The nature of individual cognitive differences has been debated for centuries. Currently, the construct of intelligence refers to an individual’s ability to use information efficiently and effectively. Ever since Stern advanced the intelligence quotient (IQ) concept in 1912, it has come to represent the estimate of an individual’s overall cognitive ability. However, IQ has proven to be a poor predictor of academic success or failure. Therefore, questions remain as to the nature of intelligence and which cognitive skills predict academic success. To address these questions in this study, structural equation modeling (SEM) was used to evaluate the relationships among variables typically used to estimate IQ, as well as those estimating executive functioning, visual-spatial/motor skills, and phonological processing. It was hypothesized that these variables would be suitable to identify a single, second-order construct, of cognitive processing, or, intelligence (g) such as those purposed by Spearman and Carroll.

For this project, intelligence was estimated via the performance of 79 children, from 58 families using a myriad of cognitive assessment tools. None of the participants exhibited evidence of seizures, head injury, birth trauma, or other disqualifying criteria such as: psychopathology, mental retardation, Autism, etc.. Overall, performance was sufficient to estimate three primary latent factors: Verbal Skill, Nonverbal Skill, and Phonological Processing ($df=17; \chi^2 = 10.13;$ GFI = 0.97; RMSEA = 0.00 [p=0.96]). These three primary factors were also
sufficient to estimate a second-order $g$-factor or single intelligence factor (Factor Loadings: Verbal=.69, Nonverbal=.55, Phonology=.83).

The results of this study were in accordance with our hypothesis that intelligence is comprised of a diverse set of cognitive skills. Our results indicate that phonological processing has as much or more influence on overall cognition ($g$) as traditional measures “intelligence.” Consequently, making statements regarding intelligence may be misleading unless done in the context of a more diverse set of cognitive performance(s). Moreover, these results contraindicate the use of IQ as a contrast to achievement for determining eligibility for academic services. The model of intelligence presented here suggests that the distinction between the constructs of achievement and intelligence (the basis of the discrepancy model) may be artificial.

INDEX WORDS: Structural Equation Modeling, SEM, Intelligence, IQ, Learning disability, Cognition, Individual differences, Academic achievement, Neurolinguistic processing
DEDICATION

This dissertation is dedicated to my friends and family who have been incredibly supportive of me over the years. Without them or their support, this would not have happened.

To my parents, you played a crucial role by inspiring me to chase my dreams and having the guts to keep asking questions. To George Hynd, you inspire, what more can I say! Knut Hagtvet, your time and mentoring both in the US and in Norway have left a profound and permanent mark on me personally and professionally - thank you! Finally, this is dedicated to my wife, truly, without her love and support I never would have made it through the last year it took to get this done. She helped me through many trials and tribulations and I will never forget her kindness or love – thank you!
ACKNOWLEDGEMENTS

I would like to thank all those who have helped to make the completion of this dissertation possible. Truly, for me, this is a dream become reality. From my earliest memories, I recall my parents fostering my inquisitiveness and desire for knowledge. Along my academic path, there have been far too many important people to list here in entirety. However, there are several key people, aside from my parents (mom & George), that I would like to thank.

I would like to extend a personal note of thanks to Nick Santilli, from the John Carroll Psychology Department who taught me that knowledge is a personal pursuit. His mentoring and unwavering support helped set me on this path. Next I must thank everyone who helped me at the Cleveland Clinic. It was there that I began to appreciate the merit of graduate education and the immense benefit that one can have on the population at large through basic and applied science. In particular, my deep thanks are owed to the Section of Neuropsychology. Every member of that section was always supportive, probing, and encouraging. The impact of their collective wisdom and guidance on my development cannot be overstated. Following my time at the Cleveland Clinic, Rainbow Babies and Children’s Hospital helped pave my initial foray into graduate school. Their immense help, guidance and support were needed as I battled my way into graduate school. My heart felt thanks to Jane – who made me realize that there can be more to life than graduate school and who forced me to contemplate this fact.

Many thanks to the Department of Psychology at The University of Memphis. It was there that I not only began my graduate training, but also where I met my wife. My time at U of M taught me that humility and creativity are necessary for scientist and clinician alike, and that self-determination and discipline are crucial for a life in academia. Armed with this information, I completed my training at UGA.

My time at UGA has been like no other. Words fail to capture the essence of my experience. Surrounded by world-class researchers, treated like family and valued as an intellectual, capture, yet fall short of describing my training. Under the tutelage of Dr. George Hynd, my training far exceeded my dreams, which I once considered fantastical. Through artful discourse and didactics, my mentors at UGA, especially Drs. Hynd and Hagtvet, taught me how to approach research with childlike fascination and to consume information as a connoisseur; appreciating the small nuances of a work that marks it as outstanding among similar studies. Art, they say, is in the details.

Finally, I must re-iterate my gratitude to my ever-supportive parents and pay special tribute to my wife – Christina. Over the years, she has been witness to many of my personal and professional triumphs and tribulations. Through them al, and despite them all she agreed to marry me. My quality of life improved the moment she entered my life and I look forward to seeing how good life can get. Thank you Christina – for every thing, your support and love mean everything to me!


### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACKNOWLEDGEMENTS</strong></td>
<td>v</td>
</tr>
<tr>
<td><strong>LIST OF TABLES</strong></td>
<td>vii</td>
</tr>
<tr>
<td><strong>LIST OF FIGURES</strong></td>
<td>viii</td>
</tr>
<tr>
<td>Chapter 1</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>6</td>
</tr>
<tr>
<td>Historical Background</td>
<td>6</td>
</tr>
<tr>
<td>Problems with the discrepancy model</td>
<td>24</td>
</tr>
<tr>
<td>Conceptualizing academic achievement</td>
<td>29</td>
</tr>
<tr>
<td>Spearman’s g revisited</td>
<td>31</td>
</tr>
<tr>
<td>Statement of the problem</td>
<td>32</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>34</td>
</tr>
<tr>
<td>Methods</td>
<td>34</td>
</tr>
<tr>
<td>Statistics</td>
<td>43</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>53</td>
</tr>
<tr>
<td>Results</td>
<td>53</td>
</tr>
<tr>
<td>Summary</td>
<td>63</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>64</td>
</tr>
<tr>
<td>Discussion</td>
<td>64</td>
</tr>
<tr>
<td>References</td>
<td>76</td>
</tr>
<tr>
<td>Appendicies</td>
<td>88</td>
</tr>
<tr>
<td>1 Research on the discrepancy model</td>
<td>88</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Wissler's 1901 results ..................................................................................................... 10
Table 2: Wallin's 1914 survey ...................................................................................................... 15
Table 3: Cognitive Performance ................................................................................................... 54
Table 4: Correlations between indicator variables........................................................................ 54
Table 5: Parameter Estimates........................................................................................................ 56
Table 6: Fit Indices ....................................................................................................................... 59
Table 7: Second-order factor loadings .......................................................................................... 60
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Spearman's model of intelligence</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Historical figures in intelligence</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>Measurement Model of Intelligence</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Structural Model</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>Full Structural and Measurement Model</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>First Alternative model</td>
<td>51</td>
</tr>
<tr>
<td>7</td>
<td>Second Alternative model</td>
<td>52</td>
</tr>
<tr>
<td>8</td>
<td>Model estimates</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>First Alternative model factor loadings</td>
<td>61</td>
</tr>
<tr>
<td>10</td>
<td>Second Alternative model factor loadings</td>
<td>62</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Ancient texts, underscore the importance of passing knowledge from one generation to another (Johnson, Dupuis, Musial, Hall, & Gollnick, 1999). Teachers have always recognized that some students learn better than do others. Understanding why all students are not created equally fueled philosophical debates for millennia (Kamphaus, 2001; Plucker, 2003). With time intelligence emerged as that construct which represents the ability to use information efficiently and effectively. Individuals with high levels of intelligence achieved/dominated and those without did not. Philosophers such as Plato, Galen and Thomas Aquinas theorized about the relationship between intelligence, teaching, and learning (Plucker, 2003). Yet, not until education became compulsory and psychology moved past introspection as a method for analyzing data, did the study of intelligence and achievement begin in earnest.

Galton was the first to scientifically study intelligence. Early studies of intelligence focused on concrete cognitive processes such as sensation, perception, and reaction time (Fagan & Wise, 1994; Kamphaus, 2001; Sattler, 1992). However, findings from those early studies were largely inconsistent and employed inadequate measures of complex cognitive processes. In 1904 Charles Spearman summarized the intelligence literature as atheoretical and in need of more scientific rigor (Spearman, 1904). In his own work Spearman used statistics, which was quite uncommon then (Brody, 2000), and urged others to do the same. Through statistics, Spearman first described the negative influence of error variance. That is, Spearman realized that latent constructs (e.g., intelligence) could not be measured perfectly; thus, every test score is composed of two parts: true score variance and error variance, After he developed a method to remove this influence, Spearman identified a single statistical entity (‘g’ for general ability) that
explained the positive manifold of correlations that remained between cognitive measures after simple computing correlations (Brody, 2000; Carroll, 1993; Matarazzo, 1972; Sattler, 1992; Spearman, 1904; Thorndike & Lohman, 1990) (see figure 1).

Figure 1: Spearman's model of intelligence

Spearman’s work was influential not only in the fields of statistics and psychology, but education as well. For example, as the concept of a “childhood” grew in societal importance, formal education became mandatory. Because compulsory education created a large and diverse student body, there arose a need to identify which children would benefit from regular classroom instruction. The most obvious characteristic by which to sort children was intelligence, yet no objective measure for this purpose was available. After consulting teachers and other psychologists, Spearman’s work influenced Alfred Binet as he developed the first scientifically approved test of intelligence in 1905 (Matarazzo, 1972). Binet’s test assessed the skills needed
for academic success, believing they best estimated an individual’s intellectual repertoire. Yet, his test did not characterize performance as a single factor, which Spearman claimed existed.

In 1912, Stern advanced the intelligence quotient (IQ) concept, operationally defined as the ratio of mental/chronological age. The IQ, which characterized performance from several measures of complex cognitive processes with a single numerical value, finally brought Spearman’s concepts to an applied setting. Terman officially incorporated the IQ in the 1916 revision of the Binet-Simon test of intelligence (French & Hale, 1990). Since then, the IQ has become deeply entrenched in our society's vernacular as being the estimate of an individual’s cognitive ability.

Although the IQ identified individuals with mental retardation accurately, it was less accurate in predicting academic success among the non-mentally retarded. For example, as far back as 1892, researchers identified students with reading difficulties yet without gross language deficits (Dejerine, 1892). Unfortunately, the true spirit behind the concept of ‘intelligence’ never seems to have been truly embraced. By 1917, Hinshelwood indicated the limited worth of the IQ to represent Spearman’s g when he reported that some students had a discrepancy between academic achievement and intelligence estimated by the IQ (Van den Broeck, 2002). Unfortunately, few attended to Hinshelwood’s work and rather than try to place his findings in a broader context of intelligence, the IQ was left to represent an individual’s overall cognitive ability, and areas of cognition outside the purview was attributed to something other than intelligence.

Burt first conceptualized and coined the term ‘underachievement’ in 1950 (Burt, 1950). By 1962, underachieving students with average or better IQs were called learning disabled (Tanner, 2001) and segregated from the regular classroom, effectively enforcing what is now
known as the discrepancy model. Educators labored to provide support services to those segregated under the belief that the IQ truly represented an individual’s overall cognitive repertoire and that those children whose achievement was not commensurate with what was predicted by their IQ scores were indeed Learning Disabled. In 1977 the discrepancy model became embedded in PL 94-142 ("Rules and Regulation Implementing Education for All Handicapped Children Act of 1975," 1977), and surprising as it seems, in the nearly 30 years since there has been little compelling evidence to support its use.

Perhaps the problems with the discrepancy model speak to a missconceptualization of the construct of intelligence and the use of the term IQ. Because many models of intelligence proposed since Spearman’s time appear incomplete, problems with policies as reflected in the use of the discrepancy model were inevitable. A growing literature is revealing that the discrepancy model is inadequate for carrying out Binet’s original charge of identifying children who could and could not learn in the regular education setting (Reynolds, 1990; Stage, Abbott, Jenkins, & Berninger, 2003; Vellutino, 2001), or even who would benefit from remedial services (Proctor & Prevatt, 2003; Reynolds, 1990).

It is in this context then that this study aims to put forth an elementary model of intelligence that demonstrates how intelligence influences all aspects of information processing, from broad verbal skills down to the processing the basic components of language, phonemes. To date, only two other authors have undertaken such a task, Spearman and Carroll; unfortunately, the success of both men are largely overlooked and the theoretical work in many related fields suffers.

The next chapter reviews the literature pertaining to contemporary misconceptions of the IQ, intelligence, and academic achievement (Carroll, 1993; Ceci, 1991; Coleman, 1968; Jencks...
et al., 1972; Schmidt, 1967). A discussion of the problematic nature of the discrepancy model precedes a discussion of Carroll’s model of intelligence derived with exploratory factor analysis. Like Spearman, Carroll found that a single entity influences nearly all facets of cognition. In response to Carroll’s call to validate his work, this chapter will conclude by proposing a testable model of intelligence where a single factor influences both the IQ and cognitive processes underlying academic achievement. Chapter three presents the methods associated with this study while Chapter four presents the results. Chapter five, the Discussion, places the results in the context of the historical literature and in contemporary perspective.
Chapter 2

Historical Background

The purpose of this chapter is to review the evolution of human measurement with a focus on how the assessment of intelligence has centered on the identification of children with learning disabilities. The review begins with developments of historical importance, which lead to issues of assessment today. These are of critical importance in how we go about conceptualizing the nature of human abilities.

Ancient times

Although the exact origins of mental tests are unknown (Anastasi, 1988), there is substantial evidence indicating that tests have been used since ancient times (Kamphaus, 2001). For example, the Chinese were using tests to fill civil service positions about 3000 years ago (DuBois, 1970 as cited in Kamphaus, p. 3). The writings of great philosophers such as Plato, Thomas Aquinas, Pascal, and Aristotle, reveal the centuries of discussion on the nature of mental abilities (Plucker, 2003). Many of those early theorists continue to influence the conceptualization of intelligence today (see Figure 2).

Plato (circa 428-347 BCE) believed that intelligence was a function of age (of the soul). According to Plato, only the soul gained true knowledge. By means of reincarnation, the soul brought this knowledge back to earth. However, the physical manifestation of the soul (i.e., the body) muted the knowledge gained (Plucker, 2003). Intelligence therefore, reflected the number of times the soul had visited earth, more meant greater intellect. Hence, knowledge came not through the senses, but through what the soul perceived and recognized (Zusne, 1957). Apparently Galen (circa 200 BCE) was the first to put forth the idea that intelligent behavior
resulted from mental processes coordinating between cortical and subcortical matter (G. W. Hynd & Willis, 1985). Centuries later, Thomas Aquinas (circa 1225-1275) believed that interactive learning developed intellect.

Note: On the web at: [http://www.indiana.edu/~intell/map.shtml](http://www.indiana.edu/~intell/map.shtml)

**Figure 2: Historical figures in intelligence**
Therefore, teachers should stimulate and engage students to facilitate learning. Further, Aquinas believed that good students also actively engage their teachers to demonstrate comprehension and mastery of the material. Consequently, teachers should remain aware of the student’s cognitive level to engage them appropriately (Plucker, 2003).

