**The Dissertation Committee for Jamison Elder Carrigan Certifies that this is the approved version of the following Dissertation:**

# Are the Effects of *g* on Achievement Smaller at Higher Ability Levels?

**Committee:**

Timothy Z. Keith, Co-Supervisor

David A. Klingbeil, Co-Supervisor

Tiffany A. Whittaker

Katie Hartmann

# Are the Effects of *g* on Achievement Smaller at Higher Ability Levels?

**by**

**Jamison Elder Carrigan**

**Dissertation**

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

**Doctor of Philosophy**

**The University of Texas at Austin**

**May 2023**

## Dedication

This dissertation is dedicated to my loving and supportive wife Kristin and to my family, without whom this paper and this degree would not have been possible.

# Acknowledgements

Many thanks to the members of my committee for their patience and support throughout my graduate studies. I also want to thank the members of Dr. Keith's lab group, especially Jackie Caemmerer and Matt Reynolds, upon whose research the present study is built. A special acknowledgment to Pearson for providing the data used in the present study.

# Abstract

## Are the Effects of *g* on Achievement Smaller at Higher Ability Levels?

Jamison Elder Carrigan, Ph.D.

The University of Texas at Austin, 2022

Supervisors:

Timothy Z. Keith

David A. Klingbeil

Spearman's Law of Diminishing Returns (SLODR) posits that the contribution of a general intelligence factor (*g*) is lower at higher ability levels in explaining individual differences in cognitive ability test scores. This phenomenon has been supported across a variety of methodologies. The present study sought first to model SLODR using factor mixture modeling wherein high- and low-ability classes were created using latent modeling instead of through a priori group selection. The linking sample of the WISC-V and WIAT-III was used ($n$ = 181). This approach was supported in the present study.

Known classes were then generated based on this latent class model; these known classes were used to model the relation between intelligence and various achievement domains. As an extension of the SLODR phenomenon, it was hypothesized that the loadings of each achievement domain onto *g* would be lower in the high-ability group. This hypothesis was based on the premise that SLODR extends to domains that are highly correlated with intelligence such as achievement. Because SLODR posits that a given battery's subtests are more intercorrelated in lower-ability groups, it would therefore follow that the correlation (and therefore the factor loading) between *g* and achievement domains in reading, mathematics, and writing similarly increase in a lower-ability group. In most cases, this hypothesis was supported; the standardized loading of the achievement domain onto *g* was lower in the high-ability class in seven of the nine measured achievement domains. These findings suggest that the relation between intelligence and achievement is not static across the ability spectrum as has previously been assumed. Further, this study suggests that psychologists conducting psychoeducational assessment should use a three-stratum model of intelligence in which broad abilities are also analyzed, especially for individuals with higher cognitive ability. Future research is needed to assess this hypothesis across batteries and methodologies.

# Table of Contents

## List of Tables

# List of Figures

# CHAPTER 1: INTRODUCTION

Test scores on cognitive measures are positively correlated with one another; a general intelligence factor (*g*) is commonly cited as the primary reason for these correlations. The presence of *g* is useful in explaining individual differences in cognitive test performance (Bartholomew, 2004). Much of the research on *g* has involved assessing its influence on real-world correlates, such as academic achievement, job performance, and even health outcomes (Gottfredson, 1997; Herrnstein & Murray, 1994). Some researchers and theorists have argued that *g* operates in a consistent manner across the ability spectrum (e.g., Wolfle, 1940); however, Spearman's Law of Diminishing Returns (SLODR) states that the contribution of *g* is lower at higher ability levels in explaining individual differences in cognitive ability test scores (Detterman & Daniel, 1989; Spearman, 1927).

Spearman's (1927) original discovery of this phenomenon took place when he noted that the intercorrelations among cognitive ability tests were larger in lower-ability children than in average-ability children. The phenomenon went largely unnoticed and unstudied until Detterman and Daniel (1989) "rediscovered" SLODR using methods similar to Spearman's original study. Since this study, SLODR has been supported in a variety of ways, including comparison of cognitive ability task intercorrelations between ability groups (Legree et al., 1996), or by comparing *g* factor (or first principal component) differences between ability groups (Deary et al., 2007; M. R. Reynolds & Keith, 2007). More recent research has used factor mixture modeling techniques to address the potential weakness of using a priori group division inherent in SLODR research (M. R. Reynolds et al., 2010).

In general, *g* acts as a strong predictor of academic achievement across a variety of domains (Caemmerer et al., 2018; S. B. Kaufman et al., 2012). General intelligence and broad achievement (that is, not domain-specific achievement) are correlated highly with one another, with some studies suggesting the correlation is higher than .8 (Deary et al., 2007; S. B. Kaufman et al., 2012), though this correlation may be lower among school-age children (Gustafsson & Balke, 1993). In higher-order models of intelligence, the effects of *g* are mediated via the broad abilities (Floyd et al., 2012; Hajovsky et al., 2014; Niileksela et al., 2016). Because of these indirect effects, *g* and the broad abilities are thought to influence achievement simultaneously. In studies that assess the influence of *g* on academic achievement with school-age populations, students' performance in the areas of reading, math, and writing are all tied to *g*, such that increases in *g* directly improve scores across domains (Caemmerer et al., 2018; Glutting et al., 2006; Niileksela et al., 2016; Taub et al., 2008). Furthermore, the influences of broad cognitive abilities (i.e., first-order factors) on academic achievement vary by domain, but generally explain additional variation in achievement outcomes above and beyond the influence of *g* (Caemmerer et al., 2018; Gustafsson & Balke, 1993).

SLODR states that the average intercorrelation among tasks of cognitive ability is smaller at higher ability levels; it follows, therefore, that the correlations between one test and another test should also be lower for high-ability individuals. Predictive validity is measured using the correlation of a test with another test; therefore, the predictive validity of one construct (i.e., intelligence) on a related construct (i.e., achievement) should be lower for high-ability individuals. Because intelligence and achievement are so strongly correlated, both at the broad achievement level and at domain-specific levels, SLODR also suggests that *g* should be less predictive of achievement at higher ability levels. Few researchers have addressed this possible extension of SLODR; results have been mixed, but potential issues with construct measurement and methodology cast some doubt on the definitiveness of these studies (Coyle et al., 2011; McGill, 2015). The present study sought to study the hypothesis that *g* has smaller effects on academic achievement at higher ability levels.

**PURPOSE OF THE STUDY**

The possibility that *g* is less predictive of academic achievement at higher ability levels is important for two main reasons. First, SLODR is a well-studied topic but little is known about its implications beyond the scope of cognitive ability. Most research about the phenomenon has not attempted to make connections between general intelligence and its correlates; academic achievement is perhaps the most well-understood correlate of intelligence. This shortcoming may be due to a lack of access to a linking sample between cognitive and achievement tests for researchers to use; the present study used the WISC-V/WIAT-III linking sample to address this hypothesis.

This study is also important as it has potential practical implications. School psychologists and other practitioners are often asked to explain discrepancies between intelligence and achievement test scores, especially in districts where a discrepancy model or patterns of strengths and weaknesses (PSW) model is used (Flanagan et al., 2011; Naglieri, 2011). If intelligence has variable effects on the achievement scores of students based on overall level of intelligence, then these models would need to be re-evaluated considering those findings. Furthermore, given that the Individuals with Disabilities Education Act (2004) states that cognitive factors must be considered in the context of a complete psychoeducational assessment, more research is needed to understand the contribution of cognitive factors on school-based outcomes like academic achievement.

### RESEARCH QUESTION AND HYPOTHESIS

The present study sought to answer the following research question: "Does a general intelligence factor (*g*) have smaller effects on achievement domains at higher ability levels?" To address this question, mixture modeling was used to create latent ability classes using the full cognitive model of the WISC-V. Those classes were then used to model differences in the relations between *g* and various domains of achievement. It was hypothesized that classes with higher *g* would have smaller factor loadings from each achievement domain onto *g*; if true, this finding would support the assumption of SLODR that *g* is less meaningful (and therefore less predictive) at higher ability levels.

