e., categories) would then be tested until an optimal fit was found. In practice, models with discrete latent traits invariably provided a closer fit and were additionally selected because of the theoretical considerations discussed in the previous chapter.

The model fit index used to select from competing models was the Bayesian Information Criterion (BIC), which is a log likelihood index based on both model-data discrepancy and model parsimony (Schwarz, 1978). Additionally, the Likelihood Ratio Chi-Squared Index (L^2), an asymptotic chi-squared index, is available when all observed indicators are discrete (as for the working memory items); when possible, this index was used to generate bootstrapping significance tests of fit of the selected models. When the L^2 index was not available, as when some observed indicators are continuous (as for everything other than the working memory items), the Log Likelihood (LL) and BIC indices only were used (Vermunt & Magidson, 2005). In all cases, the lower the index, the better the model fit.

As described in the previous chapter, generalized latent variable modeling produces classification results for each individual. When the latent trait is discrete, the posterior membership probability is first estimated for each trait level. Thereafter, the individual is assigned to the level with the highest probability. The classification process can also produce useful information regarding the overall distribution of the latent trait. Although latent trait scores are evenly spaced on a 0-1 scale (such that for three categories/levels, the scores would be 0, 0.5, and1), the frequency distribution does not necessarily take a particular shape.

Procedures

Models involving working memory tasks, Verbal, Performance, and achievement tests were tested, first separately and then in combination.

Modeling working memory: Based on the previous work by Luo (2008), working memory factors (STS, GAC, MS, and BO) were specified to load on the items of particular subtests. While STS was specified to load on every item, GAC was specified to load on only the BDS and LNS items, MS on only the LNS items, and BO on only the BDS items. Initially, only subgroups of the WM factors were estimated (FDS & LNS; LNS & BDS; FDS& BDS), in order to allow for factors to be distinguished from each other. In each grouping, models with different numbers of continuous and discrete traits were first compared. When it was found that discrete models provided a better fit for the data, models with different levels of discrete traits were then compared. Using the factor mean scores generated during model estimation, intercorrelations were then calculated to ensure that the factors were reasonably independent.

Additionally, a parametric bootstrapping procedure was used with the optimally-fitting models to investigate fit in the absolute rather than relative sense; bootstrapping checks whether the “optimal” model in fact provides a satisfactory fit for the data (Efron, 1979). Five hundred Monte Carlo replications generated the probability distribution defined by the parameter estimates of the selected model in this L^2-based procedure.

Modeling intelligence (Performance, Verbal) and achievement factors: These complex measures of cognitive ability are likely to include numerous sources of variance (Deary, 2000; Jensen, 1998); however, distinguishing all these difference sources was not the purpose of the current study. Instead, a goal was to understand the role of the working memory mechanisms in each. Therefore, item-level analyses of these measures were not attempted. Instead, factors were defined from total scaled scores of each of the relevant subtests. Models with continuous factors of varying number were first tested. After the optimal number of continuous factors was determined, models with the same number of discrete factors but varying levels were then tested. Additionally, factor loading estimates on each of the relevant subtests were obtained.

Modeling working memory with intelligence factors: The next step was to investigate the relationship between working memory and each of these factors based on the best-fitting models already identified. Path models in which the working memory factors were specified as exogenous latent variables, and the intelligence factor was an endogenous latent variable was specified in order to examine the unique contribution of each working memory mechanism (Vermunt & Magidson, 2008b).

Two methods were used to evaluate the contribution of each working memory factor to the criterion factor. First, the contribution of each exogenous variable was assessed by examining the regression weights (z-values) for the paths from each exogenous variable to the endogenous variable, along with the corresponding p-value. Thus, the relative magnitude of contribution could be assessed by comparing the size of the z-values for all factors for which $p<0.05$. Additionally, corresponding association models were estimated for each path model so that posterior probability estimates were generated. The mean factor scores were then used in hierarchical multiple regression models.

Modeling intelligence and achievement: The contribution of the intelligence factors was assessed as described previously, by examining the regression weights (z-values) and multiple regression (primarily R^2 changes). However, an additional procedure was included as a means of further validation. A series of nested models were tested, in which a path or paths to the Achievement factor were constrained to zero. When a constrained path (e.g. Verbal-to-Achievement) resulted in a markedly worse fit (as indicated by a relatively larger BIC index) when compared to the full model (in which no paths were constrained), further substantiation of the contribution of the constrained factor was obtained. On the other hand, evidence that the constrained factor did not make an important contribution was obtained when there was little change in terms of model fit.

Modeling working memory, intelligence, and achievement: Attempts to include all working memory indicators along with one or both intelligence indicators and the achievement indicators proved to be too computationally demanding. Therefore, a decision to include only subsets of working memory indicators (and therefore two instead of four working memory factors) was made. The FDS indicators were always included so that a STS factor could always be included. The LNS items were generally used in preference to the BDS items as the LNS items loaded more strongly on the GAC factor. However, some modeling with the STS and BDS subset of items was also done in order to provide further validation.

Statistical Program

The program Latent Gold 4.5 (Vermunt & Magidson 2005, 2008) was used for all generalized latent variable modeling in this study. This software allows for the inclusion of a diverse array of variables, both latent and observed, discrete (ie. nominal and ordinal) and continuous. Additionally, both linear and non-linear relations between variables may be included. For ordinal item responses (as is the case for the working memory tasks in this study), an adjacent-category logistic is the default link function. This program also allows users to generate classification results and latent trait scores based on the estimated posterior probabilities. When the latent traits are on a continuous scale, these scores are equivalent to traditional factor scores or person parameter estimates. When the latent traits are on an ordinal scale, individuals are classified according to the most probable trait level and are additionally assigned a factor mean score. These factor mean scores can then be used to evaluate inter-trait correlations. Association models were specified in graphic mode, while path models could only be specified in syntax mode.

CHAPTER THREE

RESULTS

Preliminary Analysis

Pearson's product-moment correlations were calculated and are displayed in Table 1, along with the reliability estimates, and means and standard deviations.

Table 1

Correlations, Reliability Estimates, and Descriptive Statistics

	DS	LNS	V	P	C	M
DS	(0.78)	0.49	0.45	0.35	0.43	0.39
LNS		(0.71)	0.48	0.45	0.49	0.43
V			(0.92)	0.57	0.63	0.51
P				(0.84)	0.49	0.43
C					(0.82)	0.70
M						(0.87)
Mean	16.48	9.12	41.51	42.06	19.88	19.17
SD	3.75	2.61	11.13	8.42	5.42	6.53

Note. DS is the Digit Span total score, LNS is the Letter-Number Sequencing total score, V is the sum of the four WISC-RC Verbal subtests, P is the sum of the four WISC-RC Performance subtests, C is the Chinese total score, and M is the Mathematics total score. Reliability estimates (Cronbach's Alpha Co-efficients) are displayed on the diagonal.