Kernels of truth exist in these and other theories of intelligence. Plato correctly noted that cognitive skills rapidly develop during early development. Decades of research support Aquinas’ assertions that students learn best from age appropriate, multimodal approaches. Unfortunately, those and other theories of intelligence would not influence science for centuries.

The 1800’s

Despite centuries of philosophizing about intelligence, the movement to study intelligence emerged during the Renaissance (Kamphaus, 2001). During the Renaissance, society endeavored to develop humane ways to treat individuals with mental retardation. Academics of the time identified individuals with mental retardation with tests of concrete cognitive processes such as vision training, weight discrimination, and believed that training in those skills would cure them (Kamphaus, 2001). Unfortunately, as the field of science was in its infancy, much of that early work went unsubstantiated and languished in obscurity. By the 1800s, the field of science matured enough to credibly investigate the construct of intelligence. Like those before them, early researchers also used concrete cognitive skills to theorize about intelligence (Kamphaus, 2001).

The scientific inquiry about intelligence coincided with the founding of experimental psychology (Anastasi, 1988). Those early experimental psychologists legitimized the field and pioneered developments in the areas of statistics, methodology, and psychometrics. Most intelligence theorists today can trace their pedigrees back to the "father of psychology" Wilhelm
Wundt (1832-1920). In 1879 Wilhelm Wundt opened the first laboratory in the world dedicated to experimental psychology in Leipzig, Germany (Plucker, 2003). The establishment of Wundt's laboratory made discussions about intellect less philosophical and more an empirical topic of study. Wundt's work on sensation and perception (Sattler, 1992), legitimized the scientific study of intelligence and brought others to the field.

Another pioneer of intelligence research was Sir Francis Galton (1822-1911). In addition to being a prolific writer, Galton developed the correlation statistic and later conceptualized the statistical consequences of regression towards the mean (Sattler, 1992). These two developments, though modified over time, are still widely used today. James Cattell (1860-1944) recognized that the methods for data collection were insufficient to make valid interpretations of Galton’s statistics (Kamphaus, 2001). Cattell (who first coined the term "mental tests") understood the need for improved methodology and worked to develop methods to standardize test administration and data collection (Reynolds & Kaufman, 1985).

Galton's work in statistics and Cattell's methodological improvements identified theoretical flaws in previous studies. Research showed that psychophysical measures bore little relationship to intelligent behavior (Kamphaus, 2001; Reynolds & Kaufman, 1985). Wissler (1870 - 1947) confirmed this when he found that undergraduate class standing at Columbia poorly predicted performance on Cattell's psychophysical measures of cognition. Wissler's work in 1901 also revealed that the more concrete the task (e.g., grip strength) the lower it correlated with intelligence (i.e., class standing). Conversely, tasks with a high language component correlated better with intelligence (see Table 1). Because psychophysical measures predicted intelligence so poorly, researchers identified several obstacles hindering progress.
Table 1: Wissler's 1901 results

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strength of hand and class standing</td>
<td>-.08</td>
</tr>
<tr>
<td>Fatigue and class standing</td>
<td>+.23</td>
</tr>
<tr>
<td>Reaction time and class standing</td>
<td>-.02</td>
</tr>
<tr>
<td>Association time and class standing</td>
<td>+.08</td>
</tr>
<tr>
<td>Naming colors and class standing</td>
<td>+.02</td>
</tr>
<tr>
<td>Logical memory and class standing</td>
<td>+.19</td>
</tr>
<tr>
<td>Auditory memory and class standing</td>
<td>+.16</td>
</tr>
<tr>
<td>Logical memory and mathematics</td>
<td>+.11</td>
</tr>
<tr>
<td>Logical memory and Latin</td>
<td>+.22</td>
</tr>
<tr>
<td>Latin and mathematics</td>
<td>+.58</td>
</tr>
<tr>
<td>Latin and rhetoric</td>
<td>+.55</td>
</tr>
<tr>
<td>Rhetoric and mathematics</td>
<td>+.51</td>
</tr>
<tr>
<td>German and mathematics</td>
<td>+.52</td>
</tr>
</tbody>
</table>

Lack of generalization was a primary obstacle to the study of intelligence. Early researchers used very controlled laboratory settings, and frequently failed to replicate their results due to lack of standardized testing procedures. Additionally, researchers often operationally defined the same construct differently. Without an agreed upon definition of intelligence, researchers seemed more concerned about publishing individual findings rather than develop a consistent and coherent model of intelligence (Spearman, 1904). This narrow focus made it difficult for other researchers to generalize or replicate their findings. Furthermore, as researchers often struggled to objectively evaluate the merit of what was reported, the field started to fracture.

**Spearman**

Early in his career Spearman was frustrated by the fractured development, subjectivity, and lack of mathematical objectivity employed by those in the field of experimental psychology as they studied the construct of intelligence (Guilford, 1985; Spearman, 1904). In 1904 Spearman published a critique stating that poor scientific rigor and lack of generalizability did not bode well for the longevity of intelligence research or the field as a whole. Particularly frustrating for him was the lack of statistical data in much of the published work, and the subjective
interpretations researchers made of their data (Brody, 2000). So disparate were the published findings, that Spearman said that one might believe that "there is no such thing as intelligence, but only a number of mental activities perfectly independent of one another except for this common word to designate them" (Spearman, 1904 219-220). Spearman urged that psychology abandon introspection as a method of data analysis to embrace the objectiveness of mathematics and statistics (Spearman, 1904).

Merely using statistics was not enough according to Spearman. Researchers should learn exactly what information the statistics provided, or they risked inferring too much from the results Spearman,(1904). For example, Spearman noted that the Pearson product moment correlation coefficient was vulnerable to "diversity of procedure and ignorance of organic uniformities," (Spearman, 1904 254). Spearman (1904) called upon researchers to remember that: "For having executed our experiment and calculated the correlation, we must then remember that the latter does not represent the mathematical relation between the two sets of real objects compared, but only the two sets of measurements." Taken together, Spearman understood that, left unchecked, different sources of variance obscured the true relationship between variables.

Attempting to understand these true relationships, Spearman first discussed how different sources of error variance influenced statistics. In his work, Spearman discussed at least four sources of error: test item, intra-test, repeated measures, and inter-rater reliability. Error in the test items appeared to artificially inflate or depress the absolute level of ability measured, obscuring the relationship between construct and level of performance. Furthermore, Spearman documented the phenomena where items within a test correlated higher with each other than with items and scores from other tests (Carroll, 1993; Spearman, 1904; Thorndike & Lohman, 1990;
Thurstone, 1938; Thurstone & Thurstone, 1941). This phenomenon meant that correlations between variables would fluctuate depending on the unit of analysis. Spearman’s work revealed error variance resulting from multiple observations, meaning that if precautions did not guard against this variance, the data were easily misinterpreted. Like Cattell, Spearman documented how slight variations in observer bias or methodology influenced statistical procedures. Efforts to model these sources of variance are what led Spearman to discover latent factors.

Through statistics, Spearman developed formulae to estimate the number of constructs assessed by a set of variables. Using the Pearson correlation coefficient as a starting point, Spearman used two statistical procedures to estimate whether one or more constructs influenced test performance. The first formula assumed the measurement of only one construct and should return a value smaller that the Pearson coefficient. If more than one construct was measured, then the returned value should exceed the Pearson coefficient and may easily exceed unity. The second formula assumed the measurement of multiple constructs, emphasized their commonality, and should return a value much smaller than the Pearson coefficient if only one construct was estimated. These formulae allowed Spearman to estimate whether variables share a common element beyond their simple correlations. Frequently, Spearman found that variables positively correlated even after computing the Pearson coefficient.

Spearman theorized that a latent factor existed, which accounted for the consistent "positive manifold of correlations" found after computing simple correlations (Brody, 2000). Thus, while attempting to model sources of variance, Spearman "discovered" latent factors, which would later earn the name "the father of factor analysis" (Guilford, 1985).

Spearman deduced that there are two general sources of variance in data (Brody, 2000; Carroll, 1993). The first is the amount of variance that is wholly unique or “specific” to each
variable. The other, "general" source of variance, represents a common mental function that influences performance on multiple variables (Brody, 2000). This variance indicates the presence of an inferred variable, not directly measured but estimated from the pooled residual variance of multiple variables. For example, when estimating “reading comprehension,” specific variance bespeaks to performance on each test; whereas the general variance refers to the cognitive element (latent factor) that is common to all tests of comprehension. Similarly, Spearman first identified the general intelligence factor which he called "g" (Carroll, 1993; Guilford, 1985; Spearman, 1904). Spearman claimed "... that all branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seen in every case to be wholly different from that in all the others" (Spearman, 1904 p. 284).

Interestingly, Spearman noted that the influence of g varied. That is, "the relative influence of the general to the specific function varies" (Spearman, 1904), depending on the nature of the task. Spearman found that verbally mediated tasks correlated with g more than nonverbal tasks (e.g., weight discrimination). Further, he noticed that the correlations increased with linguistic demand. That is, the tougher the language task, the higher its correlation with g. By definition, this variation implied the presence of a cognitive hierarchy linked to linguistic skill. Thus, according to Spearman, those with more developed language skills were more intelligent. This finding and implication would have immense influence on the future development of intelligence tests and the special education movement. Eventually intelligence tests would determine who could benefit from a regular education classroom, based on intelligence test scores, which were heavily influenced by linguistic adeptness. Furthermore, professionals providing remedial services would spend decades developing remedial
interventions aimed at impacting language functioning. Unfortunately, although Spearman’s theory of a single g factor appears correct, his over-emphasis on language functioning may have hindered the study of intelligence as much as it helped.

**Education in America**

Before Spearman's work influenced the study of intelligence or tests of intelligence, there had to be a need for both. That need arose from education, notably compulsory education. Even though there is evidence of textbooks dating back to 2000 BCE (Johnson et al., 1999), not until relatively recently was compulsory education viewed as important for society. A 19th century cultural revolution changed the way society viewed childhood and education, which placed both on the national agenda (Fagan & Wise, 1994). The belief was that by improving children’s lives through education, society would overcome the moral decay and economic hardships, which resulted from unrestricted immigration, urbanization and industrialization (Fagan & Wise, 1994; Mayer, 1992). These beliefs eventually culminated in federally mandated compulsory education. Compulsory education however, created many challenges for school systems, including, but not limited to providing services to children with communication difficulties, the poor physical health of the student body, and addressing the needs of students with disparate cognitive abilities.

Communication and a common background were among the first challenges faced by the school system. Compulsory education forced large numbers of rural and immigrant children into close contact, especially in the larger cities (Fagan & Wise, 1994), making communication breakdowns common because of language barriers. Furthermore, various ethnic groups held different values and expectations regarding education making it difficult for educational officials to identify and effectively communicate cohesive goals and starting points for curriculum planning to parents and the community at large (Fagan & Wise, 1994). Compounding these
problems was the unprecedented number of children with physical and psychological problems now in the system (Fagan & Wise, 1994).

Early studies of compulsory education efficacy revealed that poor physical health was a fundamental obstacle to academic success (Fagan & Wise, 1994). Physical health problems affected every race, economic, social, and demographic group. A survey by Wallin in 1914 (as cited in Fagan & Wise, 1994, p. 25) revealed that as many as 95% of school age children had at least one physical health issue (see Table 2). Consequently, education officials developed methods to address basic pediatric health care within the school setting. As physical health issues diminished, issues related to how children learn and process information gained attention.

Table 2: Wallin's 1914 survey

<table>
<thead>
<tr>
<th>Type of problems</th>
<th>% of student body afflicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severe dental problems</td>
<td>50-95%</td>
</tr>
<tr>
<td>Poor vision</td>
<td>5-20%</td>
</tr>
<tr>
<td>Problematic adenoids</td>
<td>5-20%</td>
</tr>
<tr>
<td>Nasal obstruction</td>
<td>5-20%</td>
</tr>
<tr>
<td>Enlarged or diseased tonsils</td>
<td>5-15%</td>
</tr>
<tr>
<td>Curvature of the spine</td>
<td>2-7%</td>
</tr>
<tr>
<td>Malnutrition</td>
<td>1-6%</td>
</tr>
<tr>
<td>Weak or tubercular lungs</td>
<td>1-2%</td>
</tr>
<tr>
<td>Hearing problems</td>
<td>1-2%</td>
</tr>
</tbody>
</table>

A student body with diverse cognitive abilities inhibited the ‘corporate’ model of efficiency schools emulated (Nitko & Lane, 1990). Children who otherwise appeared normal yet failed academically were an obstacle to this model. The push for success often lead educators to help children who struggled academically with unproven interventions (Fagan & Wise, 1994), oftentimes achieving little, if any success. Compulsory education combined with a need for teaching efficiency provided fertile ground for the development of a philosophy of pupil
segregation (Hearne & Stone, 1995; Johnson et al., 1999). This philosophy was clearly articulated by Hall in 1911 when he said "habits of stupidity and inertness are often more contagious than are examples are the best workers. This is why the elimination of the stupids is so urgent..." (cited in Fagan & Wise, 1994, p. 27). Contemporary intelligence tests were developed to address this need (Mayer, 1992). Alfred Binet pioneered this movement with his test to identify children with mental retardation (Thorndike & Lohman, 1990).

**Binet**

The Minister of Public Instruction asked Binet to identify children at risk for academic failure due to mental retardation (French & Hale, 1990), marking the first time intelligence research had social, political, and scientific status. To achieve this task, Binet consulted with teachers to better characterize the cognitive skills needed for school success. Binet recognized the inadequacy of psychophysical measures and eventually focused his attention on higher mental processes such as abstraction and reasoning (Brody, 2000; Sattler, 1992; Spearman, 1904). Drawing upon the statistical advances of Spearman and others, Binet, offered one of the most important contributions to psychometrics by modeling measurement error at the test item level in his test (Reynolds & Kaufman, 1985). After he developed a framework of what and how to measure cognitive skills representing intelligence, Binet sought an appropriate normative group on which to build his test.

Binet believed that age groups provided sufficient referent groups for intelligence score comparisons (Matarazzo, 1972). The original Binet test, published in 1905, identified children with mental retardation by making chronological age cohort comparisons. The rationale for these comparisons was based on the belief that children with mental retardation would perform at a cognitive level considered normal for younger children (Brody, 2000), but far below that of their
same age peers. Three years later, the 1908 revision included age-grade equivalent tables for more extensive cohort comparisons. More test items were added in the 1911 revision, which allowed examiners to estimate an individual’s mental age for still more elaborate cohort comparisons (Brody, 2000). However, making these comparisons were cumbersome and the final score did not translate well between different age cohorts.

Stern proposed the Intelligence Quotient (IQ) concept in 1912 (Sattler, 1992; Stern, 1912) to ease performance comparisons across several age ranges. The IQ concept also addressed the need to account for cognitive development (Carroll, 1993; Stern, 1912), which was not included during the initial test construction. Stern’s IQ concept was a ratio of a child's mental age divided by his/her chronological age (Brody, 2000; Sattler, 1992). This ratio helped establish a relative intelligence ranking system for each age group. The IQ concept, with a slight modification was included in Terman’s 1916 revision of the Binet test. Terman took Stern’s IQ ratio and multiplied it by 100 and dropped the decimals, which greatly simplified the comparison of scores (French & Hale, 1990). Eventually the IQ represented an individual’s overall cognitive skill by proxy of academic success, which likely added to the popularity of intelligence tests.