# CHAPTER 2: LITERATURE REVIEW

The following literature review begins by briefly examining the history of intelligence theory, including an explanation of three-stratum theory. Then, the effects of a general intelligence factor (*g*) on reading, math, and writing will be discussed, as well as the effects of broad cognitive abilities on these achievement domains. Finally, theories associated with Spearman's Law of Diminishing Returns (SLODR) will be reviewed, as well as the methodological considerations of the topic. This review will culminate by discussing SLODR's implications on the predictive ability of *g* and how these implications informed the hypotheses for the present study.

## THEORIES OF INTELLIGENCE

Charles Spearman (1904) is credited as the first researcher to propose a general intelligence factor, saying that this general factor, or *g* for short, was "something analogous to an 'energy'," or a force that could be expended to complete various cognitive tasks. He demonstrated this by showing that tasks designed to measure different aspects of cognition (e.g., reasoning, memory, abstract thought) were highly correlated with one another; this phenomenon is known as positive manifold. These intercorrelations, according to Spearman, were the result of this "mental energy" (Bartholomew, 2004). Spearman additionally noted that "for the purpose of indicating the amount of *g* possessed by a person, any test will do just as well as any other, provided only that its correlation with *g* is equally high" (1927). This idea is referred to as the theorem of the indifference of the indicator, and is still commonly used to explain high correlations among intelligence tests (Jensen, 1992).

Spearman's (1904) development of a general intelligence factor is also notable to statisticians, as it also spawned the creation of factor analysis (Cudeck & MacCallum, 2007). The development of this technique is especially notable as it came shortly after Francis Galton (1889) had introduced the concept of regression and correlation coefficients. Factor analysis is a technique that reduces many measures into fewer measures by placing items that correlate highly with each other together into factors (Keith, 2019).

### Spearman's two-factor theory

Extending upon his notion of *g*, Spearman (1927) also proposed specific factors, or *s*, that were subsumed by *g*. He described *g* as the "amount of a general mental energy," while specific factors represented the "efficiency of specific mental engines." This approach to understanding intelligence, often called two-factor theory, can also be understood by dividing the total variance in tests of intelligence into variance shared across all tests (*g*) and variance unique to each subtest (*s*) (Kamphaus, 2009).

This extension of intelligence theory drew the attention of other researchers, leading to other publications related to specific factors (e.g., Holzinger & Harman, 1938; Kelley, 1928; Thurstone, 1938). Other researchers continued studying intelligence using different approaches, such as including *g* as a lower-order factor in their models (e.g., Burt, 1949) or not including *g* at all (e.g., Guilford, 1959).

### Cattell and Horn's Gf-Gc theory

The continued development of specific factor research, especially that of Thurstone (1938), was instrumental in the formation of Raymond Cattell's Gf-Gc theory (Cattell, 1943; Schneider & McGrew, 2012). This theory posits that general intelligence – what

Spearman understood to be a single factor – is better represented by a two-factor model comprised of fluid intelligence (Gf) and crystallized intelligence (Gc). Fluid intelligence, as it was initially defined, included those traits that were physiologically inherent, while crystallized intelligence was thought to include aspects of ability that were influenced by experience (Ackerman, 2003). These domains are still relevant in modern intelligence research.

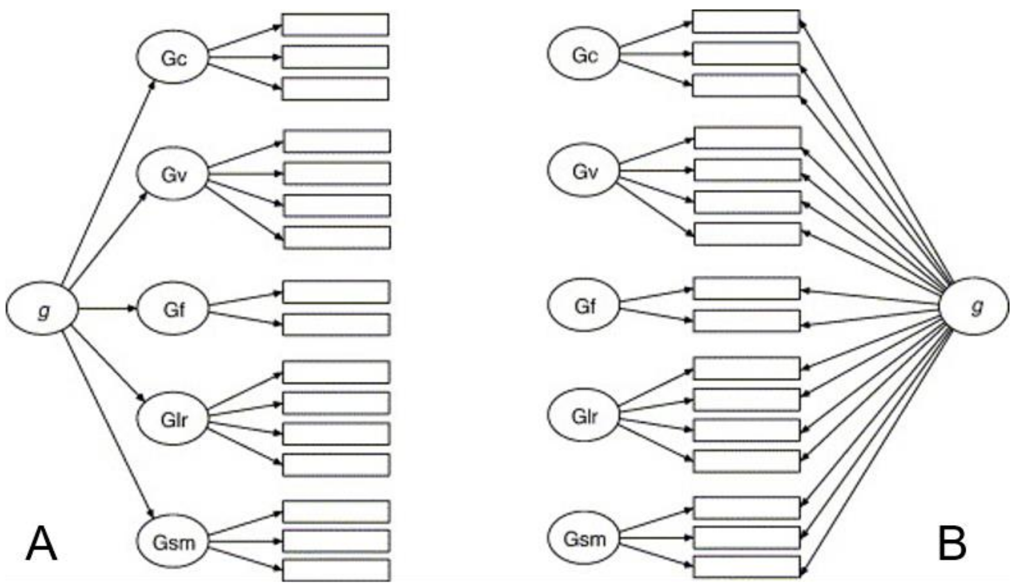
John Horn, Cattell's student, further expanded Gf-Gc theory by identifying approximately nine additional factors related to intelligence, including Gf and Gc, as well as visual intelligence (Gv), auditory intelligence (Ga), short term acquisition and retrieval (Gsm), long-term storage and retrieval (Glr), processing speed (Gs), quantitative and mathematical ability (Gq), and correct decision speed (CDS). In Gf-Gc models, these specific factors (or broad abilities) are all positively correlated with one another but lack a general intelligence factor (Horn & Noll, 1997).

**CHC theory**

John Carroll (1993) conducted a major factor analytic study in which he reanalyzed the results of more than 460 factor analytic datasets in the intelligence field spanning back to the advent of intelligence research and data collection. While the work was retrospective by nature, the result of the reanalysis was transformative to the modern understanding of intelligence: He ultimately proposed a unifying theory of intelligence, one that combined elements of Spearman's general factor and Cattell and Horn's Gf-Gc theories. This theory is variously referred to as the three-stratum theory of intelligence or Cattell-Horn-Carroll (CHC) theory (Schneider & McGrew, 2012).

In three-stratum theory, the model of intelligence consists of three strata: Spearman's *g* composes the third stratum; the second stratum is made up of approximately eight specific broad abilities; the first stratum includes several narrow abilities subsumed within each broad ability. These models are higher-order models, in which each stratum is subsumed by the preceding stratum (see Figure 1A for an illustration of this model). In this approach, *g* acts upon the narrow abilities (i.e., the subtests) indirectly via the broad abilities (Yung et al., 1999).

Occasionally, CHC theory is modeled using a nested or bifactor approach, in which *g* acts on the narrow abilities (i.e., at the subtest level) directly instead of indirectly through the broad abilities (Gustafsson & Balke, 1993; Mulaik & Quartetti, 1997). This model is less constrained than a higher-order model. However, some researchers have argued that a bifactor model of intelligence does not align with the theoretical framework of CHC theory (Keith & Reynolds, 2012; Murray & Johnson, 2013). An example illustration of this model can be found in Figure 1B.




*Note:* Adapted from Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence, 62*, 31-47.

Figure 1. Two hierarchical conceptions of intelligence: (A) the higher-order model and (B) the bifactor model

A higher-order approach to the CHC framework remains the most supported intelligence theory and is the one most often used in test development and interpretation; indeed, tests of intelligence that were created without CHC theory in mind still often conform to the CHC framework (Keith & Reynolds, 2010). Generally, CHC theory appears to be a useful explanation for a variety of cognitive processes, even those not necessarily captured in traditional intelligence tests; for example, measures of executive functioning, traditionally used in neuropsychological research and assessment, can be understood using CHC theory (Floyd et al., 2010; Jewsbury et al., 2016). Definitions of the five broad abilities used in the present study are found in Table 1; broad ability definitions in this table are based on definitions represented in Schneider and McGrew (2012).

| Broad Ability | Description |
|---|---|
| Fluid Reasoning (*Gf*) | The deliberate but flexible control of attention to solve novel, on-the-spot problems that cannot be performed by relying exclusively on previously learned habits, schemas, and scripts |
| Verbal Comprehension (*Gc*) | The depth and breadth of knowledge and skills that are valued by one's culture |
| Visual Processing (*Gv*) | The ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems |
| Short-Term Memory (*Gsm*) | The ability to encode, maintain, and manipulate information in one's immediate awareness |
| Processing Speed (*Gs*) | The ability to perform simple, repetitive cognitive tasks quickly and fluently |

Table 1. CHC Broad Abilities and Descriptions.