Working Memory Factors

An item-level analysis of the working memory indicators was performed, based on the previous generalized latent variable modeling by Luo et al. (2010). Working memory factors (STS, GAC, MS, and BO) were specified to load on the items of particular subtests; while STS was specified to load on every item, GAC was specified to load on only the BDS and LNS items, MS on only the LNS items, and BO on only the BDS items.

Initially, only subgroups of the WM factors were estimated, in order to allow for factors to be distinguished from each other. In each grouping, models with different numbers of continuous and discrete traits were first compared. When it was found that discrete models provided a better fit for the data, models with different levels of discrete traits were then compared.

FDS and LNS: In this subgrouping of WM indicators, the first factor was constrained to load only on the LNS items (thus representing GAC and MS), while the second factor loaded on all FDS and LNS items (thus representing STS). The optimal model was one in which both factors were discrete and there were seven levels of each factor. Several models were tested, including the ones shown in Table 2.

Table 2

Item-level FDS and LNS Factors

Model	df	p(bootstrap)	BIC
2 continuous factors	807	0.00	14022.53
2 discrete factors (8,8)	802	0.00	13750.40
2 discrete factors (7,8)	804	0.00	13748.59
2 discrete factors (6,8)	804	0.00	13749.05
2 discrete factors (7,7)	**804**	**0.00 (0.20)**	**13745.76**
2 discrete factors (7,6)	805	0.00	13750.12

The correlation between the two factor mean scores was significant (r=.14) at $p<0.01$, but weak, indicating the posterior estimation largely retains the orthogonal property of the priori model.

LNS and BDS: In these models, there were three factors: one for STS and GAC combined (or the variance shared by both indicators), one for MS, and one for BO. The first factor was allowed to load on all items, while the second factor was constrained to load only on the LNS items, and the third factor constrained to load only on the BDS items. Several models were tested, and the optimal model was found to have all discrete factors, with five levels for STS and GAC, six levels for MS, and four levels for BO. Model fit indices for some of the models tested are shown in Table 3.

Table 3

Item-level LNS and BDS Factors

Model	df	p(bootstrap)	BIC
3 continuous factors	803	0.00	13759.79
3 discrete factors (5,6,6)	799	0.00	13600.70
3 discrete factors (5,6,5)	800	0.00	13597.86
3 discrete factors (5,6,4)	**801**	**0.00 (0.09)**	**13577.80**
3 discrete factors (5,6,3)	802	0.00	13617.26
3 discrete factors (6,6,4)	800	0.00	13612.59

FDS and BDS: In this subgrouping, STS was allowed to load on all FDS items, while GAC with BO were constrained to load only on the BDS items. Again, the optimal model had only discrete factors with eight levels of the first, and eight of the second factor. The correlation between the two factor mean scores was again weak, and significant ($r=.13$) at $p<0.01$. Table 4 shows model fit data for some of the models tested.

Table 4

Item-level FDS and BDS Factors

Model	df	p(bootstrap)	BIC
2 continuous factors	831	0.00	11424.36
2 discrete factors (8,8)	**817**	**0.00 (0.13)**	**11314.41**
2 discrete factors (7,8)	817	0.00	11361.21
2 discrete factors (7,7)	818	0.00	11348.93

All working memory indicators: Based upon the subsets of indicators already tested and the factors distinguished, models with all working memory indicators were included. The optimal model had five levels of STS, six levels of GAC, five for BO, and five for MS. Model fit data for some of the models tested is shown in Table 5.

Table 5

Item-level FDS, BDS, and LNS Factors

Model	df	p(bootstrap)	BIC
4 discrete factors (6,6,6,6)	759	0.00	19165.90
4 discrete factors (6,5,6,6)	760	0.00	19158.42
4 discrete factors (6,5,6,5)	764	0.00	17967.14
4 discrete factors (5,6,6,6)	760	0.00	19183.74
4 discrete factors (5,6,5,6)	764	0.00	17939.05
4 discrete factors (5,6,5,5)	**765**	**0.00 (0.68)**	**17926.97**

Table 6 shows the correlations between the factor means scores obtained in the estimation of this model.

Table 6

Intercorrelations Among Working Memory (WM) Factor Mean Scores

	STS	GAC	BO	MS
STS	--	.14**	.08*	0.12
GAC		--	.13**	.13**
BO			--	.04
MS				--

*p<.05, **p<.01

Additionally, Figure 1 shows the factor loading estimates of the working memory factors, all of which are significant at p<0.05. These estimates are linear approximations of the model slope parameters, which reflect the magnitude of the relationship between the latent trait and the specific item (Vermunt & Magidson, 2005).