So successful was the Binet-Simon in France, that within a few years of publication Goddard began using it at the New Jersey Training School for Feebleminded Boys and Girls (French & Hale, 1990). By 1913, the Binet-Simon had been completely translated into English (French & Hale, 1990; Reynolds & Kaufman, 1985), and began its tradition in the American educational system. The 1916 Stanford-Binet was considered the most popular and widely used test of intelligence in this country until David Wechsler published his adult scale of intelligence in 1939 and his child version 10 years later (Thorndike & Lohman, 1990).
Despite the acclaim of the Binet tests, they were not without psychometric flaws, nor were they suitable for use with adults. As an army psychologist during World War I, Wechsler argued against extrapolating performance on the Binet to adults because of normative, and statistical restrictions (Matarazzo, 1972). Further, Wechsler criticized other measures of intelligence for their focus on speedy performance (Kuder, 1986), a concept debunked by Wissler in 1901 (Brody, 2000; Reynolds & Kaufman, 1985). When he returned to his civilian life, Wechsler worked to fill the clinical need for a better way to estimate intelligence (Kuder, 1986).

From a psychometric standpoint, Wechsler disagreed with the test administration protocol of the Binet. For example, not everyone completed the same subtests on the Binet tests. Children of the same age could complete different sets of subtests; one child may complete a series of simple tasks, taken by young children, while another could complete a set of difficult tasks commensurate to those taken by older children. This flexibility troubled Wechsler for two reasons. First, if same age children complete different tasks, inferences about intelligence could be misleading. Second, although Wechsler recognized that same age peers made a good comparison group, he was troubled by the fact that the IQ was relative to age each group independently (Matarazzo, 1972; Thorndike & Lohman, 1990). Wechsler’s finding that cognition followed a curvilinear path exacerbated this problem. Intellectual development rises sharply early in life, slows, and eventually decreases later in life (Matarazzo, 1972). Thus, the Binet tests, which assumed a simple linear function of cognitive development, as evidenced by the mental age/chronicle age ratio, became untenable (Matarazzo, 1972; Thorndike & Lohman, 1990).
Wechsler disliked several aspects of the Mental Age (MA) concept (Matarazzo, 1972). To begin, he believed that life experiences influenced the development of MA, confounding its comparison to chronological age (Thorndike & Lohman, 1990). Further, as chronological age was the expected mental age (Thorndike & Lohman, 1990) this comparison had the potential to make less of the statement about cognitive abilities, and more about their home environment. Finally, Wechsler did not believe that the concept of mental age translated well to the adult population (Kuder, 1986; Thorndike & Lohman, 1990).

Wechsler also found the method of deriving the IQ lacking (Wechsler, 1939) because cognitive skills did not develop in a linear fashion. Towards this end, Wechsler developed a logistic function of intellectual development and fit it to a normal curve (Thorndike & Lohman, 1990; Wechsler, 1939). This statistical manipulation allowed him to convert a raw score into an intelligence score ranked relative to same age peers, yet absolute in terms of position within the entire population. Wechsler accomplished this statistically by setting the overall mean score to 100 (out of convention), and set the probable error of the test to 10, which was 2/3rds of his sample standard deviation. Thus, the middle 50% of subjects fell between the scores of 90-110. Wechsler’s improvement became known as the deviation IQ and is still used today.

The culmination of these criticisms and improvements led to the introduction of the Wechsler-Bellevue test of intelligence (Fagan & Wise, 1994; Matarazzo, 1972; Sattler, 1992; Thorndike & Lohman, 1990). The Wechsler tests used norming samples closely matched to US census data unlike previous test authors. In 1955 the Wechsler-Bellevue was replaced by the Wechsler Adult Intelligence Scale (WAIS), and then in 1981, the WAIS was updated and revised to become the WAIS-R (Matarazzo, 1972). In 1949, Wechsler introduced his test of child intelligence, the Wechsler Intelligence Scale for Children (WISC). By 1974, the revised WISC,
the WISC-R, was more widely used than the Stanford-Binet (Thorndike & Lohman, 1990). To date, no other measure of intelligence, has gained the widespread use or acceptance as the Wechsler scales. No credible contenders have emerged to challenge the Wechsler scales and there have been ‘relatively few important innovations in ability measurement… since 1939’ (Thorndike & Lohman, 1990).

Although great advances had clearly taken place in regards to the measurement of cognitive abilities, the need to standardize the assessment of academic achievement became clear. Hence, efforts to assess and estimate academic achievement in a manner similar to intelligence became common.

**Achievement tests**

Compulsory education highlighted the problems of how to educate a diverse student body. Long known, yet little discussed was the arbitrary nature of grades (Johnson et al., 1999). It was also recognized that students with similar levels of knowledge may or may not pass a class depending on the criteria set by the instructor. Problems such as this were exacerbated by the mobility of society and the lack of consistent educational standards throughout the country (Johnson et al., 1999), all of which primed the education system for the achievement test movement.

Near the turn-of-the-century, several conditions sparked the achievement test revolution (French & Hale, ; Nitko & Lane). Because there were students who struggled academically, the need for a more refined method of student identification emerged. By 1918, there were 84 elementary and 25 high school standardized tests that were officially recognized and used (Nitko & Lane, 1990). So powerful was the need for standardized achievement tests, that by the time the WAIS was introduced in 1939, there were approximately 4200 test and rating scales endorsed for use in the school system (Nitko & Lane, 1990).
Standardized measures of academic achievement met several needs of the school system. Primarily, achievement tests provided a more refined method of identifying children at risk for academic failure. The net result of this ability helped to provide a measure of educational efficiency, something required by the ‘corporate’ mentality of the school system (Johnson et al., 1999; Nitko & Lane, 1990). Test results were useful for curriculum planning and provided a means of comparing one school system to another (Johnson et al., 1999; Nitko & Lane). Unfortunately, the importance of the information provided by academic achievement tests was often underestimated.

Because tests of intelligence screened out children with mental retardation, achievement tests were thought to provide information about cognition unrelated to intelligence. That belief guided curriculum planning, intervention implementation, and laws governing remedial services. The conceptual dichotomy between intelligence and academic achievement now seems illogical, especially given the origin of tests of intelligence. However, the negative impact of that belief was very real for the education system. Today we understand that achievement tests may provide robust, complementary information about overall cognitive processing, than traditional intelligence tests alone. Contemporary multidisciplinary research reveals how achievement tests provide information concerning the relative contribution of discrete cognitive processes to the whole of cognition (i.e., intelligence). When a discrete cognitive skill does not develop to expectations, there are suggestions from the neuroimaging literature that multiple cognitive systems will be affected. Thus, there is the potential that deficits in one domain of cognition will likely produce global effects. Not understanding this principle lead to the implementation of concepts such as the discrepancy model. According to this theory, only children who demonstrate achievement scores significantly below their IQ would benefit from formal
academic assistance. The next sections will review the development of the discrepancy model and some of its more prominent flaws.

**The discrepancy model**

The concept that there are students who struggle academically despite appearing normal is just over 100 years old. Dejerine (1892) was among the first to formally recognize that individuals could have reading problems without general language impairment. Extending Dejerine's work (Van den Broeck, 2002), Hinshelwood first reported on individuals with discrepant IQ and achievement scores (Hinshelwood, 1917). Hinshelwood believed that these discrepancies represented the effects of subtle cortical abnormalities gone unassessed by intelligence tests (Hammill, 1993; Wiederholt, 1974). In 1950, Burt introduced the concept of underachievement, defined by the discrepancy between educational attainment and estimated intelligence (Burt, 1950). In 1963 the Association of Children and Adults with Learning Disabilities was formed and underachievement officially became ‘Learning Disability’ (Tanner, 2001) (Kamphaus, 2001; Reynolds, 1992; Sattler, 1992; Tanner, 2001) (Frankenberger & Harper, 1987; Reynolds, 1992). Learning disabled children received academic assistance because they were at risk for academic failure. However, there was a need to develop standardized eligibility criteria.

The discord in the field of Learning Disabilities peaked in the 1970’s. Locally governed education systems used inconsistent eligibility criteria and service delivery methods. There was a need for a national set of guidelines to provide educational equity throughout the country. During the mid-1970s, the Office of Education collaborated with the top researchers in the field of education to solve this problem. The result of this collaboration was the production of Public Law (PL) 94-142 (Education for All Handicapped Children Act) in 1975 (National Joint Committee on Learning Disabilities, 1994). In 1977 the rules and regulations governing PL 94-
142 were codified and a national criteria for a Learning Disabilities was set ("Rules and Regulation Implementing Education for All Handicapped Children Act of 1975," 1977).

Overall, PL 94-142 vaguely outlined the identification and intervention procedures each school system should employ on a case-by-case basis. PL 94-142 used two broad tenants for the diagnosis of a Learning Disability (Reynolds, 1990). The first stipulated that a child with a learning disability must fail to reach an academic level commensurate with his or her age despite adequate instruction. Secondly, a child with a learning disability must present with the severe discrepancy between their estimated intelligence (i.e., IQ) and academic achievement. As a group, a multidisciplinary team must ensure that the academic failure was not due to mental retardation, sensory deficit (e.g., vision or hearing loss) or the result of some social, emotional, environmental, economic or cultural circumstance is (National Joint Committee on Learning Disabilities, 1994), using standardized instruments. Thus, the 1977 rules and regulations officially codified the use of the discrepancy model.

Despite the establishment of national criteria for diagnosing a Learning Disability, shortcomings of the discrepancy model were soon apparent. Primarily, problems with implementing PL 94-142 arose. Locally governed education meant that each state could determine its own method for implementing the law. This freedom hints at the problematic nature of the discrepancy model. Freedom of implementation reflected a conceptual ambiguity of the law. Ambiguity in the law followed conceptual ambiguity of the relationship between intelligence and academic achievement. Not understanding the nature of information provided by intelligence and achievement tests by policy-makers made problems with the discrepancy model inevitable.
Problems with the discrepancy model

In the 30 years since it was enacted, few studies have supported the ability of the discrepancy model to identify a Learning Disability (LD) or determine eligibility for academic assistance. Appendix 1 summarizes some of the discrepancy model literature related to the diagnosis of LD. As many as 70% of experts agreed that the IQ – achievement discrepancy did not belong in the LD definition (Speece & Shekitka, 2002). The President’s Commission on Excellence in Special Education concluded that the methods for LD identification were insufficient and “antiquated” and called for suggestions before congress reviews special education legislation (“Editorial: Study analyzes proposals to change LD identification," 2003). Despite the overwhelming consensus to stop using the discrepancy model, a united stand for this rationale did not exist (Speece & Shekitka, 2002). As mentioned earlier, problems with the discrepancy model were inevitable, and led to a sense of ‘chaos’ in the LD literature (Siegel, 2003). This chaos apparently stemmed from issues related to the: 1) definition, 2) application, and 3) conceptualization of the discrepancy model.

Problems defining a discrepancy

The organizational structure of education in the US almost guaranteed that ambiguity in law would prove problematic. That is, although the federal government provided direction to the field of education, federal laws were interpreted and enacted at the state and local level (Johnson et al., 1999). Thus, nearly all states interpreted PL 94-142 differently (Peterson & Shinn, 2002), and even when states developed similar LD definitions, unique variations existed (Proctor & Prevatt, 2003). For example, even though 71% of the states followed the same federal guidelines, the role of intelligence, the specific definition of academic failure, and unique exclusionary criteria existed across the country (Mercer, Jordan, Allsopp, & Mercer, 1996).
Because states developed independent LD definitions, (Peterson & Shinn, 2002), a child might have qualified as learning disabled in one state but not in another (Peterson & Shinn, 2002; Proctor & Prevatt, 2003). Consequently there were children who struggled academically with the same levels of reading achievement, yet received help only if they meet the state’s unique criteria (O'Malley, Francis, Foorman, Fletcher, & Swank, 2002). Hence, educators called LD definitions inconsistent, unpredictable, and without theoretical consensus (Peterson & Shinn, 2002; Proctor & Prevatt, 2003).

**Problems applying a discrepancy model**

Even if a universal discrepancy-based definition of LD emerged there would be problems with its application. A literature review found that meeting diagnostic criteria varied depending on the method and tests used. There were at least two popular methods to determine an IQ-achievement discrepancy, and their underlying assumptions made applying even a unified definition difficult. The Simple Discrepancy Model (SDM) and the Regression Discrepancy Model (RDM) were the two most common ways the definition of a learning disability was applied. Although the SDM and RDM shared the same goal, identify children with a learning disability (Peterson & Shinn, 2002), their underlying assumptions often led to differential diagnoses. This section provides a brief description of the primary assumptions and outcomes of each method.

Methodologically, the SDM is the most straightforward method for applying the discrepancy definition of LD. In its most basic form, a student has a learning disability if the value of IQ minus an achievement test score exceeds the state cutoff score (in the absence of exclusionary criteria). Typically, the minimum cutoff score is 15 standard score points, or a standard deviation difference between IQ and achievement. That it is possible to estimate intelligence and achievement in isolation, that no reciprocity of influence exists between the
constructs, and that one should predict the other are the underlying assumptions of this method. This method’s simplicity greatly contributes to its popularity.

The RDM is less straightforward. Unlike the SDM, which uses one calculation, the RDM employs up to six, each one using a different equation (Evans, 1992). However, only the first calculation is crucial, as it determines whether the difference between the IQ and achievement scores is significant, if not, there was no need to proceed (Evans, 1990; 1992). Regression towards the mean is the underlying assumption of RDM (Van den Broeck, 2002). That is, when variables do not correlate perfectly, subjects who score high on one variable tend to score high on other, but to a lesser degree; the same holds for subjects with lower scores.

Regardless of the exact definition of a learning disability, both application methods received criticism. The SDM seemed too flexible, simplistic, and inaccurate for identifying a learning disability. Research focusing on the RDM revealed issues relating to validity and bias that required attention (Share & Silva, 2003; Van den Broeck, 2002). Hence, even if the states used the same definition, a child might be called learning disabled in one state and not in another depending on the application method. Research comparing these methods demonstrated some of the problems with applying a discrepancy definition of a learning disability.

Studies of model efficacy, found that the SDM and RDM identified children previously diagnosed with a learning disability in excess of 60% (Peterson & Shinn, 2002; Proctor & Prevatt, 2003). However, neither outperformed the other. Peterson’s RDM study of 48 fourth-grade children reached 67% accuracy, while Proctor’s SDM study of 170 university students was 70% accurate. Clearly neither claimed superiority. Proctor, noted that even though the SDM obtained the highest identification rate, the children identified by other discrepancy models were not always a subset of the SDM group. That is, different models identified different groups of
children as learning disabled (Proctor & Prevatt, 2003). Research on minority enrollment in LD programs supported these findings. After reviewing 6036 cases, Colarusso found that while no application method balanced the racial makeup of learning disabilities programs, different groups of students had a learning disability depending on the method used (Colarusso, Keel, & Dangel, 2001).

Putting the methods aside, research indicated that the desire to help struggling children influenced test choice (Ross, 1995), which may bias eligibility determinations (Peterson & Shinn, 2002; Ward, Ward, Glutting, & Hatt, 1999). For example, Ward found that 70% of 201 children previously diagnosed with a learning disability failed to present with a non-normal cognitive profile (Ward et al., 1999). Because so many children did not meet even discrepancy criteria Ward (1999) suggests the use of some additional criteria to establish eligibility (e.g., history of failure, professional judgment, etc.). Clearly, if the diagnosis of a learning disability depended on the way the definition was applied, the conceptualization of a learning disability was problematic.