*Note:* Adapted from "The Cattell-Horn-Carroll model of intelligence", by W.J. Schneider & K.S. McGrew in *Contemporary intellectual assessment: Theories, tests, and issues*, 2012, New York: Guilford Press (pp. 99-112).

### INTELLIGENCE AND ACHIEVEMENT

In general, *g* acts as a strong predictor of academic achievement across a variety of domains (Caemmerer et al., 2018; S. B. Kaufman et al., 2012). Estimates of the correlation between general intelligence and broad achievement (that is, not domain-specific achievement) have exceeded .8 in some studies (Deary et al., 2007; S. B. Kaufman et al., 2012), though the correlation may be lower at school-age (Gustafsson & Balke, 1993). Higher-order CHC models that incorporate a second-order general intelligence factor and first-order broad abilities are often used in cognitive-achievement research, though the effects of *g* in these models is mediated via the broad abilities (Floyd et al., 2012; Hajovsky et al., 2014; Niileksela et al., 2016). Because of these indirect effects, *g* and the broad abilities are thought to influence achievement simultaneously.

Some analyses of cognitive-achievement relations include *g* as the lone predictor of academic achievement. In these studies, students' performance in the areas of reading, math, and writing are all tied to *g*, such that increases in *g* directly improve scores across domains (Glutting et al., 2006; Niileksela et al., 2016; Taub et al., 2008). In one such study, Caemmerer et al. (2018) included *g* direct effect only models in their analyses and found that for reading domains, *g*-loadings were between .51 and .61; for math domains, *g*-loadings were between .58 and .76; for writing domains, *g*-loadings were between .42 and .68. However, as was noted by the authors of this study, a common underlying issue when attempting to measure the effects of *g* on achievement is that the effects of *g* and fluid reasoning (Gf) are often indistinguishable and inseparable (Gustafsson, 1984; M. R. Reynolds & Keith, 2017). Consequently, this issue means that in the *g* direct-only models, it was impossible to determine whether the *g* effects were separate from Gf (Caemmerer et al., 2018).

**Effects of broad abilities on achievement domains**

Examining the effects of general intelligence on broad achievement, though informative on a grand scale, does not provide much utility for those interested in more focused domains like reading, math, and writing. Instead, cognitive-achievement research often focuses on the broad ability level in order to explain variance in these domains. In fact, research has shown that the effects of the broad cognitive abilities are significant above and beyond the influence of *g* (Gustafsson & Balke, 1993).

The present study addresses the five broad abilities measured by the Wechsler Intelligence Scale for Children, Fifth Edition (Wechsler, 2014): fluid reasoning (Gf), verbal comprehension (Gc), visual processing (Gv), short-term memory (Gsm), and processing speed (Gs). Table 1 provides definitions for these broad abilities (Schneider & McGrew, 2012). The following subsections review the effects of these broad abilities on the achievement domains measured by the Wechsler Individual Achievement Test, Third Edition (Wechsler, 2009). By assessing the effects of *g* on these achievement domains (rather than on broad achievement), the results will also more closely align with the categories of Specific Learning Disability outlined by IDEA (2004). Descriptions of the WIAT-III subtests can be found in Table 2.

***Basic reading skills***

The primary components of basic reading skills are decoding and word recognition skills. On the WIAT-III, basic reading skills are measured by two subtests, Pseudoword Decoding and Word Reading. Basic reading skills appears to be influenced by a variety of broad abilities; these abilities vary in their influence by age (Floyd et al., 2012; Niileksela et al., 2016). Perhaps unsurprisingly, verbal comprehension (Gc) demonstrates strong relations with basic reading skills (Cormier et al., 2017; J. J. Evans et al., 2002; Floyd et al., 2007; Garcia & Stafford, 2000; Hajovsky et al., 2014; McGrew, 1993; McGrew & Wendling, 2010; Niileksela et al., 2016; Oh et al., 2004). This finding makes conceptual sense, as language and vocabulary development are both central to Gc and required for acquiring reading skills (McGrew & Wendling, 2010). The effects of Gc on basic reading skills may be stronger in later childhood and adolescence, gradually increasing over time (Cormier et al., 2017; Hajovsky et al., 2014; McGrew, 1993). The effects of short-term memory (Gsm) on basic reading skills may also increase with age (Hammill, 2004;

McGrew & Wendling, 2010). Conversely, processing speed (Gs) appears to decline in importance as age increases for predicting basic reading skills (Cormier et al., 2017; J. J. Evans et al., 2002; Floyd et al., 2007; McGrew, 1993; Niileksela et al., 2016). Processing speed may be important early in the acquisition of reading skills as it contributes to automaticity and rapid automatic naming, skills that are foundational in early reading ability (McGrew & Wendling, 2010). Fluid reasoning (Gf) and visual processing (Gv) have not demonstrated consistent significant effects on basic reading skills (McGrew & Wendling, 2010).

***Reading comprehension***

Reading comprehension involves constructing meaning from text. On the WIAT-III, this domain is measured by the Reading Comprehension subtest. As with basic reading skills, verbal comprehension (Gc) is the strongest broad ability predictor of reading comprehension across school age (Benson, 2007; Caemmerer et al., 2018; J. J. Evans et al., 2002; Floyd et al., 2012; Keith, 1999; McGrew, 1993; McGrew et al., 1997; McGrew & Wendling, 2010; Niileksela et al., 2016; Oh et al., 2004; Vanderwood et al., 2002). These effects increase with age, as well (Cormier et al., 2017; Floyd et al., 2012; Hajovsky et al., 2014; Keith, 1999; McGrew, 1993). Furthermore, higher Gc increase the rate of growth in reading comprehension skills over time, as evidenced by multiple longitudinal studies of comprehension growth (Quinn et al., 2015; M. R. Reynolds & Turek, 2012).

Other broad abilities are less consistently attached to reading comprehension; these findings often depend on the measure used to compare ability-achievement relations (Caemmerer et al., 2018; Niileksela et al., 2016). Processing speed (Gs) and short-term memory (Gsm) have demonstrated inconsistent relations with reading comprehension; studies of these relations also vary on how age affects this pattern (Cain et al., 2004; J. J. Evans et al., 2002; Floyd et al., 2012; Hajovsky et al., 2014; McGrew, 1993). Fluid reasoning (Gf) has demonstrated some relation with reading comprehension in studies that use the Woodcock-Johnson co-norming sample (Cormier et al., 2017; J. J. Evans et al., 2002; McGrew, 1993; Niileksela et al., 2016), while a significant effect of visual processing (Gv) has been demonstrated using the Kaufman battery of tests (Hajovsky et al., 2014). However, these findings appear to be less consistent across cognitive assessments.

### *Reading fluency*

Reading fluency is the rate, accuracy, and prosody with which text is read and understood. On the WIAT-III, this domain is measured by the Oral Reading Fluency subtest. Less is known about the broad ability relations with reading fluency, but both verbal comprehension (Gc) and processing speed (Gs) have demonstrated significant effects at school age (Benson, 2007; Cormier et al., 2017; Niileksela et al., 2016). Other researchers have found relations between fluid reasoning (Gf) and reading fluency (Cormier et al., 2017), though these findings are inconsistent across studies (Caemmerer et al., 2018).