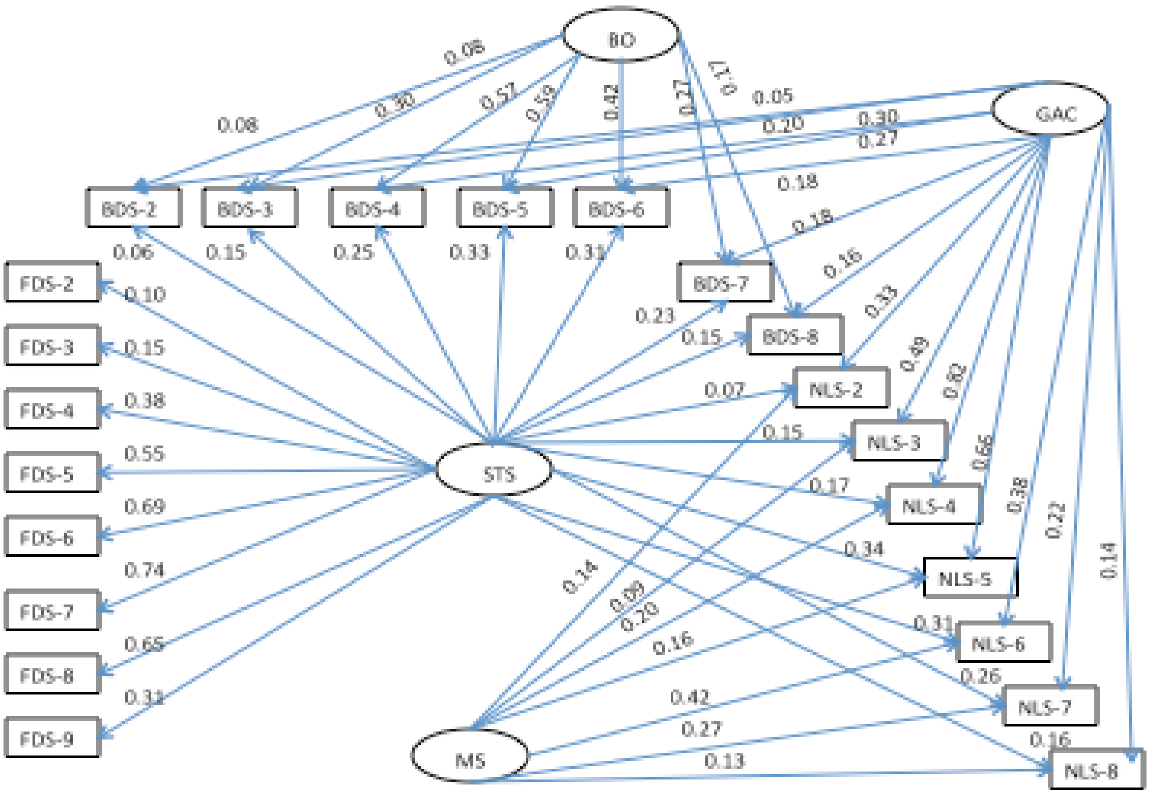

Figure 1. Factor loading estimates of working memory factors.
Note. Numbers in rectangles correspond to item numbers.

Performance Factor

As previous stated, a scaled score Performance total was used as the sole observed indicator. A model with a single factor was found to fit better than a model with two factors, and the best-fitting model had one discrete factor with five levels, as shown in Table 7.

Table 7

Task-level Performance Factor

Model	LL	BIC
1 continuous factor	-8280.8436	16643.11
2 continuous factors	-8278.2316	16665.03
1 discrete factor (3 levels)	-8275.8317	16646.66
1 discrete factor (4)	-8246.7556	16595.29
1 discrete factor (5)	**-8237.9923**	**16584.55**
1 discrete factor (6)	-8236.6639	16588.68

Note. L^2- based bootstrapping values are not available for models that include continuous variables.

Additionally, figure 2 shows the factor loading estimates of the Performance factor for each of the four subtests.

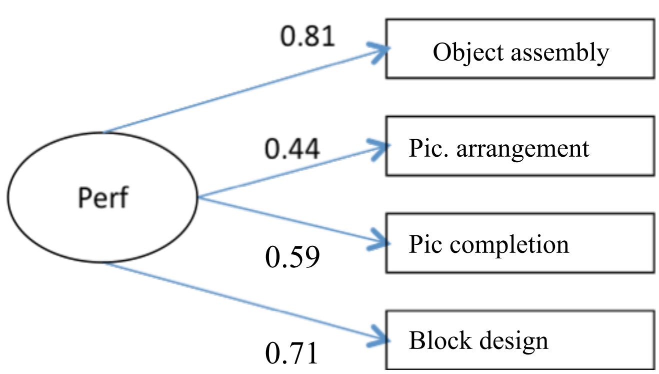

Figure 2. Factor loading estimates of Performance factor.

Working Memory and Performance Factors

An extended model that included the best fitting working memory and Performance models already identified was estimated, to investigate the unique contribution of the working memory factors to Performance. Using syntax mode in Latent Gold 4.5 (Vermunt & Magidson 2005, 2008), a path model in which STS, GAC, BO, and MS were independent exogenous variables and Performance was an endogenous variable was specified. All factors were specified as discrete, with five levels for STS, six for GAC, five for BO, six for MS, and five for Performance.

Two methods were used to evaluate contributions from the working memory mechanisms. First, the regression weights for the paths from working memory factors to Performance were examined, and are shown in Table 8. The path from STS to Performance and the path from GAC to Performance were significant at $p<0.05$.

Table 8

Paths from WM Factors to Performance

Parameter term	Co-efficient	s.e.	z-value	p-value
STS to Performance	6.94	1.19	5.83	0.00
GAC to Performance	7.13	0.99	7.22	0.00
BO to Performance	0.56	0.92	0.61	0.54
MS to Performance	0.22	0.91	0.25	0.80

The reliabilities (standard R^2) ranged from 0.54 for MS to 0.85 for STS. Additionally, the posterior trait scores for each of the factors was obtained and used to calculate Pearson correlations, as shown in Table 9.

Table 9

Intercorrelations Among WM and Performance Factor Mean Scores

	STS	GAC	BO	MS	P
STS	--	.14**	.14**	.00	.37**
GAC	--	--	.15**	.11**	.53**
BO	--	--	--	.048	.045
MS	--	--	--	--	.040
P	--	--	--	--	--

*p<.05, **p<.01

The working memory factors together accounted for 37.1% of the variance of Performance (R^2), and significant unique contributions were made by STS, GAC, and BO. STS contributed 9.4% of the variance of Performance, GAC contributed 23.2%, and BO contributed .5% above and beyond the other factors.

Verbal Factor

A Verbal factor was modeled in the same manner as the Performance factor, using the scaled score totals as observed indicators. Initially, a model with one continuous factor was compared to a model with two continuous factors. As the one-factor model fit had a better fit, it was next compared to models with various levels of a discrete factor. Overall, optimal fit was provided by the model with six levels of the discrete factor. Table 10 provides comparative information for some of the models tested.

Table 10

Task-level Verbal Factor

Model	LL	BIC
1 continuous factor	-8554.47	17198.36
2 continuous factors	-8545.61	17199.78
1 discrete factor (4 levels)	-8565.29	17232.35
1 discrete factor (5)	-8552.53	17213.61
1 discrete factor (6)	**-8539.00**	**17195.32**
1 discrete factor (7)	-8540.08	17202.28

Note. L^2- based bootstrapping values are not available for models that include continuous variables.

In addition, figure 3 shows factor loading estimates for the Verbal factor.

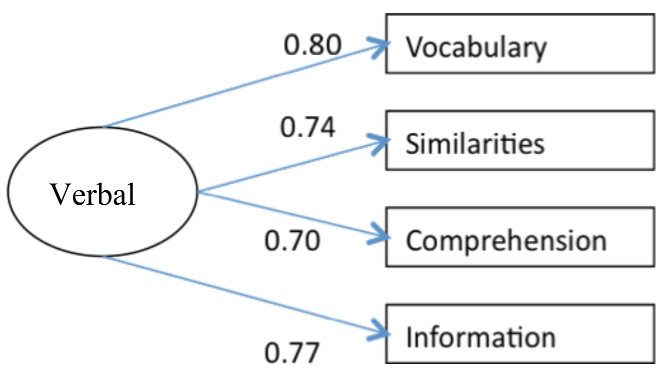

Figure 3. Factor loading estimates of Verbal factor.