**Problems conceptualizing a discrepancy**

Using an IQ-achievement discrepancy to characterize a learning disability indicates problems with conceptualizing a learning disability. In general, the learning disabilities literature appears to wrestle with three conceptual issues. The first is whether children who struggle academically are distinctly different from ‘normal’ children. Secondly, debates continue over the uniqueness of the intelligence and achievement constructs. Finally, the amount and direction of influence between intelligence and achievement remains unknown. There are conflicting reports as to whether intelligence influences achievement or vise versa. This section briefly reviews research related to these issues.
Several studies compared children with learning disabilities to low achieving readers. Seldom are significant differences between the two groups reported (D'Angiulli & Siegel, 2003; Kelly, 1998; O'Malley et al., 2002; Remschmidt, Hennighausen, Schulte-Korne, Deimel, & Warnke, 1999). Further, up to 65% of children diagnosed as learning disabled did not have unique cognitive profiles (D'Angiulli & Siegel, 2003). Predictors of reading achievement were similar for children who met discrepancy criteria and those who did not (Kelly, 1998; O'Malley et al., 2002; Tiu, Thompson, & Lewis, 2003). Further, a severe discrepancy between estimated intelligence and academic achievement was not sensitive to or predictive of response to early intervention (Stage et al., 2003).

The success of the discrepancy model rested on the presumed ability to predict academic success from estimated intelligence (Burt, 1950; Dejerine, 1892; Hinshelwood, 1917). Unfortunately, this claim seemed tenuous. Some found that the correlations between basic reading and intelligence were modest at best (Vellutino, 2001); while others reported that no relationship between intelligence and learning/achievement existed (Diseth, 2002; Kelly, 1998; Kershner, 1990; Willson & Reynolds, 2002). Still others have suggested that variables other than measures of intelligence or achievement were more predictive of academic success (Murray & Wren, 2003). For example, it was reported that low self-esteem may hinder achievement (Kershner, 1990), while success promoted further achievement (DiPerna & Volpe, 2001).

The research on the link between intelligence and academic achievement (Dandy & Nettelbeck, 2002; Horn, 1985; Robert J. Sternberg, Castejon, Prieto, Hautamaeki, & Grigorenko, 2001), is not conclusive (Ceci, 1991; Siegel, 1999). Research on different aspects of intelligence (e.g., verbal and performance) have exclusively predicted achievement (Dandy & Nettelbeck, 2002; Majoribanks, 1976; Naglieri, 2001; Tiu et al., 2003). For example, the Performance IQ of
the WISC-III predicted reading comprehension (Tiu et al., 2003). Researchers also found that intelligence uniquely influenced phonological processing (McBride-Chang, 1995), reading (Tiu et al., 2003), and vocabulary development, especially for words of Latin origin (Stage et al., 2003; Ullstadius, Gustafsson, & Carlstedt, 2002). There is even evidence that schooling influenced the development of intelligence (Ceci, 1991).

Several researchers reported a reciprocal relationship between intelligence and achievement (Stage et al., 2003; Stanovich, Cunningham, & Feeman, 1984), calling into question the prediction of one from the other. Other factors such as environment (Dickens & Flynn, 2002) and ethnicity (Dandy & Nettelbeck, 2002) reportedly influenced intelligence and achievement scores. Obviously there was much confusion regarding the conceptualization of a learning disability. The impact of that confusion meant that problems with any discrepancy model inevitable. Taken as a whole, problems with defining, applying, and conceptualizing a learning disability in terms of an IQ-achievement discrepancy apparently resulted from losing sight of the way intelligence was originally conceptualized.

**Conceptualizing academic achievement**

As noted above, conceptualizing learning disabilities as a discrepancy between intelligence and achievement was problematic. Not understanding the nature of information that each type of test provided exacerbated this problem. For example, intelligence tests purportedly estimate general cognitive abilities while achievement tests supposedly measure the attainment of a particular skill (Kamphaus, 2001; Sattler, 1992). Yet, the literature failed to fully support those claims (Carroll, 1993), and there has been no consensus as to exactly what information achievement tests provide. The literature suggested that achievement tests are broadly conceptualized in three ways: 1) separate 2) separate but coequal, and 3) the same. The lack of consensus may have
greatly contributed to the mounting criticism of the discrepancy model. The next section briefly characterizes each of the three viewpoints.

**IQ and Achievement are Separate**

The first viewpoint was that intelligence and achievement are separate constructs. An IQ score represented overall cognitive processing abilities, while an achievement score represented the mastery of discrete cognitive skills. Said differently, achievement scores represented subcomponents of the whole of cognition (i.e., intelligence), which mandated a hierarchy between intelligence and achievement. This dichotomy between intelligence and achievement gained support from the field of neuropsychology during the 1970’s and 80’s. Popular then were theories of structural/functional localization. That is, discrete cognitive functions were “housed” in particular brain locations. Further, it was considered possible that discrete brain areas could be defunct without necessarily affecting the whole of cognition. However, subsequent research has not supported localization theories, or the discrepancy model’s utility.

**IQ and Achievement are Separate but coequal**

Some conceptualized achievement skills as different but “coequal” with intellectual abilities (Kamphaus, 2001). Kamphaus did not believe that intelligence determined achievement levels (Kamphaus, 2001), and the relationship between intelligence and academic achievement was unclear. Presumably, intelligence and achievement tests measured different aspects of cognition and the former would not predict the latter. However, the logic behind this line of thought was not clear. If intelligence and achievement tests provided equal information, did they both estimate intelligence, or, perhaps achievement? To date, the term coequal remains enigmatic.
Others believed that achievement tests provided the same information as intelligence tests, that the two tests were interchangeable as they estimated nearly the exact same thing. This position was bolstered by reports that both types of tests used similar tasks and the scores were generally highly correlated. The development of the Kaufmann Achievement Battery for Children (KABC) exemplified the perceived interchangeability of these tests. Essentially, the KABC estimated academic achievement using measures of verbal intelligence from other tests (Kamphaus, 2001). Furthermore, the Peabody Picture Vocabulary Test- Third Edition (PPVT-III) (Dunn & Dunn, 1997), a test of receptive language, commonly served as an estimator of intelligence (Kamphaus, 2001). Because this conceptualization presumed intelligence influences all of cognition, it fit best with the original theories of intelligence. However, it does not fit well within the discrepancy model framework, which treated intelligence and the cognitive processes related to learning as fundamentally different things.

Clearly, the different conceptualizations of achievement test scores precluded the usefulness of a learning disabilities definition based on an IQ-achievement discrepancy. The conceptual overlap of intelligence and achievement indicated by the second and third viewpoints mentioned above, indicate that, perhaps the construct of intelligence has become too narrow since the time of Spearman.

**Spearman’s g revisited**

The previous discussion revealed how tests of intelligence and achievement apparently provide overlapping or redundant information. This overlap helps explain why “the distinction between intelligence and achievement tests has never been empirically demonstrated” (Kamphaus, 2001).
The second and third positions above indicate that the contemporary IQ provides only a provisional impression of intelligence based on how well the brain synthesizes broad-spectrum information. The nebulous nature of the IQ provides little information regarding the encoding, synthesis, and use of discrete information, limiting its worth for estimating intelligence. It might be argued that by combining the information provided by the two types of tests, a realistic approximation of intelligence may begin to emerge. Hence, using an IQ-achievement discrepancy to define a learning disability is problematic as both estimate different aspects of the same construct, intelligence.

Ironically, combining IQ and achievement information to estimate intelligence is similar to Spearman’s $g$ theory from a century ago. An exhaustive meta-analysis on cognition published in 1993 supports the single factor theory of intelligence (Carroll, 1993). Using exploratory factor analysis, Carroll found that the combined measures of intelligence and achievement best estimated the intelligence construct (Carroll, 1993). Carroll's final model characterized cognition with three strata of latent factors. Stratum I contained factors representing fairly discrete aspects of cognition (e.g., visual memory, sound discrimination, word fluency, and reaction time) (Carroll, 1993). The second-order factors of Stratum II characterized Stratum I factors more generally (e.g., fluid intelligence, general memory, and processing speed). A single third-order factor labeled general intelligence comprised Stratum III. Despite the extensive nature of this work, Carroll concluded with a call for replication using confirmatory factor analysis.

**Statement of the problem**

Although widely acclaimed, a potential weakness of Carroll’s work was the use of exploratory factor analyses. His model of intelligence resulted from a data-driven approach, devoid of theory
(Carroll, 1993). Consequently, the goal of this study is to test the theory that data from psychometrically sound intelligence and achievement tests equally contribute to the existence of a general intelligence factor. Because Structural Equation Modeling (SEM) is a dynamic way of looking at multiple constructs simultaneously with confirmatory factor analysis, it is well suited for testing this hypothesis. Should such a model prove viable, it would support g theories of intelligence such as Carroll’s. Furthermore, the viability of such a model may indicate the artificial division of the achievement and intelligence constructs as indicated by the discrepancy model. The following chapter provides a discussion of the methodology employed to examine the validity of Carroll’s model. The purposed conceptual model put forth here will be tested using SEM.
Chapter 3

Methods

Participants

The participants for this study were children seen at the Center for Clinical and Developmental Neuropsychology (CCDN) as part of a National Institutes of Health (NIH) funded grant (#HD26890-06) investigating familial dyslexia. Data collection for the grant began fall 1999, and was completed in the spring of 2003. Parents and teachers of the children typically seen at the CCDN expressed concerns about learning or behavioral problems. Grant inclusion criteria ensured that all target children had a previous diagnosis of a reading disability or demonstrated significantly impaired reading achievement accompanied by normal IQ scores. None of the target children or their siblings exhibited evidence of seizures, head injury, birth trauma, or other disqualifying criteria such as: psychopathology, mental retardation, Autism, etc..

Recruitment for the grant utilized local school districts, tutoring centers, and community organizations. Additionally, local educational administrators and school psychologists received letters and/or phone calls offering assessment services to students with reading disabilities. The Georgia Association of School Psychologists (GASP) provided a list of school psychologists working in the northern part of Georgia for a mass mailing and flyer campaign explaining the referral criteria. Similar flyers and letters of introduction were sent to local community organizations in the Athens area. Furthermore, campus groups such as the University of Georgia, Center for Family Research, assisted in publicizing the project.

Parents typically called the CCDN requesting information about the services provided. At that time, additional information was gathered from them to assure that children met the entry
criteria for participation. In the event that children did not qualify for the study, parents were referred to an appropriate assessment or intervention resource in their area or at the University of Georgia. Following the intake interview by phone, assessment appointments were scheduled. Then, a packet of materials was mailed. The packet included materials to be completed prior to the assessment by the parents and the classroom teacher.

**Measures**

It was the goal of this study to characterize the relationship between measures of intellect and academic achievement with a general intelligence factor. To accomplish this, attention was focused on four well-established and normed measures of cognitive and academic functioning. Specifically, this study assessed how well a general intelligence factor accounted for the common variance of subtests selected from the Wechsler Abbreviated Scale of Intelligence (WASI), the Wisconsin Card Sorting Test (WCST), the Test of Visual Motor Integration (VMI), and the Comprehensive Test of Phonological Processing (CTOPP). Following is a brief review of these measures.

**Wechsler Abbreviated Scale of Intelligence (WASI)**

The Wechsler scales are one of the most frequently used measures in neuropsychological batteries (Spreen & Strauss, 1998). The Wechsler scales are instruments that provide information about the overall level of intellectual functioning, and presence or absence of cognitive deficits (Spreen & Strauss, 1998). For this study, the Wechsler Abbreviated Scale of Intelligence (WASI) was chosen because it provided a short and reliable measure of intelligence (Zhu, 1999). The construction of the WASI permitted its use with individuals from age 6 up to 89 (Zhu, 1999). The WASI contained four subtests commonly seen in other Wechsler scales: Vocabulary,
Similarities, Block Design, and Matrices. In general, those subtests earned high factor loadings on a factor that represented general intelligence (Zhu, 1999) making it appropriate for this study.

The Vocabulary test assessed an individual’s ability to orally define words. It provided a good estimate of expressive vocabulary, fund of knowledge, crystallized intelligence, and general intelligence (Sattler, 1992). The similarities task required the individual to explain how items were conceptually alike; this task provided an estimate of concept formation, abstract reasoning, as well as general intelligence (Zhu, 1999). Taken together, these two tests provided a reasonable estimate of Verbal intelligence (VIQ). Numerous studies have provided strong statistical support and rationale for this abbreviated estimation of VIQ (Sattler, 1992; Zhu, 1999).

The Bock Design task assessed an individual’s ability to reconstruct complex geometric designs using multi-colored blocks. This task estimated visual spatial and visual motor skills as well as abstraction and general intelligence (Zhu, 1999). The Matrices task required individuals to identify a target picture that would complete a larger picture with a portion missing. This task purportedly estimated nonverbal fluid reasoning and general intelligence (Zhu, 1999). Taken together, these two tests offered a good estimate of Performance Intelligence (PIQ). As noted above, numerous studies provided strong statistical support and rationale for this abbreviated estimation of PIQ (Kamphaus, 2001; Sattler, 1992; Zhu, 1999).

Psychometrically, the WASI was considered a strong, appropriate tool to measure cognitive functioning. The test was normed on 2220 individuals, with the performance of about 20 people to represent each age group (Zhu, 1999). The stratified norming sample reflected the diversity of the US population (see source for details), which increased generalizability. Test-retest stability coefficients for the child sample, corrected for restriction of range, were between .77 and .93 (Zhu, 1999), well above the mark needed for clinical utility. At the subtest level, the
reliability coefficients of the child sample ranged from .86 to .93, well above the conventional .70 thresholds needed to use the instrument clinically. The reliability coefficients for the IQ scores were generally higher than the subtest coefficients (ranging from .93 to .96). This disparity, though small, was explained by the fact each subtest represents only a portion of the total IQ score (Zhu, 1999).

The standard error of measurement ($SE_m$) coefficients indicate that there is a small amount of measurement error in any given individual raw score (Zhu, 1999). $SE_m$ inversely relates to reliability. That is because the reliability coefficients were generally high, the $SE_m$ were fairly small, ranging from 2.94 to 3.99 in the children’s sample (Zhu, 1999). Further, the $SE_m$ allows the computation of confidence intervals about an individual’s raw score. That is, if the $SE_m$ were known one could construct a range of numbers where the person’s “true” score would fall with infinite measurements. Like other tests, the WASI reports IQ scores in terms of standard scores, with a mean of 100 and a standard deviation of 15 points.

Wisconsin Card sorting Test (WCST)

The WCST was originally developed to assess normal cognitive functioning (Harris, Heaton, Tuttle, Coutu, & Schniker, 1990). To complete the task, 4 stimulus cards were first arranged in front of the subject, each card with a different number and color of shape(s) on the cards. Next, the subject organized 128 response cards to match the 4 stimulus cards placed in front of them. The subject was not told what rule to use when sorting the response cards, and after each card response card was placed, the subject was told only whether their placement was correct. Once the subject deduced the correct rule by which to sort the response cards, exemplified by the correct placement of 10 consecutive cards, the examiner changed the rule
unbeknownst to the subject. This pattern continued until the subject successfully completed 5-
sets of rule changes, or ran out of response cards (whichever came first). The WCST was scored
along several dimensions including: number of rule sets completed, number or misplaced cards
(errors), perseveration (e.g., using a rule that previously was successful, but no longer
appropriate), set consistency (e.g., recognizing when a rule was successful), number of response
cards needed before the subject learns the rule, and learning efficiency.