### *Basic math skills*

Basic math skills refer broadly to arithmetic and computational skills. On the WIAT-III, this domain is measured using the Numerical Operations subtest. Three broad abilities have demonstrated consistent effects on basic math skills across studies and measures: fluid reasoning (Gf), verbal comprehension (Gc), and processing speed (Caemmerer et al., 2018; Floyd et al., 2003; Keith, 1999; McGrew & Hessler, 1995; Niileksela et al., 2016). Fluid reasoning's connection to math achievement is well studied and makes conceptual sense: Gf involves the solving of novel problems and is less related to knowledge gained through experience (McGrew & Wendling, 2010; Schneider & McGrew, 2012). Language skills are, perhaps counterintuitively, related to broad math achievement (Floyd et al., 2003; McGrew & Wendling, 2010; Taub et al., 2008); this influence could be because Gc is important in the development of number concepts (Gelman & Butterworth, 2005) and math facts retrieval (Chong & Siegel, 2008). Other broad abilities have demonstrated less consistent effects on basic math skills, though there is some mixed evidence regarding short-term memory and its influence on math achievement (Floyd et al., 2003; McGrew & Hessler, 1995).

***Math problem solving***

Math problem solving, also referred to as math reasoning, involves the application of math concepts in complex (often text-based) problems (Caemmerer et al., 2018; McGrew & Wendling, 2010). On the WIAT-III, this domain is measured by the Math Problem Solving subtest. Much like basic math skills, math problem solving appears to be consistently influenced by fluid reasoning (Gf), verbal comprehension (Gc), and processing speed (Gs) across age groups (Caemmerer et al., 2018; Keith, 1999; McGrew & Hessler, 1995; McGrew & Wendling, 2010; Niileksela et al., 2016). In terms of growth across school age, Gf and Gc tend to increase as age increases (Floyd et al., 2003; McGrew & Hessler, 1995), while Gs is more stable across the age range (Floyd et al., 2003;

Niileksela et al., 2016). The conceptual connection between Gc and math problem solving is more obvious: Since this domain is often measured using word problems, aspects of reading ability are important for problem solving (McGrew & Wendling, 2010). Interestingly, a study that used the WISC-V and WIAT-III co-normed sample did not find a significant effect of Gc on either basic math skills or math problem solving (Caemmerer et al., 2018). This discrepancy may be due to differences between batteries: The WISC-V Figure Weights task, which is a Gf measure, may have accounted for the potential variance typically attributed to Gc in other batteries. Short-term memory (Gsm) and visual processing (Gv) are perhaps related to early-age math problem solving ability (Floyd et al., 2003; McGrew & Hessler, 1995), though evidence is mixed regarding the effects of these abilities at older ages (Niileksela et al., 2016).

***Math fluency***

Math fluency involves quickly solving mathematical operations. On the WIAT-III, this domain is measured by the Math Fluency subtest. As with reading fluency, little research exists exploring the relations between broad cognitive abilities and math fluency. Limited research supports the assumptions made by the CHC model: Processing speed (Gs) exerts a significant effect on math fluency, even when accounting for the effects of basic math skills (Niileksela et al., 2016). Math and reading fluency are less often studied in the CHC literature in part because the WJ measures only include tests of achievement fluency in composites with untimed tasks. Nevertheless, math and reading fluency deserve separate consideration apart from their broader domains since the patterns of cognitive-achievement relations are different (Caemmerer et al., 2018).

### *Basic writing skills*

Basic writing skills generally include spelling and editing. On the WIAT-III, this domain is measured by the Spelling subtest. Unlike reading and math, writing domains have received far less research attention in the cognitive abilities literature (Caemmerer et al., 2018). As may be predicted based on CHC theory, verbal comprehension (Gc) exerts consistently moderate to strong effects on basic writing skills (Cormier et al., 2016; Floyd et al., 2008); these effects appear to be stable throughout the lifespan (Niileksela et al., 2016). Processing speed (Gs) and short-term memory (Gsm) have also demonstrated moderate effects throughout school-age (Cormier et al., 2017; Floyd et al., 2008; Niileksela et al., 2016). The influence of fluid reasoning (Gf) appears to increase with age throughout the schooling years (Cormier et al., 2016; McGrew & Knopik, 1993); some researchers have suggested that Gf may not be significantly related to basic writing skills until late adolescence (Floyd et al., 2008).

### *Written expression*

Written expression involves the student's ability to express his or her ideas in writing. On the WIAT-III, two subtests measure this domain: Sentence Composition and Essay Composition; in the present study, though, these subtests were analyzed separately, since their correlation with one another is not strong (Caemmerer et al., 2018). Prior research on the effects of broad cognitive abilities on written expression have primarily been conducted using the WJ batteries; these tests include a writing fluency subtest as a part of the written expression factor. As a result, studies involving the WJ tests have all found that processing speed is a significant predictor of written expression (Cormier et al., 2016; Floyd et al., 2008; McGrew & Knopik, 1993; Niileksela et al., 2016).

Other broad ability influences of written expression also appear to be heavily dependent on the measure used. For example, on older versions of the WJ like the WJ Revised and WJ-III, short-term memory was a significant predictor of written expression (Floyd et al., 2008; McGrew & Knopik, 1993), but on research with the WJ-IV, these effects are much weaker or entirely nonsignificant (Cormier et al., 2016; Niileksela et al., 2016). This difference could be due to the fact that the Gsm factor on the WJ IV is more focused on working memory instead of short-term memory (Caemmerer et al., 2018).

Because broad ability influences of written expression appear to be more measure-dependent, understanding the Wechsler scales is especially important for the present study. The only study that has used the co-normed sample of the WISC-V and WIAT-III found that broad writing skills were significantly influenced by short-term memory; this finding supposes that the ability to store and manipulate information in mind (such as grammar and spelling rules) led to better spelling and composition scores (Caemmerer et al., 2018). Processing speed was also not significantly related to written expression, unlike previous studies of the WJ or Kaufman tests; this finding is likely because the Caemmerer et al. (2018) study included no writing fluency task in their measurement of broad writing ability. This study also highlighted a pattern of broad ability interactions: Gc influenced the Spelling subtest (i.e., basic writing skills) and Gf was more important for written expression subtests like Essay Composition; this same pattern has been demonstrated in other studies (e.g., Cormier et al., 2016), though Gf's influence on written expression is not wholly consistent across batteries (Niileksela et al., 2016). Overall, more cognitive-achievement research is needed using the Wechsler scales as the interactions between broad abilities and achievement domains appears to be highly dependent on the measures selected; in fact, cross-battery approaches show promise for alleviating these concerns (Caemmerer, 2017).

### THEORIES OF SPEARMAN'S LAW OF DIMINISHING RETURNS (SLODR)

Spearman's (1904) theory of general intelligence as a mental "energy" contributed to his proposal that *g* operated in the same way as in engine mechanics. He posited that the more *g* a person possessed (that is, more mental "energy"), the less advantage is added with constant, increased amounts. In order to illustrate this concept, he provided a set of correlation matrices of 12 cognitive ability tasks each (Spearman, 1927). The first matrix measured what he called "normal" children (i.e., children from a nonclinical sample); the second matrix included what he called "mentally defective" children (i.e., children from a clinical sample). The mean intercorrelation among the set of tasks for the "normal" children ($n$ = 78) was .47; for the "defective" children ($n$ = 22), the mean intercorrelation was .78.

78 NORMAL CHILDREN (Corrected for Attenuation).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Opposites- | — | 75 | 78 | 71 | 62 | 64 | 72 | 78 | 57 | 40 | 46 | 33 |
| 2. Observation | 75 | — | 72 | 58 | 60 | 58 | 67 | 56 | 58 | 56 | 52 | 29 |
| 3. Absurdities | 78 | 72 | — | 53 | 41 | 44 | 79 | 68 | 41 | 46 | 34 | 29 |
| 4. Memory sentences | 71 | 58 | 53 | — | 54 | 61 | 54 | 37 | 54 | 55 | 19 | 43 |
| 5. Crossing o's | 62 | 60 | 41 | 54 | — | 73 | 48 | 54 | 38 | 36 | 52 | 35 |
| 6. Geometrical figs. | 64 | 58 | 44 | 61 | 73 | — | 45 | 48 | 30 | 42 | 48 | 35 |
| 7. Discrim. length | 72 | 67 | 79 | 54 | 48 | 45 | — | 56 | 49 | 30 | 31 | 06 |
| 8. Crossing patterns | 78 | 56 | 68 | 37 | 54 | 48 | 56 | — | 30 | 21 | 27 | 18 |
| 9. Memory form | 57 | 58 | 41 | 54 | 38 | 30 | 49 | 30 | — | 24 | 31 | 29 |
| 10. Tapping | 40 | 56 | 46 | 55 | 36 | 42 | 30 | 21 | 24 | — | 29 | 18 |
| 11. Strength of grip | 46 | 52 | 34 | 19 | 52 | 48 | 31 | 27 | 31 | 29 | — | 28 |
| 12. Interpret. pictures | 33 | 29 | 29 | 43 | 35 | 35 | 06 | 18 | 29 | 18 | 28 | — |

Mean =0·466.