Working Memory and Verbal Factors

A path model was specified in order to evaluate the unique contribution of each working memory factor to the Verbal factor. Based on previous modeling, all factors were specified as discrete, with five levels of STS, six levels of GAC, five levels of BO, six levels of MS, and six of Verbal. In the first stage of the evaluation, the parameter estimates from the working memory factors to the Verbal factor were examined, and are shown in Table 11 below. The path from STS to Verbal, the path from GAC to Verbal, and path from MS to Verbal were all significant at $p<0.05$.

Table 11

Paths from WM Factors to Verbal

Parameter term	Co-efficient	s.e.	z-value	p-value
STS to Verbal	12.19	1.72	7.08	0.00
GAC to Verbal	18.78	3.32	5.65	0.00
BO to Verbal	2.69	1.93	1.39	0.16
MS to Verbal	6.06	1.88	3.22	0.00

The reliabilities (Standard R^2) of each factor were as follows: STS=0.85, GAC= 0.81, BO=0.65, MS= 0.53, and Verbal = 0.84. Correlations between factor mean scores used in this model are shown in Table 12.

Table 12

Intercorrelations Among WM and Verbal Factor Mean Scores

	STS	GAC	BO	MS	V
STS	--	.15**	.10**	.01	.39**
GAC	--	--	.11**	.16**	.41**
BO	--	--	--	.038	.10**
MS	--	--	--	--	.07
V	--	--	--	--	--

*p<.05, **p<.01

The working memory factors together accounted for 25.15% of the variance of Verbal (R^2). The unique variance contributed by STS was a significant 7.7% of Verbal. GAC contributed a significant 12.8% of the variance above and beyond the other working memory factors. Neither MS nor BO made any significant unique contribution.

A nested model was used to further investigate whether MS makes a significant unique contribution to Verbal. In the nested model, the MS-to-verbal path was constrained to zero; the fit of this model was compared with the full model with no paths constrained. The BIC (LL) fit index for the full model was 34108.65, while that of the nested model was only marginally worse, at 34126.91. The Akaike Information Criterion 3 (AIC3) index for the full and nested models were also compared, with that of the nested model being smaller and therefore preferable (33612.27, compared to 33615.14). The weight of the evidence therefore indicated that MS does not make a meaningful contribution to Verbal.

Achievement Factor

The total scaled scores for mathematics and Chinese were used as observed indicators for an achievement factor. After testing competing models (as shown in Table 13), a model with four levels of a discrete factor was found to provide the best fit.

Table 13

Task-level Achievement Factor

Model	LL	BIC
1 continuous factor	-5305.10	10644.07
2 continuous factors	-5305.10	10650.84
1 discrete factor (3 levels)	-5198.18	10450.55
1 discrete factor (4)	**-5193.97**	**10448.90**
1 discrete factor (5)	-5193.17	10454.08
1 discrete factor (6)	-5193.22	10460.96

Note. L^2- based bootstrapping values are not available for models that include continuous variables.

However, the model with a continuous factor could not be identified without some additional constraints, because the number of factor loading parameters to be estimated exceeded the single observed covariance to be fitted to. One possible solution is to add an additional constraint so that the model will be just identified, such as by constraining the two loadings to be equivalent or choosing an arbitrary value for one of the two values. Therefore, the two loadings were first constrained to be equivalent, and the BIC (LL-based) estimate was 10654.22. Next, the loading estimate for Chinese was fixed (using the loading estimate obtained previously), with the resulting BIC (LL-based) estimate of 10644.07. While these models fit worse than those with discrete factors, they might reasonably be considered over-constrained.

Another possible solution to this identification problem is to demonstrate that the Chinese and Math indicators can be considered mixtures of distributions rather than homogeneous distributions, thereby showing that a continuous-factor model (which is based on the assumption that the observed indicators follow homogeneous normal distributions) is invalid. The homogeneity of the distributions was tested by modeling various numbers of clusters, as shown in Table 14.

Table 14

Testing Homogeneity of Achievement Indicator Distributions

Indicator	Model	LL	No. of parameters	BIC
Chinese	1-cluster	-2719.56	2	5452.67
	2-cluster	**-2640.14**	**5**	**5314.15**
	3-cluster	-2631.51	8	5317.21
Math	1-cluster	-2892.02	2	5797.59
	2-cluster	**-2794.30**	**5**	**5622.49**
	3-cluster	-2784.91	8	5624.04

The results shown in Table 14 indicate that neither observed indicator follows a homogeneous univariate normal distribution. As the assumption about the univariate distributions represented by the one-cluster models does not appear valid, the assumption of a homogeneous bivariate normal distribution for the two indicators also appears implausible. Therefore, it was concluded that the model of a normally distributed latent factor underlying a homogeneous bivariate normal distribution is inappropriate for the two indicators.

It should be noted that the models with a discrete factor could be more readily identified. This is because a discrete-factor model is a mixture model in which the number of known quantities exceeds the number of factor loadings to be estimated. In these models, there are multiple levels (i.e. multiple latent subgroups), each of which give rise to their own observed covariances, while the number of factor loadings is two, no matter how many levels there are.

Figure 4 shows the factor loading estimates for the Achievement factor and the Chinese and Mathematics indicators.

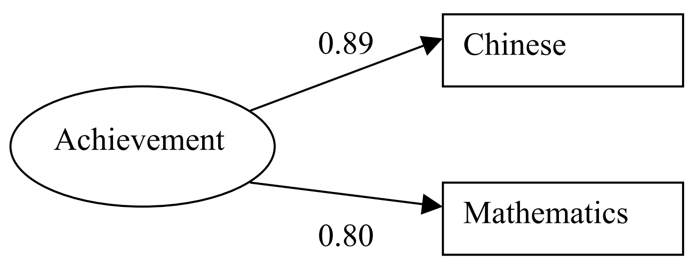

Figure 4. Factor loading estimates of Achievement factor.

Working Memory and Achievement Factors

In order to evaluate the unique contribution of each working memory factor to achievement, a path model was specified in Latent Gold's syntax mode, with independent working memory factors were specified as exogenous, and the Achievement factor specified as endogenous. Based on previous modeling, only discrete factors were used, with five levels for STS, six for GAC, five for BO, six for MS, and four for achievement. The parameter estimates are shown in Table 15, showing STS, GAC, and BO make a substantial unique contribution to achievement ($p<.05$).

Table 15

Paths from Working Memory Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
STS to Achievement	6.95	1.11	6.24	0.00
GAC to Achievement	6.12	0.72	8.46	0.00
BO to Achievement	2.02	0.63	3.19	0.00
MS to Achievement	0.59	0.82	0.72	0.47

As the significance of the BO-to-Achievement path was unexpected, significance was further tested by comparing the fit of the full model and a nested model in which this path was constrained to zero. The BIC (LL) fit index for the full model was 27776.34, compared to 27784.28 for the constrained model. Since constraining the path resulted in a fit that was only very marginally worse than the full model, it was concluded that BO does not make a substantial unique contribution to achievement.