Despite its nonverbal nature the WCST provided information about the cognitive skills
needed for more than simple sorting. It was thought that success at the WCST resulted from the
subjects ability to abstract and deduce the appropriate sorting rule, and then, maintain that
cognitive set for an appropriate time before shifting cognitive sets as needed to learn another
sorting rule. Because of its low use of language, the WCST seemed useful as a nonverbal tool to
estimate the cognitive skills of abstraction, deduction, and set-shifting, including the ability to
use simple feedback appropriately to guide the thought processes necessary for success.

The computer version of the WCST (used in this study) was normed across several
populations: Normal (N=65), Brain Injured (N=52), Frontal Damage (N=11), Nonfrontal
Damage (N=11), Diffuse Damage (N=30), and returned values met criteria for reliability and
validity. That is, individuals with no overt brain damage continually performed better than those
with cortical defect. Furthermore, for purposes of establishing construct reliability, performance
on the WCST was compared, and commensurate to that of performance on two other
neuropsychological tests of cognitive set-shifting: The Trail Making Test and the Stroop Test.

The WCST was chosen for this study because of its ability to measure an individual’s
ability to think abstractly and flexibly without the overt use of language. As such, it was believed
that performance on this test, along with that of the Block Design portion of the WASI and the

38
VMI served as good indicators of nonverbal information processing skill. Further, because performance on this test called upon abstraction and problem solving skill for success, its inclusion in a model of intelligence that contains both verbal and nonverbal measures seemed reasonable.

Test of Visual Motor Integration (VMI)

The VMI is a task that was designed to measure developmental changes in hand-eye coordination (Beery, 1997). As such, subjects were required to copy designs of increasing geometric complexity. Research on the VMI indicated that the mean internal split-half correlation of the test is .85 (range = .76 to .91) with a coefficient alpha of .82, and a standard error of measurement across the age range (3-17) not exceeding a standard score of 6 points (test mean = 100, SD = 15), all of which indicated high levels of internal consistency.

Correlations between age and performance on the VMI exceed 0.8, and have come close to 0.9 (see manual), indicating its suitability for estimating the development of the cognitive skills needed to integrate multimodal information. The test authors also claimed that because of the low amount of language demand needed to be successful on the VMI, performance was more related to nonverbal aspects of intelligence than verbal. Studies on the VMI have revealed that performance on the VMI correlated less well with verbal aspects of intelligence (Verbal IQ = .48) than nonverbal aspects (Performance IQ = .66). Similarly, performance on the VMI was more related to cognitive skills such as reading than, say mathematics. Performance on the VMI has not been as useful in distinguishing low achieving from learning disabled children, suggesting that a variety of experiences (e.g., slow maturation, low socioeconomic conditions and poor academic environment etc) may affect success on this task.
Interestingly, low performance on the VMI has been shown to be useful in identifying boys in kindergarten, who later struggled with reading. Furthermore, VMI performance has been tentatively linked to poor psycholinguistic performance. These results suggest, that while the VMI is a nonverbal task, and related to nonverbal intelligence, it is sensitive to deficits in integrative multimodal information. As such, VMI performance may well prove a good indicator of problem solving skills across several domains of cognition. This may account for why poor performance on the VMI in kindergarten has been predictive of academic difficulty by the second grade. The VMI was chosen for this study because of its ability to potential estimate an individual’s intelligence nonverbally. That is, because the VMI has been shown to be predictive of an individuals’ ability to integrate multimodal information, this variable, in conjunction with other measures of intelligence with low linguistic demand will serve as a good indicator of nonverbal intelligence.

Comprehensive Test of Phonological Processing (CTOPP)

Skill in phonological processing is considered crucial for the development of reading (Lombardino, Riccio, Hynd, & Pinheiro, 1997). As such, phonological processing is a cognitive skill most often related to academic achievement and, only rarely, related to general intelligence. Skill in phonology was included as one of the primary factors in Carroll’s three-stratum theory, so it can be argued that it should play a role in the present model of intelligence.

Phonological processing consists of the analysis and synthesis of the simplest parts of speech (phonemes). Skill in phonological processing includes support skills in phonemic awareness, phoneme segmentation and the use of phonological memory (Lennon & Slesinski, 2001). Skill in phonological processing allows an individual to combine sounds and place them in a meaningful context through the formation of words, eventually leading to comprehension.
The CTOPP estimated phonological skill. First published in 1999, the CTOPP is a relative new measure of phonological skill designed to be an improvement and serve as an extension over existing measures (Lennon & Slesinski, 2001). The development of the CTOPP was theoretically driven and supported through factor analytic and empirical studies. In general, the CTOPP covers a broad age range (5 – 24) and has superior psychometric properties to many other commercially available tests of phonological processing (Lennon & Slesinski). Normative data were collected from over 1600 individuals stratified to match US census data from 1997 on variables including: gender, race, socioeconomic status, and parental education etc. (see source for details) (Lennon & Slesinski, 2001). Reliability estimates generally exceed .80, which is considered more than adequate for clinical use (Lennon & Slesinski, 2001). Test–retest estimates range from a low of .70 to above .90. Such consistently high reliability coefficients suggest the presence of a relatively small amount of measurement error and support its use clinically. The test authors offer two versions of the test, one for 5 and 6 year olds and another for 7 – 24 year olds inclusively, both with separate norm tables. For this study, it was the later of the two versions used. Additionally, only a few of the subtests of the CTOPP were selected for this study: Phonological Reversal, Blending, and Segmenting nonwords, because it is anticipated that those were the most sensitive measures of the phonological skills needed to learn the principal components of language; as well as potentially representing the core deficit faced by individuals with dyslexia (Lombardino et al., 1997).

The phoneme reversal was a task that assessed a child’s ability to encode and manipulate individual phonemes presented verbally. The task required subjects to listen to a set of phonemes presented on an audio-tape, and repeat them aloud in the reverse order. The task progressed from
easy (e.g., hear ‘er’ and says ‘re’) to hard (e.g., hear ‘le’ and says ‘el’) (Richard K. Wagner, Torgesen, & Rashotte, 1999).

In the Blending nonwords task, the child was asked to put together a string of sounds to form a complete nonword (e.g., ‘skree – doo’ makes ‘skreedo’) (Richard K. Wagner et al., 1999). To avoid the possibility that a child recognized the component sounds of a real word, the nonword task was chosen for this study.

For the Segmenting nonword task, the child first heard a nonword, and was asked to repeat the made up word, one sound at a time. “For example, the examinee listened to the audiocassette-recorded sounds of ‘ren,’ then repeated the nonword in entirety, and finally said the nonword one sound at a time. The correct response is ‘r -ē- n’” (Richard K. Wagner et al., 1999)

Taken as a group, these three tasks appeared to resemble the core component processes that a child must master in order to develop normal reading ability. Thus, these three tasks served as indicator variables for the Phonology factor in the present model of intelligence. Children who experienced difficulty with any or all of these component processes likely experienced difficulty acquiring the core phonological awareness skills to reach age appropriate levels of academic achievement. Thus, children who struggled with phonological processing were probably at risk for being identified as having a learning disability (Hiemenz & Hynd, 2000; Lombardino et al., 1997).

The goal then was to employ structural equation modeling to complete multiple confirmatory analyses simultaneously, which met several methodological goals in the most parsimonious manner possible. Namely, to replicate the expected findings that verbal and nonverbal measures had strong factor loadings on a single general intelligence factor. Further,
the present model could extend the findings reported by Carroll in 1993, which clearly demonstrated that cognitive skills typically ascribed to the realm of academic achievement (i.e., phonology) had a robust, and direct, factor loading on a single index of intelligence.

Statistics

Brief overview of structural equation modeling (SEM)

Structural equation modeling (SEM) was used to test the hypothesis that a general intelligence factor predicts/influences performance on both tests of intelligence and academic achievement. Before reporting the results, a brief overview of structural equation modeling (SEM) is beneficial. “Structural equation modeling may best be defined as a class of methodologies that seeks to represent hypotheses about the means, variances, and covariance of observed data in terms of smaller number of “structural” parameters defined by a hypothesized underlying model” (Kaplan, 2000). SEM attempts to reach these goals by estimating models of construct measurement and structural integrity. The measurement model evaluates how well the indicator variables represent a latent construct. One usually assumes that indicator variables are unidimensional. Unidimensionality implies that a set of indicator variables represent not more than one latent factor. When a set of indicator variables are positively and similarly weighted (Anderson & Gerbing, 1988), there is evidence of unidimensionality. Therefore, it is extremely important that the indicator variables are chosen with care.

The structural model evaluates the relationships among the latent variables. That is, given the integrity of the measurement model, how well does the structural model explain the nature of the data (Kelloway, 1998). Thus, SEM provides information about the nature of variance from several sources at once.
Assumptions of SEM

As with many other statistical techniques, SEM has a few basic underlying assumptions. Four assumptions are of particular relevance. First, the model should evolve from theory. Although it is possible to use a data-driven approach for model generation, this is not a method of choice (Hagtvet, 2003). Secondly, the data should not be mutually influential. When conducting cross-sectional studies, researchers should ensure that the data arrive from independent observations. Thirdly, as error is ever present, one should ensure that it is not systematically correlated with the model variables. Fourth and finally, the data should be normally distributed, as deviations will seriously affect model estimation.

Because this study used family data, meeting the second assumption was potentially problematic. Because there are a number of siblings in the present data set, individual scores potentially violate assumption of independence. Nevertheless, there are methods available to deal with this situation. The method used to overcome this limitation was to shift the unit of analysis from the individual to the family. Care was taken during this process so as not to lose critical information through a reduction in variance in the data. Although not a new concept, the overall variance in the data was retained by using both the mean and standard deviation scores as indicator variables in the analyses. Hence, for this study, when more than one child from a family was assessed, scores were averaged, and that averaged score and the associated standard deviation score were used (Hagtvet, 2003).

Advantages

For the purpose of this study, SEM was particularly useful for several reasons. Because SEM is a multivariate technique, it is well suited for analyzing complex constructs such as intelligence; univariate techniques are limited because they provide only a narrow view of the relationships
present in data (Crowley & Fan, 1997). Because of the complex nature of cognition, a
dmultivariate technique considerably improves the study of intelligence. Indeed, many researchers
have called for an increase in the use of multivariate studies in both the intellectual and
achievement fields (Carroll, 1993; Daneman, 1991; Matarazzo, 1972; Spearman, 1904;
Thurstone, 1938). An additional advantage of SEM was the ability to specify and estimate a
theoretically derived model, as well as alternative models, which might fit the data equally well.
This ability provided greater flexibility in examining the complex and dynamic nature of
intelligence. SEM also proved flexible enough to accommodate the use of different item scale
scores for indicator variables. For example, in a review of 15 studies, it was found that
"intercorrelations averaged between .4 and .5 when scales were used as indicators, and between
.1 and .3 when items were used" (Cohen, Cohen, Teresi, Marchi, & Velez, 1990). Scale scores
were used in this study to enhance the reliability and lower the mean error of the indicator
variables (Knight, 2000).

Building a model

Generally, SEM examines the relationships among variables according to a pre-specified,
theoretical model. The present model originated from a review of the relevant intelligence and
achievement literature. Previous work validated the constructs of Verbal and Performance IQ
factors of the Wechsler scales, supporting their use in the present study. A third latent factor,
Phonology, was also included in the model. As mentioned in the “statement of the problem”
section, it was believed that a general intelligence factor can adequately account for the variance
and covariance in measures of intelligence and academic achievement. As such, the factor
loadings of all three first-order factors (VIQ, PIQ, and Phonology) on the second order factor,
general intelligence, were estimated.
The variables previously mentioned were used in the following manner. The measurement model estimated the presence of three primary factors. The Vocabulary and Similarities tasks represented a Verbal factor. Similarly, the Block Design and Matrices tasks represented a Performance factor, while the Phonological Reversal, Blending and Segmenting Nonwords tasks represented a Phonology factor. Then, the structural model estimated the relative contribution of each of these three factors to a second-order general intelligence factor. The combined measurement and structural model created the full SEM model.

In the measurement model (Figure 3), the lines from the circles to the squares, indicates the direction of influence (i.e., the latent factor influences performance on the indicator variables) and reveals which indicator variables represented the latent construct. Properly chosen indicator variables should have had fairly high and uniform factor loadings demonstrating their ability to represent the latent construct. Each set of indicator variables and purposed latent factor constituted an independent measurement model, individually estimated and evaluated via confirmatory factor analysis. The short arrows pointing to the indicator variables represent measurement error, which should have been minimal with the use of appropriate indicator variables. Again, the adequacy and precision of the indicator variables was crucial. Estimating a latent construct using imprecise measures challenges not only the validity of that particular construct (Maruyama, 1998), but the viability of the hypothesized structural model.

Figure 4 depicts the hypothesized structural model for this study. The structural model outlines the theoretical nature and direction of influence between the three first-order (primary) constructs (VIQ, PIQ, and Phonology) and a second-order construct (General Intelligence). This particular structural model was similar to the measurement model in that it estimated how well the primary factors represent the general intelligence factor. As with the measurement model, the
lines from the general intelligence factor to the primary factors represent latent factor loadings, or the relative contribution of the primary factor to the second-order factor. Unexplained variance at the latent level became error terms and was represented by the short lines pointing to the circles.

Although the full model of intelligence shown in Figure 5 resulted from an extensive review of the literature, it was possible that data fit another model of intelligence equally well or better. Therefore two alternative models of intelligence were examined to determine whether the competing models explained the data any better. If these alternative models of intelligence fit the data as well or better than the hypothetical model put forth earlier, then the viability of the initial model would have been called into question and another model would have taken its place.

Figures 6 & 7 depict the alternative models. In Figure 6, Phonology no longer had a direct factor loading on \( g \). Instead, Phonology was said to load onto the Verbal factor. The rationale for this model was that it may be plausible that the Verbal factor can reasonably absorb the variance present in the Phonology factor. Should this model of intelligence adequately fit the data, the results may have implied that the Phonology factor did not contribute anything unique above and beyond the information provided by the Verbal factor. The model of intelligence depicted in Figure 6, investigates the possibility that the Phonology factor may actually moderate the influence of \( g \) onto the Verbal factor. That is, by allowing the Phonology factor to load onto \( g \) and the Verbal factor simultaneously, it was possible to determine whether the Phonology factor provided any unique information in the estimation of either \( g \) of Verbal performance. If, say, a significant factor loading between the Verbal or Phonology factors and \( g \) (in the hypothesized model) was no longer significant in the second alternative model, one implication would be that one of the factors moderates the influence of intelligence on the other factor. That is, there would...
be sufficient overlap of information provided by the two factors, that when allowed to covary, only one of them was significantly influenced by intelligence.