22 DEFECTIVE CHILDREN (Corrected for Attenuation).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Absurdities | — | 1·0 | 1·0 | 98 | 97 | 1·0 | 1·0 | 1·0 | 98 | 94 | 94 | 79 |
| 2. Opposites- | 1·0 | — | 97 | 95 | 87 | 91 | 85 | 76 | 85 | 87 | 70 | 72 |
| 3. Crossing patterns | 1·0 | 97 | — | 91 | 80 | 88 | 68 | 92 | 74 | 78 | 76 | 67 |
| 4. Crossing o's | 98 | 95 | 91 | — | 85 | 77 | 84 | 67 | 76 | 81 | 73 | 55 |
| 5. Memory sentences | 97 | 87 | 80 | 85 | — | 73 | 90 | 68 | 88 | 65 | 78 | 68 |
| 6. Observation | 1·0 | 91 | 88 | 77 | 73 | — | 76 | 83 | 71 | 86 | 59 | 65 |
| 7. Memory form | 1·0 | 85 | 68 | 84 | 90 | 76 | — | 65 | 67 | 70 | 77 | 75 |
| 8. Interpret. pictures | 1·0 | 76 | 92 | 67 | 68 | 83 | 65 | — | 74 | 80 | 80 | 59 |
| 9. Geometrical figs. | 98 | 85 | 74 | 76 | 88 | 71 | 67 | 74 | — | 65 | 60 | 62 |
| 10. Discrim. length | 94 | 87 | 78 | 81 | 65 | 86 | 70 | 80 | 65 | — | 51 | 45 |
| 11. Tapping | 94 | 70 | 76 | 73 | 78 | 59 | 77 | 80 | 60 | 51 | — | 61 |
| 12. Strength of grip | 79 | 72 | 67 | 55 | 68 | 65 | 75 | 59 | 62 | 45 | 61 | — |

Mean =0·782.

Figure 2. Spearman's (1927) original matrices illustrating his "law of diminishing returns"

Spearman's hypothesis went largely unstudied until Detterman and Daniel (1989) seemed to rediscover the phenomenon while analyzing the Wechsler Adult Intelligence Scale—Revised (WAIS-R) and the Wechsler Intelligence Scale for Children (WISC-R). For each of these measures, subtest intercorrelations were similarly related to ability group such that the intercorrelations were highest in the low ability groups, declining as *g* increased. For both the WAIS-R and WISC-R, these intercorrelations were twice as high in the low *g* groups than in the high *g* groups. However, this study noted that Spearman's

(1904) emphasis on "positive manifold" seemed to imply that the positive manifold was "uniformly distributed over the full range of ability," seemingly unaware of the fact that Spearman also hypothesized this phenomenon (Detterman & Daniel, 1989). In fact, Deary and Pagliari (1991) noted that indeed this finding was almost identical to one published decades earlier by Spearman. Detterman (1991) subsequently acknowledged that their finding was evidence of Spearman's Law of Diminishing Returns (SLODR). Regardless of the oversight, Detterman and Daniel (1989) understood the importance of their analysis, stating that "if the finding that correlations between mental tests vary systematically by level of ability is found to be a general one, not specific to certain tests, then the implications of this finding are substantial." Since this rediscovery, many researchers have replicated these findings using the same methodology (or one similar) on various test batteries (e.g., Deary et al., 1996; M. G. Evans, 1999; Legree et al., 1996). Since Detterman and Daniel's (1989) study, the term Spearman's Law of Diminishing Returns has come to be associated with the notion that the contribution of *g* in explaining variation in cognitive ability declines as an individual possesses more *g* (Jensen, 2003).

Anderson's (1992) theory of minimal cognitive architecture is a common conceptualization of the SLODR phenomenon in the literature. This theory states that a single general processing mechanism acts as a constraint on the effectiveness of more specific domains of cognition. This theory is useful for explaining why, for example, different cognitive domains are correlated (i.e., Spearman's "positive manifold"). In individuals whose general processing mechanism is more efficient, these domains are less constrained; this lack of constraint means that the various cognitive domains are less correlated with one another. This theory seems to map well onto Spearman's (1927) original conceptualization of the phenomenon: In the parlance of engine mechanics, *g* is like fuel for engines to complete domain-specific functions, but this fuel provides diminishing returns in engine efficiency as incremental increases are made. Detterman and Daniel (1989) also seemed to align with this conceptualization, saying:

> [If] central processes are deficient, they limit the efficiency of all other processes in the system. So all processes in subjects with deficits tend to operate at the same uniform level. However, subjects without deficits show much more variability across processes because they do not have deficits in important central processes. (p. 358)

An alternative framework for the SLODR phenomenon is Thurstone's (1938) sampling theory, which posits that low ability individuals have fewer cognitive resources on which to base their behaviors, thereby limiting their ability to express complex behaviors. In this conceptualization, higher intercorrelations among cognitive tasks in lower ability individuals is due to ability constraints. This theory is not necessarily incompatible with any of the previously mentioned theories; Spearman (1927), Detterman and Daniel (1989), and Anderson (1992) all suggest that, at lower ability levels, fewer cognitive resources constrain a wide range of different behaviors, leading to the statistical phenomenon of SLODR.

### METHODOLOGICAL CONSIDERATIONS WHEN STUDYING SLODR

Traditionally, research on SLODR has involved dividing samples into high and low ability groups and either comparing the intercorrelations of various cognitive tasks or the proportion of variance accounted for by a common *g* factor (Tucker-Drob, 2009). Detterman and Daniel's (1989) classic study unintentionally mirrored the method first used by Spearman (1927): Subtest intercorrelations were compared between intellectually disabled individuals and their same-age college-going peers, as well as between low and high IQ high school students, on the WAIS-R. The low IQ groups had substantially higher intercorrelations than the high IQ groups. Detterman and Daniel (1989) also tested for the same outcomes with more than two groups by dividing the standardization sample from the WAIS-R and the WISC-R into five ability groups. These groups were created by dividing individuals based on their Vocabulary subtest scores, and, as a replication, their Information subtest scores; the Full Scale IQ composite was not used for group division as that would have introduced spurious negative correlations among subtests (since those subtests contribute to the calculation of the FSIQ composite score). The Vocabulary and Information subtests were chosen because they had the highest correlations with FSIQ. The lowest ability groups' IQ equivalent was less than 78; for these groups, the average subtest intercorrelations were about .7. In the highest ability groups, where the IQ equivalent was greater than 122, these intercorrelations were about .4 (Detterman & Daniel, 1989).

Since the Detterman and Daniel (1989) study, this phenomenon has been studied using increasingly more sophisticated methodologies. For example, Deary et al. (1996) used principal component analysis to compare groups. Two similarly distributed subsamples of low and high IQ 14- to 17-year-olds were created from a larger sample of 10,500 school-age children; the first principal component accounted for about 2% less variance in the high IQ groups than the low IQ groups. Abad, Colom, Juan-Espinosa, and

García (2003) used a similar approach on a sample of 3,430 university applicants, as well as the Spanish standization sample of the WAIS-III (*n* = 823). In the university sample, the first principal component accounted for about 2% more variance in the low IQ groups, a finding that mirrored the Deary et al. (1996) study; in the WAIS-III sample, this difference was over 12%. Kane, Oakland, and Brand (2006) demonstrated an even larger variance gap: Using high and low IQ groups from the WJ-R standardization sample (*n* = 6,359, ages 2 to 95), prinicipal component analyses found that a single common factor accounted for 52% of the variance in the test scores for low IQ individuals and only 29% in the high IQ group—a difference of 23%.