The reliabilities of the factors (standard R^2) were: STS = 0.85, GAC= 0.81, BO=0.64, MS= 0.58, Achievement= 0.88.

Intercorrelations between the factor mean scores are shown in Table 16.

Table 16

Intercorrelations Among WM and Achievement Factor Mean Scores

	STS	GAC	BO	MS	A
STS	--	.46**	.16**	.09**	.62**
GAC	--	--	.16**	.15**	.47**
BO	--	--	--	.021	.17**
MS	--	--	--	--	.13**
A	--	--	--	--	--

*p<.05, **p<.01

The working memory factors together accounted for 33.1% of the variance of Achievement. The contribution of GAC above and beyond the other factors was 16% of the variance of achievement, while the contribution of STS was 8.8%. Both MS and BO contributed a significant .09%.

Achievement and Non-Working Memory Factors

Paths between Achievement and Performance, and between Achievement and Verbal were estimated separately. In the first of these, a path model with Performance as an exogenous factor, and Achievement as an endogenous factor was specified, and Performance was found to make a significant unique contribution to Achievement.

The correlation between the two factor mean scores was .49, significant at the 0.01 level. Similarly, a path model with Verbal as an exogenous factor, and Achievement as an endogenous factor was specified. In this model, Verbal was found to make a substantial unique contribution to Achievement. Table 17 shows the parameter estimates for each of the non-working memory paths. The correlation between the Verbal factor mean score and the Achievement factor mean score was .59, also significant at the 0.01 level.

Table 17

Paths from Non-WM Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
Perf. to Achievement	9.96	1.18	8.45	0.00
Verb. to Achievement	19.75	2.05	9.62	0.00

To investigate further the role of each exogenous factor, the fit of full and nested models was compared, and the results are shown in Table 18. Reliabilities of all factors were in the 0.80 to 0.96 range.

Table 18

Fit of Full and Nested Models: Verbal, Performance, and Achievement

Model	LL	BIC (LL)
Full (no paths constrained)	-17637.75	36745.32
Perf. path constrained	-17654.75	36780.73
Verb. path constrained	-17722.22	36921.29

Performance and Verbal together were found to account for 51.7% of the variance of Achievement. Verbal provided a significant 6.6% contribution above and beyond Performance, while Performance accounted for 14.6% above Verbal.

Working Memory, Performance, and Achievement Factors

A subset of working memory items (FDS and LNS) were included in a path model in which the working memory factors and Performance were specified as associated (ie. correlated) predictors for Achievement. The paths from all three factors were significant at $p<0.5$, and are shown in Table 19.

Table 19

Paths from Working Memory and Performance Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
STS to Achieve.	6.37	1.14	5.61	0.00
GAC+MS to Achieve.	6.56	1.20	5.48	0.00
Perf. to Achieve.	10.15	1.62	6.28	0.00

Correlations between the mean factor scores were significant at 0.01 for STS and GAC+MS ($r=.27$), STS and Performance ($r=.45$), STS and Achievement ($r=.42$). Additionally, significant correlations at the same level were found between GAC+MS and Performance ($r= .63$), GAC+MS and Achievement (.55), and Performance and Achievement (.73).

In multiple regression analyses, the three indicators together accounted for 57.4% of the variance of Achievement. The working memory factors together contributed a significant 12.3% of the variance of Achievement above and beyond Performance, while Performance contributed a unique and significant 9.9%. The unique contribution of STS was 4.6% of the variance, while the unique contribution of GAC+MS was 7.5%.

The fit of a series of nested models was then compared, as shown in Table 20.

Table 20

Fit of Full and Nested Models: LNS, FDS, Performance, and Achievement

Model	LL	BIC (LL)
Full (no paths constrained)	-16150.693	32963.66
Perf. path constrained	-16189.077	33033.67
STS path constrained	-15848.56	33017.84
GAC+MS constrained	-15854.81	33030.86
Both WM paths constrained	-15864.34	33050.70

A similar path model that differed only in that it included BDS items in place of LNS items was also tested. In the full model, factor reliabilities (R^2) ranged from 0.77 for GAC+BO to 0.95 for Achievement. Correlations were significant at the 0.01 level for STS and GAC+BO (r=.22), STS and Performance (r=.41), and STS and Achievement (r=.39). In addition, correlations that were significant at the 0.01 level were found between GAC+BO and Performance (r=.36), GAC+BO and Achievement (r=.50), and Performance and Achievement (.62). The paths from all three factors to Achievement were significant at $p<0.05$, as is shown in Table 21.

Table 21

Paths from Working Memory and Performance Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
STS to Achieve.	3.75	0.89	4.24	0.00
GAC+BO to Achieve.	4.28	0.69	6.17	0.00
Perf. to Achieve.	8.39	1.07	7.81	0.00

The fit of a series of nested models was then compared with that of the full model, as shown in Table 22.

Table 22

Fit of Full and Nested Models: FDS, BDS, Performance, and Achievement

Model	LL	BIC (LL)
Full (no paths constrained)	-14929.44	31103.03
Perf. path constrained	-14976.91	31204.56
STS path constrained	-14624.22	31115.37
GAC+BO constrained	-14954.11	31154.39
Both WM paths constrained	-14963.56	31174.08

Working Memory, Verbal, and Achievement Factors

The first model included only FDS and LNS of the working memory factors, and the following factors were specified: STS, GAC+MS, Verbal and Achievement. A path model was specified, with Achievement as the only endogenous factor for the associated exogenous factors. Reliabilities (Standard R^2) for all four factors were in the 0.8-0.9 range.

Correlations between the mean factor scores were significant at 0.01 for STS and GAC+MS (r=.29), STS and Verbal (r=.48), STS and Achievement (r=.48). Additionally, significant correlations at the same level were found between GAC+MS and Verbal (r= .57), GAC+MS and Achievement (.61), and Verbal and Achievement (.76).

The paths from the Working Memory factor and Verbal factor to Achievement were all significant at $p<0.05$, and are shown in Table 23.

Table 23

Paths from Working Memory and Verbal Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
STS to Achieve.	5.19	1.23	4.24	0.00
GAC+MS to Achieve.	4.73	0.90	5.25	0.00
Verbal to Achieve.	16.44	2.04	8.05	0.00

The fit of a series of nested models was then compared with the full model. As shown in Table 24, the full model provided the optimal fit.