Figure 3: Measurement Model of Intelligence
Figure 4: Structural Model
Figure 5: Full Structural and Measurement Model

Note: V= Vocabulary test, S= Similarities test, B= Block Design test, W= Wisconsin Card Sort Test, VM= Test of Visual Motor Integration, PR= Phonological Reversal test, BN= Blending Nonwords test, SN= Segmenting Nonwords test
Figure 6: First Alternative model
Note: V= Vocabulary test, S= Similarities test, B= Block Design test, W= Wisconsin Card Sort Test, VM= Test of Visual Motor Integration, PR= Phonological Reversal test, BN= Blending Nonwords test, SN= Segmenting Nonwords test

Figure 7: Second Alternative model
Chapter 4

Results

Structural equation modeling was used to evaluate the relationships among variables traditionally used to measure intelligence, executive functioning, visual-spatial/motor skills, and linguistic variables that support reading (i.e., phonological processing) hypothesized to contribute nearly equally well to a single construct of intelligence ($g$). An indicator was set to 1.0 to establish a metric for each latent variable. Chi-square difference tests along with several ad-hoc fit indices were used to evaluate the viability of the proposed model.

As mentioned previously, three models of intelligence were to be examined in this study. The first hypothetical model was the one proposed (factor loadings shown in Figure 8), and the second and third models examined whether alternative models of intelligence, using the same indicator variables and latent factors fit the data equally well or better. Regarding the two alternative models of intelligence, the relevant fit indices (outlined below) indicated that the alternative models fit the data reasonably well. However, the second-order factor loadings of both of the alternative models were negative (see Figures 9 & 10). Implicit in these findings is that as performance on the primary latent variable improves, overall intelligence declines. Because the factor structure of the alternative models was not consistent with any expectation of how the data reflect cognition, and the purposed hypothetical model did conform to expectations, the results discussed below focus solely on fit and relevance of the purposed model of intelligence shown in Figure 8.
Measurement model

Initially the descriptive statistics for each indicator variable were examined to determine the normality and variability and can be found in Table 3. For the most part, the variables were normally distributed with and adequate range to indicate that there were neither floor nor ceiling effects for any of the tests. It has been suggested that variables with values for skewness of ± 2 and values for kurtosis of ± 7 can be considered adequately normal for maximum likelihood estimation, especially if measurement of those variables is continuous (West, Finch, & Curan, 1995). No observed variable in this study had values that exceed either of those values.

Correlations among the indicator variables are shown in Table 4.

### Table 3: Cognitive Performance

<table>
<thead>
<tr>
<th></th>
<th>N=79</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
<th>Min</th>
<th>Max</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td></td>
<td>96.20</td>
<td>16.32</td>
<td>77</td>
<td>58</td>
<td>135</td>
<td>-0.10</td>
<td>-0.01</td>
<td>1.84</td>
</tr>
<tr>
<td>Similarities</td>
<td></td>
<td>102.37</td>
<td>13.49</td>
<td>81</td>
<td>57</td>
<td>138</td>
<td>1.47</td>
<td>0.00</td>
<td>1.52</td>
</tr>
<tr>
<td>Block Design</td>
<td></td>
<td>100.38</td>
<td>15.89</td>
<td>78</td>
<td>67</td>
<td>145</td>
<td>0.39</td>
<td>0.70</td>
<td>1.79</td>
</tr>
<tr>
<td>WCST</td>
<td></td>
<td>94.18</td>
<td>17.03</td>
<td>67</td>
<td>50</td>
<td>117</td>
<td>-0.50</td>
<td>-0.68</td>
<td>1.92</td>
</tr>
<tr>
<td>VMI</td>
<td></td>
<td>95.16</td>
<td>13.99</td>
<td>69</td>
<td>68</td>
<td>137</td>
<td>0.42</td>
<td>0.52</td>
<td>1.57</td>
</tr>
<tr>
<td>Segmenting Nonwords</td>
<td></td>
<td>7.06</td>
<td>3.40</td>
<td>18</td>
<td>0</td>
<td>18</td>
<td>1.67</td>
<td>0.57</td>
<td>0.38</td>
</tr>
<tr>
<td>Blending Nonwords</td>
<td></td>
<td>6.80</td>
<td>2.99</td>
<td>13</td>
<td>0</td>
<td>13</td>
<td>-0.39</td>
<td>-0.01</td>
<td>0.34</td>
</tr>
<tr>
<td>Phonological Reversal</td>
<td>3.72</td>
<td>3.18</td>
<td>13</td>
<td>0</td>
<td>13</td>
<td>0.74</td>
<td>0.95</td>
<td>0.36</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Correlations between indicator variables

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.29</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>0.06</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
<td>0.14</td>
<td>0.58</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.35</td>
<td>0.44</td>
<td>0.32</td>
<td>0.09</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.19</td>
<td>0.14</td>
<td>0.35</td>
<td>0.19</td>
<td>0.23</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.22</td>
<td>0.25</td>
<td>0.29</td>
<td>0.11</td>
<td>0.14</td>
<td>0.44</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ideally, variables hypothesized to load on the same latent factor will be more highly correlated with one another than with other observed variables hypothesized to load on a different latent factor. The correlations of the observed variables purported to measure the verbal latent factor were more highly correlated with each other than any other factor. However, inspection of the correlation matrix revealed a couple of potential measurement problems for the latent variables. The first concern was that the measures of the nonverbal factor were correlated with those of the phonology factor. Although the correlations indicated that the measures of nonverbal ability were, for the most part, strongly related to each other, they also demonstrated some overlap with measures of phonological processing. This may be due to the fact that the tasks for both factors require the analysis of part to whole relationships.

To determine the extent to which each observed variable contributed to the construct of interest, a confirmatory factor analysis was run on each measurement model. Results of this analysis are found in Table 5. Each construct was examined for unidimensionality, which is indicated by similarly positively weighted factor loadings. The observed variables for each latent variable had adequate factor loadings that were significantly different from zero. The squared multiple correlation (SMC) for each observed variable indicated the amount of variance in that variable that is uniquely attributable to the latent factor (Bollen, 1989). For example, the SMC for the segmenting nonwords task is .64, meaning that 64% of the variance in this measure is explained by the latent factor phonology. The influence of the latent factors on the observed variables was sufficient for use in the measurement model and subsequent analyses. Information specific to each latent variable is described below.
Table 5: Parameter Estimates

<table>
<thead>
<tr>
<th>Latent construct</th>
<th>Observed Measure</th>
<th>Loading</th>
<th>Error</th>
<th>SMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>Vocabulary</td>
<td>.72(^a)</td>
<td>.48*</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Similarities</td>
<td>.80*</td>
<td>.34*</td>
<td>0.65</td>
</tr>
<tr>
<td>Nonverbal</td>
<td>Block Design</td>
<td>.99(^a)</td>
<td>0.01</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>VMI</td>
<td>.58*</td>
<td>.67*</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>WCST</td>
<td>.36*</td>
<td>.87*</td>
<td>0.13</td>
</tr>
<tr>
<td>Phonology</td>
<td>Segmenting Nonwords</td>
<td>.80(^a)</td>
<td>.36*</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Blending Nonwords</td>
<td>.61*</td>
<td>.63*</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>Phonological Reversal</td>
<td>.57*</td>
<td>.67*</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Note: \(^a\) variable used to set metric for latent variable; \(^*=p<.05\); Loadings are completely standardized estimates; SMC= squared multiple correlations

Verbal skill. All of the factor loadings on the verbal skill factor were strong (WASI Vocabulary .72, WASI Similarities .80).

Nonverbal Skill. Regarding Nonverbal information processing, the WASI Block Design clearly had the strongest factor loading (.99) than the other two indicator variables. The VMI had a more moderate factor loading (.58), whereas the WCST had a significant, but relatively weaker factor loading on the Nonverbal skill factor (.36). Factor loadings are comprised of the shared variance among the indicator variables for a given latent variable. Because the Block Design task required the manipulation of blocks in 3D space, it may be more related to nonverbal information processing than the other two variables, therefore having less common variance with them. Regarding the relatively weak contribution of the WCST, it is possible that individuals may use “internal self-talk” as they attempt to determine what rule to use to sort the cards correctly. It is important to remember, that, although the WCST had a weaker factor loading, it was still statistically significant.

Phonological processing. The factor loading for the phonological processing factor ranged moderate to high. Segmenting nonwords (.80), which requires the deconstruction of a
nonword into its component sounds, had a strong factor loading on the phonological processing construct. Blending nonwords (.61) and Phoneme reversal (.57) both had moderate factor loadings on the phonological latent factor. The factor loadings for the Blending Nonwords and Phoneme Reversal may be slightly lower because they require more knowledge and familiarity with the linguistic rules, compared to the deconstruction of nonwords employed by the Segmenting Nonwords task.

Structural Model

The structural model consists of the measurement model, which links the observed and latent variables, plus the structural model, which specifies the relations among the latent variables. The results of the full structural model are reported below.

The model was initially evaluated for overall fit, that is, how well the data were able to reproduce the covariance matrix implied by the hypothesized model (Hoyle, 1995). The overall fit indices included in this study fall into two categories: absolute fit and comparative fit. Goodness-of-fit (Tanaka & Huba, 1984) is a measure of absolute fit that measures the amount of variance explained by the model, a statistic analogous to $R^2$. Values greater than .90 are generally considered to indicate acceptable fit. The other index of absolute fit is the root mean square error of approximation [RMSEA (Steiger, 1990)]. RMSEA is a measure of the amount of error between the model and the data. Because it is measuring the degree of error, a value of 0 would indicate a perfect fit. Some have suggested that RMSEA values less than .05 indicate a close fit of the model, .08 a reasonable fit, and .1 an acceptable fit (Browne & Cudeck, 1993). In addition to the RMSEA point estimate, a confidence interval for it is produced which estimated the probability of fit in the population at large. This statistic can serve as an indication of the likelihood that the model could be replicated in a cross-validation study.
Two measures of comparative fit are also included. These indices measure the closeness of fit between the null hypothesis with no relationships and the model being tested. The nonnormed fit index [NNFI; (Bentler & Bonett, 1980)] has been shown to perform well when the maximum likelihood estimator is used as the method of estimation; however, it is sensitive to sample size. The comparative fit index (CFI) (Bentler & Bonett, 1980) performs similarly to the NNFI index, but is less sensitive to sample size (Hu & Bentler, 1995). Models with values greater than or equal to .90 have acceptable fit. Using multiple fit indices permits a broader and more comprehensive evaluation of the fit of the model. The residual matrix is another indication of the closeness of fit between the covariance matrix implied by the hypothesized model and the sample covariance matrix. An analysis of the residuals can reveal any discrepancy between the two. For ease of comparison, the residuals are standardized, and any value greater than 2.58 are considered large (Byrne, 1998).

The fit indices shown in Table 6 indicated close overall fit of the structural model. The RMSEA was 0.0 with the probability of close fit ranging from 0.0 to 0.04; note that the upper bound value of the confidence interval is less than the value suggested by Browne and Cudeck (1993) as appropriate for a close fitting model. All of the measures of absolute and comparative fit exceeded the .90 criterion level. No residuals exceeded the 2.58 criterion level. The fit indices and small residuals suggested that the hypothetical model of intelligence reproduced the covariance matrix reasonably well. Of note, however, is that the NNFI value (1.06) exceeded 1.0, this may indicate some misspecification of the model, or more likely, because the value is so small, is due to random error.
Table 6: Fit Indices

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>17</td>
</tr>
<tr>
<td>(\chi^2)</td>
<td>10.13</td>
</tr>
<tr>
<td>GFI</td>
<td>0.97</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.00</td>
</tr>
<tr>
<td>RMSEA 90%</td>
<td>0.0 - .04</td>
</tr>
<tr>
<td>RMSEA p Value</td>
<td>0.96</td>
</tr>
<tr>
<td>NNFI</td>
<td>1.06</td>
</tr>
<tr>
<td>CFI</td>
<td>1.00</td>
</tr>
</tbody>
</table>

In addition to overall fit, the path coefficients were examined to evaluate the strength of the relationships among the latent variables. Paths were considered statistically significant when the test statistic was greater than or equal to \(\pm 1.96\). Taken together, the three primary factors: Verbal Skill, Nonverbal Skill, and Phonological processing, sufficiently identified the presence of the second-order \(g\) factor. However, the pattern of factor loadings was surprising.

**Overall Intelligence (g).** Second-order factor loadings are presented in Table 7. Verbal skill had a strong indicator of intelligence, with a factor loading of (.69) on \(g\), the second order latent factor representing overall intelligence. Traditionally, verbal skill has provided a significant contribution, and estimate of intelligence. In keeping with tradition, Nonverbal skill had a more moderate factor loading on \(g\) (.55). That the Nonverbal factor loading is somewhat smaller than Verbal skill is not very surprising, as the use of language is the easiest cognitive skill to assess. For this reason alone, the majority of cognitive assessment tools rely heavily on one’s proficiency in language. The factor loading attributed to Phonology (.83) was surprising. Although it has been established that proficiency with the general use of language is generally highly correlated to intelligence, finding that the discrete cognitive skills such as phonological processing would have such a high factor loading was not expected. The present findings raise several questions about the nature of information processing and defining intelligence.
Table 7: Second-order factor loadings

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>0.69</td>
</tr>
<tr>
<td>Nonverbal</td>
<td>0.55</td>
</tr>
<tr>
<td>Phonology</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Note: V= Vocabulary test, S= Similarities test, B= Block Design test, W= Wisconsin Card Sort Test, VM= Test of Visual Motor Integration, PR= Phonological Reversal test, BN= Blending Nonwords test, SN= Segmenting Nonwords test
Note: V= Vocabulary test, S= Similarities test, B= Block Design test, W= Wisconsin Card Sort Test, VM= Test of Visual Motor Integration, PR= Phonological Reversal test, BN= Blending Nonwords test, SN= Segmenting Nonwords test

**Figure 9: First Alternative model factor loadings**
Note: V= Vocabulary test, S= Similarities test, B= Block Design test, W= Wisconsin Card Sort Test, VM= Test of Visual Motor Integration, PR= Phonological Reversal test, BN= Blending Nonwords test, SN= Segmenting Nonwords test

**Figure 10: Second Alternative model factor loadings**
Summary

The relationships among latent variables in a second-order model of intelligence were evaluated with structural equation modeling. In a two-step approach, the measurement model was examined first before considering the overall fit of the structural model. In general, the indicator variables for the latent factors had adequate unidimensionality and factor loadings to estimate the constructs they were purported to measure.

Overall fit of the full model was adequate, however the fact that one of the fit indices exceeded 1.0 may be of some concern. This value maybe simply due to random error present in the data, and that with either another or larger data set the value for the fit index would meet the previously mentioned fit index criteria. The pattern of relationships of the Verbal and Nonverbal skill to $g$ was in line with expectations. However, finding that Phonological processing had a larger factor loading than either Verbal or Nonverbal skill was unexpected.
Chapter 5

Discussion

The goals of this study were twofold. The first was to examine a theoretical model of intelligence consistent with the first one proposed nearly 100 years ago (Diseth, 2002; Spearman, 1904). This exercise would help validate the model of intelligence Carroll identified using exploratory factor analyses (Carroll, 1993). The second and subsequent goal of this study was to empirically demonstrate how the viability of such a model as the one tested does not support the use of the discrepancy model or the logic behind comparing IQ and academic achievement scores (Colarusso et al., 2001; Kershner, 1990). The results of this study were largely in support of the proposed model. All three primary latent variables were sufficient to identify the presence of a second-order $g$ factor of intelligence. Even though all three primary factors were significant indicators of the $g$ factor, the pattern of factor loadings were unexpectedly gratifying. In addition to the relationships among the latent factors, the implications of the observed pattern of factor loadings are of interest.