Despite the popularity of principal component analysis for studying SLODR, the approach fails to take into consideration other influences of cognitive performance besides a single general factor (i.e., *g*); a CHC approach to intelligence research would suggest that the effects of broad cognitive abilities should also be considered in any conceptual model (Keith & Reynolds, 2012). Instead, a confirmatory factor analytic (CFA) approach to studying SLODR can be used; by doing so, the theoretical models of intelligence (which are hierarchical) match the analytic models (M. R. Reynolds et al., 2010). When using CFA to detect SLODR, second-order *g* factor variance is expected to be smaller as ability increases; this would be evidence that SLODR operated at the *g* level. If first-order factor variances also shrink as ability increases, SLODR would be demonstrated at the broad ability level, or across all factors (M. R. Reynolds et al., 2010). Furthermore, the *g* loadings (that is, the pathways from *g* to the various broad abilities) can be examined on a standardized CFA solution: If the *g* loadings decrease as ability increases, SLODR is operating at the level of *g*. Likewise, if the loadings of the subtests on the broad abilities decrease as ability increases, SLODR is operating at the broad ability level (M. R. Reynolds et al., 2010).

M.R. Reynolds and Keith (2007) were among the first to use CFA in SLODR research, thereby addressing the influence of the broad abilities. Using the standardization sample of the KABC-II (ages 6 to 18), high ($n$ = 594) and low ($n$ = 593) ability groups were selected based on Anderson-Rubin factor scores obtained from principal component analysis; this approach minimized selection effects. Then, CFA was used to fit two different hierarchical ability models, a higher-order model and a bifactor model (see Figure 1); because the Gf factor only had two indicators in the bifactor model (and was therefore underidentified), it was subsequently removed from the model (M. R. Reynolds & Keith, 2007).

In the higher-order model investigation, *g* explained more variance, on average, in the subtests for the low ability group (.16) than for the higher ability group (.07); this finding suggests that "*g* is less important in explaining individual variation in performance on cognitive tests for higher ability children," just as SLODR claims (M. R. Reynolds & Keith, 2007). However, the average proportion of variance explained in the subtests by the broad abilities did not vary significantly between the high and low ability groups (.23 and .22, respectively). Additionally, the average proportion of residual (that is, unexplained) variance was .62 for the low ability group and .70 for the high ability group. These findings suggest that the smaller amount of variance explained by *g* in the high ability group is not explained adequately by the broad abilities but instead by a greater proportion of unique variance than in the lower ability group. Put another way, SLODR appears to act on *g* and not on the broad abilities; however, M. R. Reynolds and Keith (2007) did note that the contribution of *g* to the various broad abilities was lower in the higher ability group. Furthermore, these findings were more or less replicated when using a bifactor model, suggesting that SLODR operates similarly for either hierarchical model of intelligence. The finding that SLODR appears to operate on *g* and not on the broad abilities is inconsistent with the findings in Carlstedt (2001), who found that, at higher ability levels, Gc and Gv explained more variance in ability. M. R. Reynolds and Keith (2007) posited that the inconsistency may be due to differing age ranges between the studies, the different number of groups used (Carlstedt, 2001, used more than two groups to divide ability), or because their study used multiple broad abilities as first-order factors.

The M. R. Reynolds and Keith (2007) study also assessed the higher-order model with all first-order intercepts invariant between the high and low ability groups (with the exception of Gc, as it was determined that the two ability groups differed in Gc beyond their difference in *g*); in doing so, they tested whether variances in *g* differed across ability groups. This model fit significantly worse than the specific factor means model, and the high ability group showed significantly lower variability in *g* than the low ability group. M. R. Reynolds and Keith (2007) state that this difference in variances shows that *g* is "more homogenous in the high ability group," just as SLODR would predict; this finding was replicated in the bifactor model, as well. Broad ability factor variances were then set to be invariant across groups, but the model fit was not significantly worse than the previous scalar invariant model (in which equality constraints were added to the subtest intercepts). However, when the *g* variances were constrained to be equal across groups, the model fit was significantly worse. Again, this finding suggests that SLODR operates on *g*, but not on the broad ability level. Furthermore, M. R. Reynolds and Keith (2007) suggest that SLODR is not a function of the magnitude of *g* loadings, since SLODR was not produced primarily by subtests with low *g* loadings in either hierarchical model.

Inherent in most SLODR research is the need to split the sample by ability level. This task has traditionally been accomplished by splitting the sample in half at the mean of an overall IQ score (e.g., Jensen, 2003), a subtest not included in the analysis (e.g,. Detterman & Daniel, 1989), or a score on the first principal component (e.g., M. R. Reynolds & Keith, 2007). These a priori group divisions are potentially problematic for three reasons well-documented by M. R. Reynolds et al. (2010): First, the decisions about the cut-point and the number of groups are often arbitrary and not directed by a particular model or theory. Second, by dichotomizing a continuous distribution, range restriction issues are likely, making it difficult to fully detect the presence of SLODR (Cohen, 1983; MacCallum et al., 2002). Third, groups that are formed using some researcher-selected cut-point are frequently assumed to be similar in terms of factor invariance, even when they may not be (Nesselroade & Thompson, 1995). This third limitation has been overcome in the past using multi-group mean and covariance structure analysis (MG-MACS) to test whether ability groups are similar on a construct level (Meredith, 1993; Nesselroade & Thompson, 1995; M. R. Reynolds & Keith, 2007). However, this approach still fails to address the first two limitations of a priori group formation. The practice of forming researcher-selected ability groups may indeed be impossible to do using theory, since there is no theory that dictates how many ability groups exist within the continuous *g* distribution (M. R. Reynolds et al., 2010).

In order to address the limitations of a priori group selection within a CFA framework, researchers have used factor mixture modeling (FMM; Bauer, 2005; M. R. Reynolds, 2008; M. R. Reynolds et al., 2010). FMM combines CFA and latent class modeling; in this approach, continuous latent factors and categorical classes are modeled (Dolan & van der Maas, 1998; Lubke & Muthén, 2005; Yung, 1997). Whereas MG-MACS still relies on the researcher to create groups to test for invariance, FMM creates probabilities of class (or cluster) membership for each individual in the dataset based on the overall substantive model; in other words, classes are formed from model probabilities (M. R. Reynolds et al., 2010). Traditionally, class membership is indicative of qualitative differences within a given sample; however, FMM can also be used indirectly to assess various mixture components (Lubke & Spies, 2008). The primary difference between these two uses of FMM is a matter of theory: In SLODR research, it is not assumed that multiple, heterogeneous classes exist within the overall sample, but rather that there is potential non-normality in the distributions of the latent variables (M. R. Reynolds et al., 2010).

Putting FMM methodologies to use, M. R. Reynolds et al. (2010) tested a set of hypotheses regarding the presence of SLODR using the standardization sample from the KABC-II. First, it was predicted that a single class model would have worse model fit than a model with more than one latent class. This was indeed the case, as a two-class model fit better than a one-class; a three-class model improved model fit further. Second, classes should differ in their latent *g* means and variances, and the class with the highest *g* mean should show smaller *g* variance; this was also the case in their study. Finally, standardized subtest residual variances should be larger in the higher ability group; this is because less subtest variance would be explained by *g*. Again, this proved to be the case in the M. R. Reynolds et al. (2010) study.

One potential confound of any investigation of SLODR is the influence of age differentiation (Facon, 2004; Kane & Brand, 2006). M. R. Reynolds et al. (2010) addressed this using FMM by including age as a covariate in order to explain differences between and within classes; in other words, the latent class variable was regressed on age (between classes), and *g* was regressed on age (within each class). However, they found that age was neither a significant predictor of class membership, nor that it explained a significant amount of variation in *g* within each class.