Table 24

Fit of Full and Nested Models: FDS, LNS, Verbal, and Achievement

Model	LL	BIC (LL)
Full (no paths constrained)	-15234.24	31738.27
Verb. path constrained	-15293.74	31861.96
STS path constrained	-15241.07	31752.23
GAC+MS constrained	-15246.6	31763.74
Both WM paths constrained	-15253.32	31777.74

In the multiple regression analysis using the mean factor scores, 62.9% of the variance of Achievement was accounted for by Verbal, STS, and GAC+MS. Verbal accounted for a significant 22.6% of the variance above and beyond the working memory factors, while GAC+MS accounted for a significant 1.4% above and beyond the other factors, and STS a significant .3%. The two working memory factors together uniquely contributed a significant 1.8%.

Working Memory, Verbal, Performance, and Achievement Factors

A model including LNS and FDS indicators along with Verbal, Performance and Achievement was estimated, with Achievement as the sole endogenous variable. The Standard R^2 reliabilities of all factors included in the model (STS, GAC+MS, Performance, Verbal, and Achievement) were above 0.80, and the paths from all exogenous variables to Achievement were all significant at $p<0.05$, as shown in table 25.

Table 25

Paths from Working Memory, Performance, and Verbal Factors to Achievement

Path	Co-efficient	s.e.	z-value	p-value
STS to Achieve.	5.43	1.20	4.54	0.00
GAC+MS to Achieve.	5.20	0.98	5.31	0.00
Perf to Achieve.	5.29	1.33	4.06	0.00
Verbal to Achieve	13.95	1.89	7.39	0.00

Table 26 compares the fit of the full and nested models tested.

Table 26

Fit of Full and Nested Models: FDS, LNS, Verbal, Performance, and Achievement

Model	LL	BIC (LL)
Full (no paths constrained)	-23155.16	48239.92
STS path constrained	-23168.49	48267.68
GAC+MS constrained	-23161.35	48252.81
Perf path constrained	-23161.34	48252.80
Verbal path constrained	-23199.69	48332.68
Both WM paths constrained	-23176.20	48283.74

Table 27 shows the correlations between the mean factor scores.

Table 27

Intercorrelations Among Factor Mean Scores in Final Model

	STS	GAC+MS	P	V	A
STS	--	.23**	40**	.39**	.46**
GAC+MS	--	--	.53**	.46**	.58**
P	--	--	--	.59**	.64**
V	--	--	--	--	.71**
A	--	--	--	--	--

**p<.01

Multiple regression using the mean factor scores showed that STS, GAC+MS, Performance and Verbal accounted for 64.2% of the variance of Achievement. All factors made a significant unique contribution to the variance of Achievement (R^2 change), with STS contributing 2.1%, GAC+MS contributing 4.3%, Performance contributing 2%, and Verbal contributing 10.6%. The unique contribution of the two WM factors together was 6.2%, compared to 19.2% for Verbal and Performance together. Additionally, when controlling for Verbal, WM had greater explanatory power than Performance, with Performance adding 7.5%, and the two WM factors adding 11.7%.

Posterior Classification of Participants

Finally, the frequency count of participants into the different levels of the latent factors was examined. The seven levels of STS and GAC + MS can be labeled 1-7 with 1 representing that group of participants with the lowest ability level and 7 representing the group with the highest ability level. For each of the two working memory factors, both the second and third levels of the factors were completely devoid of participants. In other words, there was a marked gap between the first and fourth levels in the distribution of participants.

CHAPTER FOUR

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary of Findings

This study provided clarification of the factor structure of working memory, and its relationship with intelligence and achievement. Factors that are common across multiple working memory tasks (specifically, STS and GAC) were found to make an important contribution to both intelligence and achievement, while task-specific factors (BO and MS) were not. Furthermore, in this study the working memory factors combined were clearly better predictors of achievement than traditional measures of intelligence (Verbal and Performance). At the same time, results of this study indicate that both of these traditional measures of intelligence make significant and unique contributions to academic achievement above and beyond those of the working memory factors.

The most unique aspect of this study was the examination of the relationship between *independent* latent working memory factors and a latent achievement factor. Unlike previous studies which did not differentiate between the role of STS and the role of GAC, this study was able to provide more precise information about the nature of working memory's contribution. Thus, it was possible to discern that the general factors of GAC and STS *both* made substantial unique contributions and that the contributions of the more specific mechanisms were much lesser.

The use of an item-level analysis facilitated this greater level of precision. Each trait was identified by multiple items, and multiple traits were identified through the concurrent analysis of items with different task demands. Such an approach would not have been possible had the analysis occurred at the task level, because of the difficulties inherent in designing the range of task demands necessary to identify each trait.

The use of generalized latent variable modeling allowed for the identification of latent traits according to the axiom of local independence for item responses. This axiom states that any association (linear correlation or otherwise) existing between item responses is solely attributable to a latent trait. In traditional structural equation modeling, traits are identified only indirectly through the pattern of correlations between tasks. Thus, in the vast majority of previous latent-level investigations into the relationship between working memory and important criteria such as intelligence and achievement, the working memory factors have been correlated rather than orthogonal. For example, Oberauer et al. (2003) defined storage in the context of processing, updating, short-term memory and shifting as correlated factors, while Miyake et al. (2000) defined shifting, updating, and inhibition. In such designs, variance that is *shared* between different kinds of tasks cannot be examined. However, it is clear that each kind of task has multiple sources of variance, some of which are shared with other working memory tasks, and some of which are task-specific. Moreover, that portion of variance shared with other kinds of tasks is likely to come from multiple sources; for example, shifting and updating tasks are likely to share some variance based on their shared demands of storage, and some variance based on their shared demands on GAC. However, due to the limitations of this correlated factors approach, it is impossible to distinguish between these two sources.

Theoretically, an orthogonal factors approach in which each factor is defined by multiple tasks sharing particular task demands is also possible. However, this approach is somewhat impractical due to difficulties inherent in task design. Each orthogonal factor would need to be defined by at least three observed indicators, and most observed indicators would have at least three underlying factors (STS, GAC, and a more specific mechanism). Designing such a range of tasks to reliably measure such a range of factors is inherently difficult.

As previously stated, this study builds on the earlier work by Luo et al. (2010) in which orthogonal factors are defined by items rather than tasks. Many of the earlier results were replicated, and additional clarification of the role of BO was obtained. Additionally, this study was able to examine the role of the working memory components in academic achievement, independent of the contributions of traditional measures of intelligence. By specifying a path model in which the working memory factors, Performance, and Verbal were exogenous variables, and Achievement was a dependent variable, the present study was also able to evaluate the role of working memory as a whole relative to Performance and Verbal ability. Specifically, it was found that Verbal is the superior predictor of academic achievement, but that the combined working memory factors are superior to Performance.