The discussion that follows will briefly review the stated intent of this study before considering the viability of the hypothetical model of intelligence. Following that, the patterns of relationships among the latent variables are evaluated in light of the principal research questions addressed in this study. Finally, the potential implications resulting from findings such as those reported here, are discussed are they relate to the fields of education and neuroscience.

The attempt to accurately describe and quantify intelligence is centuries old (Plucker, 2003). The theoretical and empirical advances of the early 1900s made it possible for researchers from the fields of psychology and mathematics to develop and test hypothetical models of intelligence (Spearman, 1904; R. J. Sternberg, 2000). Spearman put forth the first empirically
supported model of intelligence in 1904. According to his theory, a single latent factor
influenced the whole of cognition. This latent factor, he termed \( g \) for \textit{General Intelligence}. Since
Spearman first introduced his model of intelligence, rival models of intelligence have emerged.

Each successive model of intelligence is purportedly more comprehensive because they
often assess a large sample of cognitive skills (Carroll, 1993). Typically, these skills form
smaller cognitive processes represented by latent factors (Bollen, 1989). However, in nearly all
of the successive models of intelligence, the latent factors are either correlated and/or mutually
influential to each other (Carroll, 1993; Guilford, 1985). This means that a single latent factor
that could account for the inter-relations between those latent factors. The lack of appreciation
for this concept by policy makers in the field of education has lead to the development of the
discrepancy model; a set of procedures employed to determine whether a child is eligible for
academic assistance in the school setting.

Although there are many ways to employ the policy and procedures of the discrepancy
model, the general concept among all of the procedures is the same (Colarusso et al., 2001;
Kershner, 1990; Peterson & Shinn, 2002; Proctor & Prevatt, 2003). A child completes an IQ test,
and through any number of procedures, their IQ score is compared to their level of academic
achievement. Again, depending on the exact requirements of the child’s locale, if their IQ score
is sufficiently discrepant from their level of academic attainment, they are eligible for academic
assistance. Research on the efficacy of the discrepancy model is typically disparaging (Colarusso

Regardless of the method, or standards used to establish an IQ achievement discrepancy,
there is a growing body of literature suggesting that the discrepancy model is unable to meet the
needs of its design (Vellutino, 2001). In spite of the methodological and computational criticisms
leveled at the discrepancy model, the fact that IQ and achievement test scores are compared to each other is suggestive of a lack of appreciation about the nature of information processing. As conceptualized by Spearman, intelligence is a single entity that represents an individual’s overall ability to encode, decode, synthesize, and use information (Carroll, 1993; Diseth, 2002; Spearman, 1904). This concept should apply to all forms of information processing. Therefore, comparing IQ and achievement test scores may not be a valid exercise from a theoretical perspective and may help explain why the research on the utility of the discrepancy model is not encouraging. However, given that Spearman’s model of intelligence is 100 years old, contemporary policy makers in the field of education may feel that his model is dated. In his effort to understand the nature of intelligence, Carroll, in 1993, conducted a meta-analysis on information processing and intelligence. Using exploratory factor analyses, he found that indeed, a single latent factor influenced all of the facets of cognition in his study. However, he was cautious in interpreting his model of intelligence, calling for confirmatory factor analytic studies to help continue to validate the $g$ model of intelligence (Carroll, 1993).

Viability of the purposed model. Among the primary goals of this study was to validate Carroll’s 1993 model of intelligence. Towards this end, a second-order factor analysis was conducted using structural equation modeling. In keeping with the spirit of both Spearman and Carroll, the current model of intelligence included measures from a broad range of cognitive processes. Specifically, individual performances on eight different tasks served as estimators of three primary latent factors: Verbal skill, Nonverbal skill, and Phonological processing. These three primary latent factors then estimated the presence and influence of a single second-order latent factor – $g$. Results of this study are consistent with Spearman, Carroll, and others who have considered intelligence a single global entity. According to the SEM fit indices, the
hypothesized model of intelligence tested here conforms to the data reasonably well (Bollen, 1989; Browne & Cudeck, 1993; Byrne, 1998; Cohen et al., 1990). However, even though many of the fit indices meet criteria for a well fitting model, there are a few indications that there is room for the current model of intelligence to be improved.

Despite the strength of the fit indices, the final estimates of the hypothetical model of intelligence suggest areas of improvement. Beginning with the indicator variable error terms, there are indications that the present model of intelligence could be improved. Error terms are but one indication of how representative the indicator variable is of the latent construct (Browne & Cudeck, 1993). The more variance that is attributable to a latent construct, the smaller the error term (Browne & Cudeck, 1993). In the present study, half of the indicator variables had error terms greater than .60, suggesting that some other variable may be a better indicator of that latent construct. Relatively low indicator variable loadings are another area of improvement for the current model of intelligence. Although all of the indicator variables used in this study had significant factor loadings, it is preferred that the factor loadings should have values equal to or greater than .70. Factor loadings below .70 suggest that there are might be other variables that may be better estimators of the latent factor. The same concepts hold true when discussing the error terms and beta weights of the primary latent factors. If the indicator variables are poor estimators of the latent factor, there will in turn be more error associated with that latent factor. In the present study, the WCST, although significant, proved to be a relatively poor estimator of Nonverbal skill. This may be attributable to the possibility that subjects use internal speech to guide their responses as they complete the task. Thus, despite the low demands on overt language necessary to complete the task, it may be that there is a moderate amount of internal
dialogue occurring during the testing session. Methodological improvements are another way to enhance the validity of the purposed model of intelligence (Hagtvet, 2003).

The assumptions made of the data in the present study may call into question the methods used to estimate the hypothesized model of intelligence. Subject data used in the present study came from a grant investigating the familial components of developmental dyslexia. Consequently, the data from the 79 individuals used to evaluate the purposed model of intelligence originated from 58 families. Because the present model and methodology employed did not employ a tactic to account for the common or shared variance known to be present amongst family members, it is possible, that when this shared variance is reconciled, the indices related model of intelligence evaluated here would change (Hagtvet, 2003). One potential method for accounting for the shared variance between siblings is to use averaged family scores (Hagtvet, 2003). However, there is no consensus as to how best include the within family variance in SEM when the siblings are not twins.

From a statistical perspective, it seems reasonable to conclude that the model of intelligence examined in this study, while not definitive, is a valid example of how human cognitive processes may be organized. The hypothetical model of intelligence in this study found that a single, second-order factor of intelligence adequately explained the amount of variance present in three primary latent factors representing disparate aspects of information processing consistent with earlier work (Carroll, 1993; Stage et al., 2003; Ullstadius et al., 2002; Ward et al., 1999). The present model of intelligence appears to extend the first scientifically acceptable one put forth a century ago (Spearman, 1904). In his model, Spearman characterized intelligence with just one primary latent factor (Carroll, 1993; Diseth, 2002; Spearman, 1904). Further, the present study helps to validate Carroll’s model of intelligence, derived via exploratory factor
analyses. The model of intelligence tested in this study demonstrated the validity and reliability of characterizing intelligence as a single entity using higher order confirmatory factor analyses (Carroll, 1993; Stage et al., 2003; Ullstadius et al., 2002; Ward et al., 1999).

**Pattern of factor loadings.**

The pattern of factor loadings associated with the hypothetical model of intelligence presented here are worthy of comment. Overall, the strength and pattern of factor loadings of the Verbal and Nonverbal latent factors were consistent with expectations. The verbal factor had a slightly stronger relationship with the second-order latent factor $g$ than the Nonverbal factor. This was expected due to the historical nature of intelligence test construction, which has relied heavily on the use of language to exemplify successful cognitive processing skills (Carroll, 1993; D'Angiulli & Siegel, 2003; Murray & Wren, 2003; Remschmidt et al., 1999; Robert J. Sternberg et al., 2001; Ullstadius et al., 2002). Consequently, intelligence as currently conceived, is to a greater degree, a measurement of how well an individual can use language efficiently and effectively (Carroll, 1993; Spearman, 1904; Tiu et al., 2003; Ullstadius et al., 2002). This is not to discount the information provided by the Nonverbal factor.

The Nonverbal factor represents that aspect of cognition that is not amenable to linguistic functioning, but nonetheless provides crucial information about how well the individual incorporates and acts upon information (Anastasi, 1988; Tiu et al., 2003). Historically, nonverbal information processing represents novel problem solving skills (Kamphaus, 2001; Sattler, 1992). Because it is easier, practical, and is similar to everyday experience, the use of language has evolved as a relatively stronger predictor of intelligence than the ability to solve problems in novel situations. Despite this expected pattern of latent factor loadings for the Verbal and Nonverbal factors (Carroll, 1993; D'Angiulli & Siegel, 2003; Kamphaus, 2001; McBride-Chang,
1995), the strength of the relationships of the Phonology factor with the intelligence factor was somewhat surprising.

This study attempted to demonstrate once again that it is reasonable to conceive of intelligence as a single latent factor. The fit indices (Bentler & Bonett, 1980) of the models tested here appear consistent with previous models of intelligence. A single second-order, latent factor influenced two primary factors in predictable ways. Specifically, the data here supported the presence of verbal and performance factors that appear consistent with both the popular VIQ/PIQ model and the Gf/Gc models of intelligence (Carroll, 1993; Horn, 1985). Thus, this single intelligence factor presumably influences nearly all domains of cognition as first demonstrated by Spearman in 1904, and then again by Carroll in 1993. However, unlike previously investigated models of intelligence (Anastasi, 1988; Horn, 1985; Sattler, 1992), the hypothetical model of intelligence examined here revealed that intelligence may have a direct influence on the processing of such discrete information as phonemes.

This last finding was somewhat surprising. Although, overall intelligence should influence phonological processing (McBride-Chang, 1995), there is little historical evidence to support such a notion of this proportion. It previously stood to reason that discrete cognitive skills such as phonological processing should either support broad based language functioning (such as estimated by the Verbal factor), or be a cognitive skill so different from intelligence that it was not credited as being able to provide information about an individual’s level of intellect (Knight, 2000; Majoribanks, 1976). Because the ability to use language purposefully is the most overt cognitive skill humans possess, it comes as no surprise that effective language use is the most prominent feature in the assessment of intelligence (Horn, 1985; 1994; Spearman, 1904). Perhaps for that reason alone, measures of broad language skills have traditionally had relatively
high factor loadings on overall intelligence (Kamphaus, 2001; Sattler, 1992). Broad measures of language aside, skill in diverse nonverbal information processing, or novel problem solving skills, historically have the second strongest relationship with overall intelligence (Carroll, 1993; Horn, 1985; 1994; Kamphaus, 2001; Sattler, 1992). This pattern of factor loadings is present in nearly all models of intelligence, whether in the popular VIQ/PIQ model, the Gf/Gc model, or any derivation of the two (Guilford, 1985). According the psychometric landscape of today, brief measures of broad language and nonverbal skills appear sufficient to provide a reasonable estimate of an individual’s intelligence (Zhu, 1999). Yet, the results presented here indicate that discrete cognitive skills such as phonological processing may provide a rich amount of information about the way the brain processes information from many different perspectives, each with many possible implications.

Possible implications

Because the Verbal factor represents broad language functioning (Matarazzo, 1972), it should have the highest factor loading on the second-order g factor of intelligence (Reynolds & Kaufman, 1985; Sattler, 1992; Thurstone & Thurstone, 1941). As the phonology factor represents a set of more discrete cognitive skills (Lombardino et al., 1997; R. K. Wagner et al., 1987; R. K. Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993), it might be that the factor loading would, at best be equal to, but probably less than that of the Nonverbal factor. Surprisingly though, the factor loading for Phonology exceeded that of the Verbal factor. Whereas phonological processing was expected to significantly contribute to the formation of the intelligence factor (McBride-Chang, 1995), is was not thought to have such a high factor loading because it traditionally represents such a small amount of the total information processing skills.
an individual possesses (Carroll, 1993; Horn, 1985). If validated, the possible implications for such a finding have the potential to dynamically change the field of intelligence testing and educational placement (Kershner, 1990; O'Malley et al., 2002; Peterson & Shinn, 2002; Proctor & Prevatt, 2003). In the following section some of those possible implications are discussed in terms of the conceptualization of intelligence, the incorporation of findings from neuroscience (B. A. Shaywitz et al., in press; B. A. Shaywitz et al., 2002; S. E. Shaywitz & Shaywitz, 1999; S. E. Shaywitz et al., 1998) are incorporated into the fields of education (G. W. Hynd & Cohen, 1983; G. W. Hynd & Semrud-Clikeman, 1989; Lombardino et al., 1997) and intervention, and finally educational policy ("Editorial: Study analyzes proposals to change LD identification," 2003; "Learning disabilities, dyslexia, and vision: a subject review. Committee on Children with Disabilities, American Academy of Pediatrics (AAP) and American Academy of Ophthalmology (AAO), American Association for Pediatric Ophthalmology and Strabismus (AAPOS)," 1998).

To begin, if indeed intelligence influences discrete cognitive skills through refined statistical techniques (Anderson & Gerbing, 1988), then the intelligence paradigm put forth by Spearman needs revisiting and fully embraced. The clear implication is that, truly, intelligence is the sum total of an individual’s ability (Wechsler, 1939) to incorporate and efficiently and effectively use information from a wider range of cognitive domains than previously thought. It would no longer make sense to hold up one cognitive skill as a standard for all other cognitive skills, as each one uniquely contributed to one’s overall intelligence (Horn, 1985; Spearman, 1904; Terman & Oden, 1947; Wechsler, 1939). Research from the field of neuroscience is, more so than ever before, making great strides in revealing strong brain/behavior relationships, showing how many different parts do contribute the whole of intellect (Eckert & Leonard, 2000; Eckert et al., 2002; Eckert et al., 2003; Heimenz & Hynd, 2000; G. W. Hynd & Cohen, 1983;
George W. Hynd, Marshall, & Semrud-Clikeman, 1991; G. W. Hynd & Semrud-Clikeman, 1989; G. W. Hynd, Semrud-Clikeman, Lorys, Novey, & Eliopulos, 1990; Leonard et al., 2001; Lombardino et al., 1997; B. A. Shaywitz et al., in press; B. A. Shaywitz et al., 1995; B. A. Shaywitz et al., 2002; S. E. Shaywitz et al., 1999; S. E. Shaywitz & Shaywitz, 1999; S. E. Shaywitz, Shaywitz, Fletcher, & Escobar, 1990).

During the last several decades, advances in the field of neuroscience, and in neuroimaging in particular, have provided additional evidence of brain behavioral relationships, especially in regards to cognition (Eckert & Leonard, 2000; Eckert et al., 2002; Eckert et al., 2003; Heimenz & Hynd, 2000; G. W. Hynd & Cohen, 1983; George W. Hynd et al., 1991; G. W. Hynd & Semrud-Clikeman, 1989; G. W. Hynd et al., 1990; Leonard et al., 2001; Lombardino et al., 1997; B. A. Shaywitz et al., in press; B. A. Shaywitz et al., 1995; B. A. Shaywitz et al., 2002; S. E. Shaywitz et al., 1999; S. E. Shaywitz & Shaywitz, 1999; S. E. Shaywitz et al., 1990).