### IMPLICATIONS OF SLODR ON THE PREDICTIVE ABILITY OF *G*

Despite the increasingly sophisticated methods of investigating SLODR, the phenomenon still boils down to a central premise: The mean intercorrelation between tests is lower at higher ability levels. Given this, it would follow that the correlation between one test and another test should also be lower for high-ability individuals. Predictive Validity may be measured using the predictive correlation of a test with another test; therefore, the ability of one construct (i.e., intelligence) to predict a related construct (i.e., achievement) should be lower for high-ability individuals. Given the well-documented correlations between intelligence and achievement, SLODR seems to indicate that *g* would be less predictive of achievement at higher ability levels.

Little research has investigated whether this extension of SLODR is correct. One study has assessed whether scores on the SAT predicted college grade-point average (GPA) differently for high ability individuals (Coyle et al., 2011); this study divided participants based on scores on the Armed Services Vocational Aptitude Battery (ASVAB), a measure of *g*. Interestingly, the high ability group demonstrated a higher predictive validity of the SAT on GPA. These findings are counter to the SLODR hypothesis; however, the unexpected results may be attributable to several factors, including poor indicators of *g* such as the SAT, unexamined factorial invariance between ability groups, or potential age effects.

Another study has investigated the incremental predictive validity of the WJ-III COG broad cluster scores on WJ-III ACH scores (McGill, 2015). In this study, the SLODR hypothesis was supported: GIA (a general factor score) accounted for less variance in achievement in higher ability classes than in lower ability classes. Here, ability classes were simply derived from the descriptive ranges on the WJ-III COG (e.g., Below Average, Average, Above Average). This rudimentary way to create ability groups may suffer from the previously mentioned limitations of a priori group formation. Furthermore, this study did not attempt to build a higher-order factor analytic model of intelligence; rather, it relied on observed scores. In sum, further research that examines SLODR's hypothesis is needed as there is little extant research on the topic.

# CHAPTER 3: METHOD

## PARTICIPANTS

The sample for this study came from the WISC-V and WIAT-III co-norming sample (*n* = 181). Students ranged from 6 to 16 (*M* = 11.82, *SD* = 3.07) years old. The sample was 55% male; 50% of the sample was White, 21% was Hispanic/Latinx, 20% was African American/Black, 7% was “Other,” and 2% was Asian. The mean testing interval between administrations of the WISC-V and WIAT-III was 16 days. All individuals were administered all subtests of the WISC-V (with the exception of one individual who did not complete the Digit Span subtest). For the WIAT-III, the percentage of missingness for each subtest ranged from 0% to 16.6% for the Essay Composition subtest. More information about the numbers of individuals who were administered each subtest can be found in Table 4.

## MEASURES

### WISC-V

The WISC-V (Wechsler, 2014) is an individually administered test of intelligence for children and adolescents ages six through 16 years, 11 months old. Generally, the test has received broad support from the research community (e.g., Benson, 2017; Keith, 2017; M. R. Reynolds & Hadorn, 2016). The test was designed to be consistent with CHC theory and purports to measure five cognitive broad ability factors: Fluid Reasoning (Gf), Verbal Comprehension (Gc), Visual Spatial Reasoning (Gv), Processing Speed (Gs), and Working Memory (Gsm). Table 2 provides descriptions of each of the 16 subtests on the standard WISC-V battery, as well as their corresponding broad abilities. Reliability estimates for each subtest are reported in the test's technical manual and are generally quite high (Wechsler, 2014). Average test-retest reliability coefficients across the subtests ranged from .72 to .91 in the standardization sample with a mean interval of 26 days. Average split-half reliability coefficients across ages ranged from .80 to .96, evidence of internal consistency. Scores on the WISC-V are converted from raw scores to age-standardized scores for interpretation; the age-standardized scores were used in the present study.

**WIAT-III**

The WIAT-III (Wechsler, 2009) is an individually administered achievement test for ages four through 50. The present analysis focused on ten subtests that measure aspects of reading, math, and writing achievement; Table 2 provides descriptions of each of these subtests, as well as their corresponding achievement domain. The Essay Composition subtest was not administered to children under the age of 8 in the present sample. Reliability estimates are reported in the test's technical manual and are quite high (Breaux, 2010). Average test-retest coefficients were reported for Math Fluency, Oral Reading Fluency, Sentence Composition, and Essay Composition, and ranged between .87 and .95 (Breaux, 2010). Average split-half reliability estimates were reported for the other subtests and ranged from .88 to .97 (Breaux, 2010). Age-standardized achievement scores were used in the present study.

| **WISC-V** | | |
|---|---|---|
| Subtest | Task Description. The child: | Broad Ability |
| Similarities | Describes how two words or concepts are similar | Gc |
| Vocabulary | provides a definition for given words or names a picture | Gc |
| Information | Answers general knowledge questions | Gc |
| Comprehension | Answers questions based on general knowledge and social conventions | Gc |
| Block Design | Reproduces two-dimensional patterns using blocks in a specified amount of time | Gv |
| Visual Puzzles | Mentally manipulates images to form a complete picture | Gv |
| Matrix Reasoning | Selects an image that completes a picture with a missing portion | Gf |
| Figure Weights | Is given a key and selects a response that balances a scale | Gf |
| Picture Concepts | Selects from a set of pictures to create a group that has common characteristics | Gf |
| Arithmetic | Solves orally presented arithmetic problems in a specified amount of time (without use of pencil and paper) | Gsm |
| Digit Span | Completes two tasks: In Digit Span Forward, the child repeats back an increasingly long set of digits in the order the child heard them. In Digit Span Backward, the child repeats digits back in the reverse order presented by the examiner | Gsm |
| Picture Span | Is shown pictures and is asked to recall them in sequential order from a response page | Gsm |
| Letter-Number Sequencing | Listens to a set of numbers and letters and is then asked to repeat the numbers back from smallest to largest and the letter back in alphabetical order | Gsm |
| Coding | Uses a key to copy symbols that correspond to numbers in a specified amount of time | Gs |
| Symbol Search | Determines whether a specified symbol is present or absent in a group of other symbols in a specified amount of time | Gs |
| Cancellation | Is shown arrays of pictures and must select target symbols under timed conditions. | Gs |

| **WIAT-III** | | |
|---|---|---|
| Subtest | Task Description. The child: | Ach. Domain |
| Pseudoword Decoding | Is asked to sound out nonsense words | Basic reading skills |

| | | |
|---|---|---|
| Word Reading | Identifies letters, sounds, or words from a list | Basic reading skills |
| Reading Comprehension | Reads short passages and then is asked questions about them | Reading comprehension |
| Oral Reading Fluency | Must accurately read passages aloud in a specified amount of time | Reading fluency |
| Numerical Operations | Solves written math problems involving addition, subtraction, multiplication, and division | Basic math skills |
| Math Problem Solving | Solves orally presented math word problems that are often multistep and related to real-world concepts such as time or money | Math problem solving |
| Math Fluency | Answers simple addition, subtraction, and multiplication problems in a specified amount of time | Math fluency |
| Spelling | Is asked to spell orally presented words based on definitions and their use in a sentence | Basic writing skills |
| Sentence Composition | Builds sentences using target words and combines multiple sentences into a single sentence | Written expression (I) |
| Essay Composition | Writes words, sentences, or paragraphs essay in response to prompts | Written expression (II) |

*Note:* Adapted from Wechsler (2009) and Wechsler (2014).

Table 2. A description of the subtests on the WISC-V and WIAT-III.

## DATA ANALYSES

Prior to analysis, data were cleaned and prepared using the tidyverse suite of packages (Wickham et al., 2019) in R, version 4.0.2 (R Core Team, 2020). All subsequent analyses were conducted using Mplus, version 7.4 (Muthén & Muthén, 2012). Mplus is advantageous for testing structural equation models as it handles missing data through the Full Information Maximum Likelihood (FIML) procedure which is the recommended procedure for handling missing data (Enders & Bandalos, 2001; Schafer & Graham, 2002). In the present study, all 181 individuals in the sample completed the full WISC-V battery (with the exception of one missing score for the Digit Span subtest). Missingness on the

WIAT-III subtests ranged from 1.1% to 16.6%, though there was no missing data on the Numerical Operations, Math Problem Solving, or Spelling subtests. Table 4 provides the number of individuals who completed each subtest.