The analysis began with a four-factor model of working memory; however, it is worth noting that the key results are based on a two-factor model. This modification was necessitated by practical considerations: the high number of latent factors and observed indicators simply overwhelmed the capacity of the current software. Therefore, it became necessary to include a more streamlined working memory, which included only two factors and two types of tasks. The FDS items were selected because of their status as putatively pure indicators of STS; while the LNS items were selected over the BDS items because of their higher loadings on the GAC factor. Together, the FDS and LNS items therefore draw on both of the general working memory factors more intensively than had another combination of indicators been selected.

At the same time, it should be emphasized that basing the most important results on a two- rather than four- factor model did not result in compromised results. Modeling of subsets of working memory factors indicated that neither of the specific factors made a significant contribution to any of the criterion variables. This fact was validated and cross-validated by the multitude of methods used to assess the unique contributions of each component. Therefore, contributions from what has been termed the GAC+MS factor can be assumed to result from the GAC factor almost exclusively.

Important clarification of the role of Performance was also provided by this study. Luo, Thompson and Detterman (2003) had earlier reported that Performance played a negligible unique role in predicting academic achievement, relative their other predictors. More specifically, in the multiple regression part of their analysis, they found that Performance explained the lowest proportion of achievement (R^2change= 0.05, compared to 0.15 for working memory and 0.18 for processing speed). Similar results were found in the structural equation modeling part of their analysis, with Performance uniquely contributing only 0.12 of the variance of their achievement factor. Important similarities exist between the current study and the study by Luo et al. (2003); both used the same indicators for the Performance, and Verbal factors. However, the Luo study had a number of limitations relative to the current study: it was a task-level analysis that treated working memory as a unitary construct, and it used traditional structural equation modeling. The reliance on a unitary model for working memory means that the Luo study was unable to identify differential roles for multiple working memory mechanisms. Additionally, latent factors were addressed by means of traditional structural equation modeling; therefore this study was unable to provide tangible individual latent trait scores, and is open to the usual criticisms of structural equation modeling, which were noted in the first chapter.

Consistent with the earlier study by Luo et al. (2010) that used the same sample and many of the same measures, an important finding of this study is that the general working memory mechanisms are predictive of intelligence and achievement in a way that the more specific mechanisms are not. While the earlier study found some suggestions that BO may have some predictive power, this was not replicated in the current study. Therefore, the current investigation provides substantiation to the idea that working memory's importance in relation to intelligence and achievement rests upon the general mechanisms rather than task-specific ones. If this indeed is the case, a better understanding of the relationships among these important psychological constructions may be facilitated by a research focus on variance shared by multiple tasks of working memory than by a focus on more specific executive functions (e.g., switching). Only future studies that analyze a greater range of working memory tasks in a manner whereby both shared and specific sources of variance can be measured will allow us to determine conclusively if this is the case.

As previously stated, the present study builds on the previous investigation by Luo et al. (2010). Both studies found a clear role for the generalized mechanisms of STS and GAC in predicting the criterion variables. Luo et al.'s study also reported some ambiguity with regard to the role of BO, and speculated that this ambiguity might be due to the lower reliability of the BO factor.

Another interesting finding was the apparent gap in the distribution of the latent traits. As previously stated, the process of posterior classification of participants assigns participants to ability levels based on their individual response patterns; examination of the resulting frequency distribution revealed zero-frequencies in the second- and third-lowest ability levels. The apparent gap between the first (ie. lowest) and fourth levels of the trait is somewhat consistent with the earlier study by Luo et al. (2010), in which the two lowest levels of GAC were adjacent to an almost empty (N=1) third category and thus “outliers” with regard to the rest of the distribution.

These results should be interpreted with caution, given that both studies used the same sample and may therefore merely reflect sample-specific characteristics. However, it is also possible that the apparent gap reflects a discontinuity in the distribution of the trait. This would be consistent with the so-called law of diminishing returns in which higher between-test correlations have been observed at the lower end of the ability continuum. Therefore, those at the low ability end may be isolated from the rest of the population in terms of their latent ability traits, and these significant differences in trait levels may have implications for performance on a number of different tests.

Compared to the earlier study by Luo et al. (2010), this study provides a clearer picture of the role of specialized working memory mechanism vis-à-vis generalized working memory mechanisms. Luo et al. reported inconsistency in his findings, such that nested models that included all working memory factors along with the criterion *g* factor suggested that the BO and MS factors were expendable, while the path estimates for these factors to the *g* factor were statistically significant, as were correlations between the mean factor scores of the predictor and criterion variables. In contrast, there is less ambiguity about the lesser importance of these specialized mechanisms in the current study. While the lower reliabilities of these mechanisms (in both studies) may contribute to the inconsistency of the results, overall results of the current study provide substantiation to the hypothesis that it is the more broadly applicable working memory mechanisms that provide working memory tasks with their predictive power.

Conclusions and Recommendations

The use of generalized latent variable analysis allowed for comparison of the fit of discrete and continuous latent variables in relation to the observed data. The fact that discrete traits, which may best be represented as distinct clusters separated by varying intervals, better accounted for the observed data on a uniform basis is an interesting finding in itself. As discrete traits may reflect a heterogeneous population, it is possible that qualitative differences exist between different ability groups. As previously noted, the only two studies that have used general latent variable analysis to investigate working memory trait distribution have relied upon the same sample. Therefore, it is important that future investigations in this area include different ages and other demographic characteristics.

Some researchers in the psychometric tradition have been critical of a reductionist approach to studying intelligence. A major charge is that this line of inquiry has been "generally limited to laboratory demonstrations" (Ackerman & Beier, 2005, p 125) and not found any real –world applications. However, the current study illustrates the potential of a reductionist approach in terms of enabling us to predict individual differences in an academic setting. If the mechanisms of working memory described in this study were not predictive of real-world performance, they would indeed be of limited interest; however, this study clearly shows that GAC and STS account for a substantial proportion of the variance in the achievement factor. Therefore, identifying ways to improve academic performance must include further micro-level analysis. As the working memory mechanisms themselves become better validated, they in turn may be used to explain more complex and ambiguous behaviors such as those found in workplaces and schools (Deary, 2001; Conway, 2005). A very practical implication of this work is that improving performance on memory span tasks through the explicit teaching of rehearsal strategies may result in improved performance on a variety of verbal tasks, such as mathematics and comprehension. Further, learning effective rehearsal strategies may be necessary foundation in order to fully benefit from standard learning and reading interventions. For example, children who are unable to memorize the phonemic sequences of unfamiliar words will be at a disadvantage in vocabulary acquisition (Conners et al., 2008).

It is worth emphasizing that working memory, intelligence, and academic achievement are all very broad constructs, and this study defined these constructs in a very circumscribed way. Further research into the role of working memory mechanisms in academic achievement should include a wide variety of working memory tasks—perhaps including dichotic listening as well as other span tasks (Heitz, Unsworth, & Engle, 2005) such that a fuller picture of the mechanisms determining performance can be obtained. Similarly, future research could substitute different kinds of achievement tasks.