Among the more relevant findings are those related to phonological processing and developmental dyslexia (Eckert & Leonard, 2000; Eckert et al., 2002; Eckert et al., 2003; Heimenz & Hynd, 2000; G. W. Hynd & Cohen, 1983; George W. Hynd et al., 1991; G. W. Hynd & Semrud-Clikeman, 1989; G. W. Hynd et al., 1990; Leonard et al., 2001; Lombardino et al., 1997; B. A. Shaywitz et al., in press; B. A. Shaywitz et al., 1995; B. A. Shaywitz et al., 2002; S. E. Shaywitz et al., 1999; S. E. Shaywitz & Shaywitz, 1999; S. E. Shaywitz et al., 1990).

Investigations from multiple fields have confirmed that the temporo-parietal region, and the planum temporale in particular relate to the accurate processing of neurolinguistic information. Neuroanatomical research has documented that ‘normal’ readers have a larger planum temporale on the left, compared to the right hemisphere (Geschwind, 1979; Geschwind & Galaburda, 1987; Geschwind & Galaburda, 1985; Geschwind & Levitsky, 1968). Conversely, students diagnosed
with dyslexia are more likely to have either symmetrical or rightward asymmetry of the planum temporale. As there is a growing consensus that phonological processing difficulty is the core deficit associated with dyslexia, it would appear that structural abnormalities associated with dyslexia (e.g., rightward asymmetry) might have functional correlates associated with processing the basic components of speech (Barry & De Bastiani, 1985; Bosman, van Leerdam, & de Gelder, 2000; Demonet et al., 1992; Hanley & Gard, 1995; Leonard et al., 2001; Petryshen, Kaplan, Liu, & Field, 2000; B. A. Shaywitz et al., in press; B. A. Shaywitz et al., 2002; S. E. Shaywitz et al., 1998). Combined with the findings presented here, the implication would be that the structural abnormalities associated with phonological deficits might also negatively influence overall cognition, or intelligence. As these findings may well and truly be replicated in future studies, studies such as this one may in turn have an impact on the field of education and the way interventions are developed and delivered.

The implications from the apparent validity of the hypothesized model of intelligence presented here combined with those from the field of neuroscience should lead to practical changes in the field of education. One of the most practical changes that the field of education may make is the way children qualify for academic services. Using simple cutoff scores on standardized measures may prove to be a better method of serving students in need (Mercer et al., 1996; Peterson & Shinn, 2002; Proctor & Prevatt, 2003; Siegel, 1999, 2003). For example, as the link between structure and function becomes clearer, it might be more practical to deliver remedial services to any child struggling with a particular cognitive skill (e.g., phonological processing). Recent fMRI research hints at the success of this approach especially combined with specific, empirically validated interventions (B. A. Shaywitz et al., in press). Recent research has demonstrated that with a target, intensive, and empirically validated intervention, it may be
possible to actually alter the function of a particular cognitive structure known to support the cognitive deficit in question (B. A. Shaywitz et al., in press). Armed with such knowledge, policy makers may eventually change the psychometric and financial landscape associated with special education.

Carried to an administrative level, the eventual implications of future studies reporting findings consistent with those reported here may help shape future education policy. Currently, the discrepancy model determines whether a child receives academic services, and research suggests that it is not meeting the needs of students or educators alike. Clearly there is a need to employ another method to determine which students should receive academic services. If the findings reported here were extrapolated, one might see that discrete cognitive functions influence overall intellectual development. Furthermore, those structures carrying out those discrete cognitive skills may prove sensitive to intensive and proven interventions – thus altering their native function, and in essence enhancing the individual’s overall intelligence (B. A. Shaywitz et al., in press). Therefore, it would behoove policy makers to enact improved legislation that would see no child left behind in their quest for academic mastery because there was an insufficient difference between two aspects of cognition, when both contribute to that elusive attribute called intelligence. After all, no matter how cognition is conceptualized, only one organ’s function is assessed – the brain – and all intervention efforts are geared to improve the life of one individual at a time.
References


Wiederholt, J. L. (1974). Historical perspectives on the education of the learning disabled. In L. Mann & D. A. Sabatino (Eds.), *The second review of special education* (pp. 103-152). Austin, TX: PRO-ED.


<table>
<thead>
<tr>
<th>Author (date)</th>
<th>Sample size</th>
<th>Measures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colarusso (2001)</td>
<td>6036 eligibility reports</td>
<td>No discrepancy method balanced the racial makeup of LD programs. Different discrepancy models identified different groups of children.</td>
<td></td>
</tr>
<tr>
<td>D'Angiulli (2003)</td>
<td>121 normal (nor) 143 reading disabled (RD) 100 arithmetic disabled (AD)</td>
<td>WRAT WISC-R WRMT-R Gilmore Oral Reading Test VMI PPVT-R</td>
<td>Nor &gt; RD &amp; AD on all verbal tests All 3 groups demonstrated significant discrepancies between VIQ and PIQ in both directions 65% or more LD did not show typical patterns of performance A large % of normal did not have typical patterns of performance WISC-R patterns are not helpful in looking for LD</td>
</tr>
<tr>
<td>DiPerna (2001)</td>
<td>394 students 104 teachers 21 schools</td>
<td>Academic competence evaluation scale</td>
<td>There was support for a model of achievement that included academic enablers.</td>
</tr>
<tr>
<td>Diseth (2002)</td>
<td>89 Norwegian undergraduates</td>
<td>WAIS vocabulary Verbal analogies test Sandefjord Rybakoff spatial test Approaches and study skills inventory End of semester essay</td>
<td>Factor analysis returned a 1-factor solution of intelligence There is a negative correlation between surface approach and academic achievement Vocabulary correlates with the surface approach of studying A curvilinear relationship exists between surface studying and end of term essay grade There is not much support for the relationship between IQ and achievement</td>
</tr>
<tr>
<td>Kershner (1990)</td>
<td>19 boys 6 girls</td>
<td>WISC-R Cooper SEI WRAT TOWL</td>
<td>IQ is not related to learning ability Individual self concept predicts academic achievement Offer a model of self concept as related to achievement</td>
</tr>
<tr>
<td>Majoribanks (1976)</td>
<td>400 children (age 12)</td>
<td>Verbal reasoning</td>
<td>Verbal reasoning is a better predictor of achievement than nonverbal reasoning</td>
</tr>
<tr>
<td>Author (date)</td>
<td>Sample size</td>
<td>Measures</td>
<td>Results</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>McBride-Chang (1995)</td>
<td>136 3rd &amp; 4th grade students</td>
<td>WISC-III</td>
<td>Phonological awareness is composed of general intelligence, verbal memory, and speech perception</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phoneme deletion</td>
<td>Need more research to determine what exactly is phonological awareness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phoneme position</td>
<td>Need to work on the interplay between speech perception and phonological awareness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Phoneme segmentation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Memory for words</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Memory for rhyming words</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Memory for nonsense rhyming words</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Digit span</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech perception [bath-path]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech perception {slit-split}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Speech perception [ba-wa]</td>
<td></td>
</tr>
<tr>
<td>Murray (2003)</td>
<td>84 students with LD</td>
<td>College GPA</td>
<td>Most cognitive and achievement measures were not correlated with GPA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nelson Denny Reading</td>
<td>Most cognitive and achievement measures were correlated with each other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WRAT-R</td>
<td>FSIQ and delay/avoidance predicted GPA - but only 14%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WJ-R</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WAIS-R</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study habits questionnaire</td>
<td></td>
</tr>
<tr>
<td>Author (date)</td>
<td>Sample size</td>
<td>Measures</td>
<td>Results</td>
</tr>
<tr>
<td>--------------------</td>
<td>-------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>O'Malley (2002)</td>
<td>3 cohorts</td>
<td>Letter flash cards</td>
<td>Linear growth in VMI and phonemic awareness</td>
</tr>
<tr>
<td></td>
<td>N=379</td>
<td>CTOPP</td>
<td>All other variables had curvilinear growth</td>
</tr>
<tr>
<td></td>
<td>1=156</td>
<td>Rapid automatized naming</td>
<td>No evidence that IQ-ach discrepancy and low achievers are different groups</td>
</tr>
<tr>
<td></td>
<td>2=132</td>
<td>Recognition discrimination task</td>
<td>Non-impaired individuals performed better in 1st grade</td>
</tr>
<tr>
<td></td>
<td>3=91</td>
<td>VMI</td>
<td>Low achievers and the IQ-discrepant group had different trajectories (nonsig) but maybe significant with larger sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Word flash cards</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Spelling</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WJ-R</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WISC-R short form</td>
<td></td>
</tr>
<tr>
<td>Peterson (2002)</td>
<td>48 4th grades students with LD</td>
<td>WISC-II</td>
<td>The absolute achievement model correctly identified about 50% of students with LD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WJ-BRC</td>
<td>The RDM was accurate to 60%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R-CBM</td>
<td>The SDM was accurate to about 57%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The relative discrepancy model was accurate up to 95%</td>
</tr>
<tr>
<td>Proctor (2003)</td>
<td>170 university clinic cases</td>
<td>WAIS-III</td>
<td>Examined 4 LD models</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WJ-III (COG &amp; ACH)</td>
<td>Simple discrepancy model identified the most LD students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>COG-ACH discrepancy identified the least LD students</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Highest concordance was between intra-individual differences &amp; COG-ACH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Not all models identified that same kids</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No model is truly better than another</td>
</tr>
<tr>
<td>Remschmidt (1999)</td>
<td>32 children</td>
<td>Culture Fair intelligence test</td>
<td>Children with spelling and reading problems had a high level of first degree relatives with problems</td>
</tr>
<tr>
<td></td>
<td>31 moms</td>
<td>Grade appropriate spelling test</td>
<td>Low achievers typically had large IQ-achievement splits</td>
</tr>
<tr>
<td></td>
<td>26 dad</td>
<td>Grade appropriate reading comprehension</td>
<td>No support of 2 types of LD children</td>
</tr>
<tr>
<td>Author (date)</td>
<td>Sample size</td>
<td>Measures</td>
<td>Results</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>12 sisters</td>
<td></td>
<td>Study habits questionnaire</td>
<td>Questionnaires can be used to identify spelling deficits in adults</td>
</tr>
<tr>
<td>14 brothers</td>
<td></td>
<td>Study attitudes questionnaire WISC-R German version</td>
<td>A small group of dyslexics had visual processing problems</td>
</tr>
<tr>
<td>Share (2003)</td>
<td>Age 5 = 991</td>
<td>WISC-R</td>
<td>Boys had lower IQ intercepts than girls</td>
</tr>
<tr>
<td></td>
<td>Age 7 = 954</td>
<td>Burt word reading test</td>
<td>A combined regression equation overestimates boys abilities and underestimates girls abilities</td>
</tr>
<tr>
<td></td>
<td>Age 9 = 955</td>
<td>Rutter's adversity index</td>
<td>Reading problems are equal in boys and girls</td>
</tr>
<tr>
<td></td>
<td>Age 11 = 925</td>
<td>IL test of psycholinguistic abilities (verbal comprehension)</td>
<td>Children with reading problems have lower VIQ and same PIQ</td>
</tr>
<tr>
<td></td>
<td>Age 13 = 850</td>
<td>IL test of psycholinguistic abilities (verbal expression) Dunedin articulation check Progressive achievement test Basic motor ability test Neurological exam Rey ALVT COWA Trail making test Grooved pegboard Rey-O</td>
<td>Both genders show the same pattern of deficits Gender effects are still seen if IQ isn't used</td>
</tr>
<tr>
<td>Speece (2002)</td>
<td>113 editors</td>
<td>19 item questionnaire</td>
<td>IQ is not relevant for reading achievement Phonemic awareness is important to assess Treatment validity is important IQ should only be used to rule out MR</td>
</tr>
<tr>
<td>Stage (2003)</td>
<td>73 boys</td>
<td>Letter writing</td>
<td>Phonology, orthographic skill and attention predicted response to intervention</td>
</tr>
<tr>
<td></td>
<td>55 girls</td>
<td>Forced choice word pairs Phoneme deletion Syllable deletion Rapid automatic naming</td>
<td>VIQ is not the only predictor of response to intervention VIQ - achievement discrepancy did not predict response to intervention</td>
</tr>
<tr>
<td>Sternberg (2001)</td>
<td>326 us students</td>
<td>Sternberg triarchic abilities test</td>
<td>Provides support for the structural validity of his test</td>
</tr>
<tr>
<td>Author (date)</td>
<td>Sample size</td>
<td>Measures</td>
<td>Results</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Tiu (2003)</td>
<td>61 kids with reading problems</td>
<td>WISC-III PIQ</td>
<td>Processing speed does not influence reading</td>
</tr>
<tr>
<td></td>
<td>63 kids with no problems</td>
<td>Finding A's</td>
<td>No interaction between decoding, listening comprehension or processing speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO perceptual speed test</td>
<td>PIQ has a significant effect on reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT listening comprehension</td>
<td>PIQ does not mediate decoding, listening comprehension or processing speed on reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT reading comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WRMT-R word attack, word ID</td>
<td></td>
</tr>
<tr>
<td>Ullstadius (2002)</td>
<td>9001 male conscripts Subjects from Sweden</td>
<td>8 vocabulary/language tests 5 spatial tests</td>
<td>Found that $g$ and $Gc$ both influence performance on vocabulary tests The influence was stronger for classical words [of Latin origins]</td>
</tr>
<tr>
<td>Ward (1999)</td>
<td>201 LD students</td>
<td>WISC-III</td>
<td>70% of the sample have a typical core profile for the WISC-III and WIAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT</td>
<td>31.2% = below average IQ &amp; achievement 35.5% = low IQ and underachievement Support for a multivariate approach to eligibility Many LD children have 'normal' profiles consistent with the WISC &amp; WIAT linking sample</td>
</tr>
</tbody>
</table>

- Analytical, Practical, Creative
- Processing speed
- 8 vocabulary/language tests
- 5 spatial tests
- WISC-III PIQ
- Finding A's
- WIAT listening comprehension
- WRMT-R word attack, word ID
- Subjects from Sweden
- WISC-III
- WIAT

**Table:**

<table>
<thead>
<tr>
<th>Author (date)</th>
<th>Sample size</th>
<th>Measures</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiu (2003)</td>
<td>61 kids with reading problems</td>
<td>WISC-III PIQ</td>
<td>Processing speed does not influence reading</td>
</tr>
<tr>
<td></td>
<td>63 kids with no problems</td>
<td>Finding A's</td>
<td>No interaction between decoding, listening comprehension or processing speed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CO perceptual speed test</td>
<td>PIQ has a significant effect on reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT listening comprehension</td>
<td>PIQ does not mediate decoding, listening comprehension or processing speed on reading</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT reading comprehension</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WRMT-R word attack, word ID</td>
<td></td>
</tr>
<tr>
<td>Ullstadius (2002)</td>
<td>9001 male conscripts Subjects from Sweden</td>
<td>8 vocabulary/language tests 5 spatial tests</td>
<td>Found that $g$ and $Gc$ both influence performance on vocabulary tests The influence was stronger for classical words [of Latin origins]</td>
</tr>
<tr>
<td>Ward (1999)</td>
<td>201 LD students</td>
<td>WISC-III</td>
<td>70% of the sample have a typical core profile for the WISC-III and WIAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WIAT</td>
<td>31.2% = below average IQ &amp; achievement 35.5% = low IQ and underachievement Support for a multivariate approach to eligibility Many LD children have 'normal' profiles consistent with the WISC &amp; WIAT linking sample</td>
</tr>
</tbody>
</table>

**Notes:**

- The table lists various studies and their methodologies, focusing on intelligence factors and their impact on reading and cognitive abilities.