Perhaps more importantly, the present study illuminates the potential of the blending of item-level analysis with the new technology of generalized latent variable modeling. Just as the current study was able to identify some particular working memory mechanisms instrumental in academic performance, analysis using other markers of intelligence—such as the WISC-IV subscales – and other markers of achievement will increase our understanding of the relationships among working memory, intelligence, and achievement even more. Furthermore, future research may include currently ambiguous constructs such as fluid intelligence (as indicated by item scores on the Performance subtests or Ravens Progressive Matrixes, for example). When combined with models of better-understood constructs such as working memory, such work has the potential to shine new light on very complex psychological constructs.

References

Ackerman, P. L., & Beier, M. E. (2005). Knowledge and intelligence. In O. Wilhelm & R. W. Engle (ed.) *Handbook of understanding and measuring intelligence.* London: Sage.

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (ed.) *The psychology of learning and motivation (Volume 2)*. New York City: Academic Press.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

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

Berliner, D. C. (2006). Educational psychology: Searching for essence throughout a century of influence. In P. A. Alexander and P. H. Winne (ed.). *Handbook of educational psychology*. American Psychological Association.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York City: John Wiley and Sons.

Bryant, F. B. & Yarnold, P. R. (1995). Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis. In L. G. Grimm & P. R. Yarnold (ed.) *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory≠attention. *Intelligence, 33,* 251-272.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*, 229-249.

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

Colom, R., Rebollo, I., Palacios, A., Juan-Espinoza, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence, 32,* 277-296.

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

Conway, A. R. A. (2005). Cognitive mechanism underlying intelligence: Defense of a reductionist approach. In O. Wilhelm & R. W. Engle (ed.). *Handbook of understanding and measuring intelligence.* London: Sage.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30,* 163-183.

Conway, A. R. A., Kane, M. J, & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547-552.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory, 4,* 577-590.

Dai, X., & Lynn, R. (2001). Gender differences in intelligence among Chinese children. *Journal of Social Psychology, 134,* 123-125.

Dan, L., Jin, Y., Vandenber, S. G., Yuemei, Z., & Caihong, T. (1990). Report on Shanghai norms for the Chinese translation of the Wechsler Intelligence Scale for Children-Revised. *Psychological Reports, 67,* 531-541.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Experimental Psychology: General, 128,* 309-331.

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

Deary, I. J. (2000). *Looking down on human intelligence.* New York: Oxford University Press.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5,* 164-170.

Deary, I. J., Strand, S., Smith, S., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13-21.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence, 13,* 349-359.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics, 7,* 1-26.

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

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence, 32,* 391-401.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science, 7,* 237-241.

Kelley, T. L. (1928). *Crossroads in the mind of man*. Stanford, CA: Stanford University Press.

Klem, L. (2000). Structural equation modeling. In L. G. Grimm & P. R. Yarnold [Ed.].*Reading and understanding more multivariate statistics*. Washington, DC: American Psychological Association.

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37,* 347-364.

Guildford, J. P., & Dallenback, K. M. (1925*). Determination of memory span by the method of constant stimuli.* Chicago, IL: University of Chicago Press.

Hannary, H. J. (1988). *Experimental techniques in human neuropsychology*. New York City: Oxford University Press.

Heinen, T. (1996). *Latent class and discrete latent trait models*. Thousand Oaks, CA: Sage.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (ed.). *Handbook of understanding and measuring intelligence.* London: Sage.

James, W. (1890). *The principles of psychology*. Boston, MA: Harvard University Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CN: Praeger.

Jensen, A. R. (2003). Regularities in Spearman's law of diminishing returns. *Intelligence, 31,* 95-105.

Jensen, A.R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Amsterdam, the Netherlands: Elsevier.

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

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

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

Kenny, D. A. & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165-172.

Klem, L.(2000). Structural equation modeling. In L. G. Grimm & P. R. Yarnold (ed.) *Reading and understanding more multivariate statistics.* Washington, DC: American Psychological Association.

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

Krumm, S., Schmidt-Atzert, L., Buehner, M., Ziegler, M., Michalczyk, K., & Arrow, K. (2009). Storage and non-storage components of working memory predicting reasoning: A simultaneous examination of a wide range of ability factors. *Intelligence, 37,* 347-364.

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education, 22,* 77-94.

Luo, D., Chen, G., Zen, F., & Murray, B. (2010). Modeling working memory tasks on the item level. *Intelligence*, *38,* 66-82.

Luo, D., Thompson, L. A., & Detterman, D.K. (2003). The causal factor underlying the correlation between psychometic g and scholastic performance. *Intelligence, 31,* 67-83.

Magidson, J. & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (ed.). *The Sage handbook of quantitative methodology for the social sciences.* Thousand Oaks, CA: Sage.

Miyake, A., Friedman, N. P., Rettinger, D.A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130,* 621-640.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (ed.) *New developments and techniques in structural equation modeling.* Mahwah, NJ: Lawrence Erlbaum Associates.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H.(2005). Working memory and intelligence—their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 61-65.

Oberauer, K., Süß, H., Wilhelm, O., & Wittmann, W.W. (2003). The multiple facets of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31,* 167-193.

Peterson , J. (1925). *Early conceptions and tests of intelligence*. Yonkers-on-Hudson, NY: World Book Company.

Rohde, T. E., Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35,* 83-92.

Salthouse, T., Babcock, R., Mitchell, D., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence, 14,* 187-230.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Schweizer, K., & Moosbrugger, H. (2004). Attention and working memory as predictors of intelligence. *Intelligence, 32,* 329-347.

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* Boca Raton, FL: Chapman & Hall/CRC.

Spearman, C. E. (1927). *The abilities of man, their nature and measurement.* New York: Macmillan.

Sternberg, R. J. (2002). *Cognitive psychology*. Boston, MA: Wadsworth.

Süß, H., Oberauer, K., Wittmann, W.W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability – and a little bit more. *Intelligence, 30,* 261-288.

Tillman, C. M., Nyberg, L. & Bohlin, G. (2008). Working memory components and intelligence in children*.* *Intelligence, 36,* 394-402.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. J. A. Hagenaars and A. L. McCutcheon (eds.), *Applied Latent Class Analysis,* 89-106. Cambridge: Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2008a). *Latent GOLD 4.5*. Belmont, MA; Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2008b). *LG-Syntax User's Guide: Manual for Latent Gold 4.5 Syntax Module.* Belmont, MA: Statistical Innovations Inc.

Wechsler, D. (1986). *Wechsler Intelligence Scale for Children-Revised for China.* Changsha, China: Hunan Medical University.