Analysis began by validating the WISC-V scoring model (M1; see Figure 3). This model included *g*, as well as the five CHC broad ability factors measured by the WISC-V. The Arithmetic subtest was cross-loaded onto both Gf and Gsm, as indicated in the test manual (Wechsler, 2014). Each broad ability factor was indicated by two to four subtests.

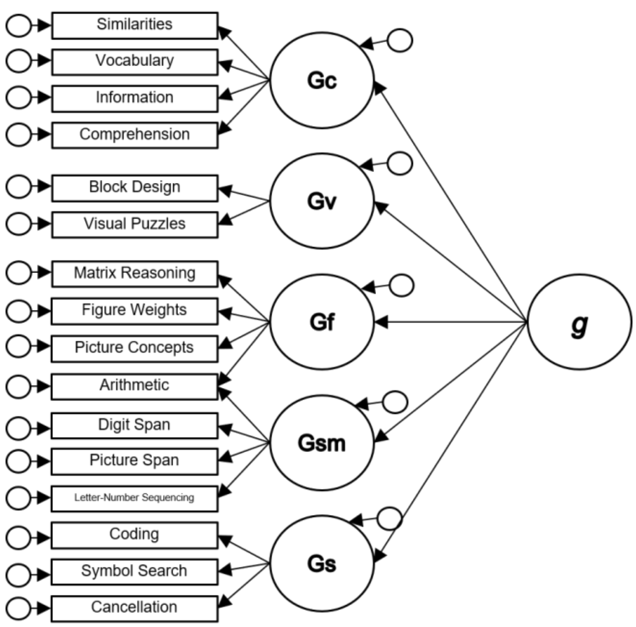



Figure 3. WISC-V validation model (M1)

Global fit of the WISC-V validation model was examined with two indexes: the root mean square error of approximation (RMSEA) and the comparative fit index (CFI).

RMSEA values around or less than .05 and CFI values around or greater than .95 are considered to indicate "good" global fit (Schermelleh-Engel et al., 2003).

A mixture CFA model was then tested by adding a latent class factor to the WISC-V validation model. Two-, three-, and four-class solutions were evaluated and compared to each other using relevant fit statistics and model parameter estimates. Fit indices used include the Akaike Information Criteria (Akaike, 1987), the Bayesian Information Criteria (Schwarz, 1978), and the adjusted Bayesian Information Criteria (aBIC), which is adjusted for sample size (Sclove, 1987). The aBIC tends to perform better in latent variable mixture models (Henson et al., 2007). For each of these fit statistics, a lower value indicates a better fitting model.

The adjusted likelihood ratio test (aLRT) was also used to compare nested models with increasing numbers of classes (Lo et al., 2001), although this test has yet to be evaluated for use with complex models (Bauer & Curran, 2003). The aLRT provides a *p*-value that specifies whether the present model shows a statistically significantly better fit than a model with one fewer class. Entropy was also evaluated as a measure of "fuzziness" between classes. Entropy values closer to one indicate that the model predicted class membership with more accuracy, while values closer to zero indicate that the model struggled to predict class membership.

Latent mean differences and latent variances were also interpreted across models. There is no theoretical difference between a two-, three-, or four-class model; rather, running models with iteratively more classes simply provided the ability to "replicate" the findings in models with varying numbers of classes. Nevertheless, model comparisons were made to determine which among the various mixture models of intelligence was the best fitting using the appropriate fit statistics (Nylund et al., 2007).

The process by which latent class models were created follows the procedure outlined in M. R. Reynolds (2008). The following latent class models were first performed using a two-class solution, then all steps were repeated for three- and four-class solutions. Figure 4 shows a basic illustration of the inclusion of a latent class factor.

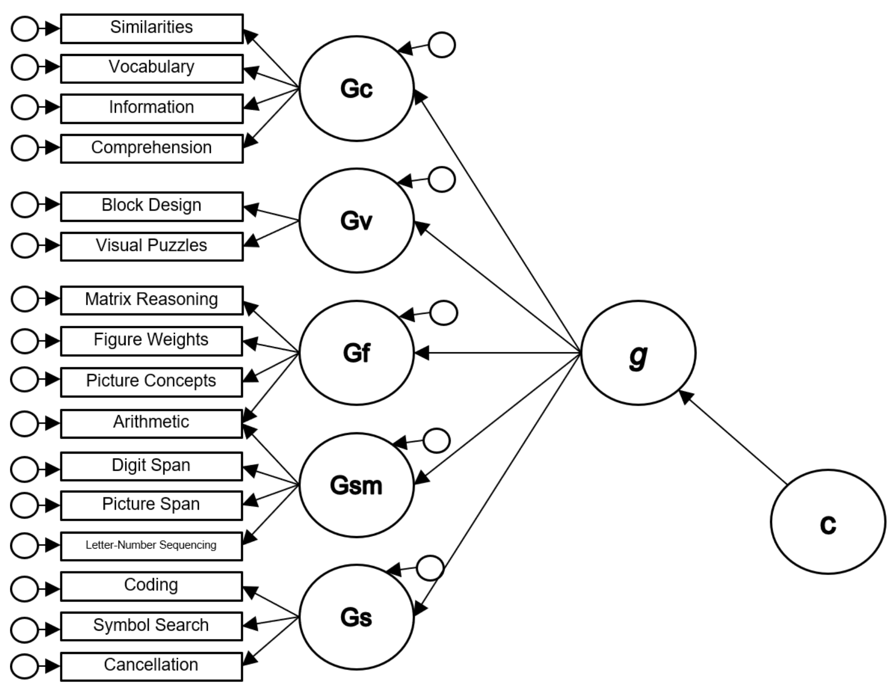



Figure 4. Cognitive model with more than one latent class factor specified (M2)

In the first latent class model (M2; model M1 was the WISC-V validation model with no latent classes), only the second-order *g* mean was freely estimated between classes, with all other model parameters specified as class-invariant. Therefore, the only difference between each class in this model was the level of *g*. In these models, latent means are not properly identified, so one class serves as the reference class with the latent *g* mean set to zero (Byrne & Stewart, 2006). Other class *g* means are therefore freely estimated, meaning their estimates represent the difference from the reference class *g* mean.

In the next latent class model (M3), the constraints on *g* variance were also freely estimated within each class, along with the free estimation of *g* means between classes. As noted in M. R. Reynolds (2008), this model is of particular importance as it allows classes to differ in both *g* means and *g* variances.

The final model (M4) built upon the previous model by releasing the constraints of subtest residual variances. A better fitting model here would indicate that the average standardized subtest residual variance varies between classes (Lubke & Muthén, 2005). In M. R. Reynolds (2008), the average standardized subtest residual variance was higher in the high ability classes, indicating that *g* explains less variance in the subtests. This model is of particular importance to this study because, if SLODR exists, model fit should be highest for this model.

M2-M4 were repeated using three- and four-class solutions, thereby creating M5-M7 for the three-class solutions and M8-M10 for the four-class solutions. Table 3 provides a brief summary of each of the models evaluated in the study.

Once all latent class models were created, fit statistics were compared to determine the best-fitting and most appropriate model upon which to create known classes; these classes were then used for subsequent models. These known classes were used to assess differences in the influence of *g* on various achievement outcomes.

| Model | # of classes | Description |
|---|---|---|
| M1 | 1 | WISC-V confirmatory factor model |
| M2 | 2 | *g* means differ |
| M3 | 2 | *g* means and *g* variances differ |
| M4 | 2 | *g* means, *g* variances, and subtest variances differ |
| M5 | 3 | *g* means differ |
| M6 | 3 | *g* means and *g* variances differ |
| M7 | 3 | *g* means, *g* variances, and subtest variances differ |
| M8 | 4 | *g* means differ |
| M9 | 4 | *g* means and *g* variances differ |
| M10 | 4 | *g* means, *g* variances, and subtest variances differ |

Table 3. A summary of latent class models used in the study.

Each of the ten WIAT-III subtests measures a different domain of reading, math, or writing achievement, with the exception of the Pseudoword Decoding and Word Reading subtests, which both appear to measure basic reading skills and have been strongly correlated in previous studies (Caemmerer et al., 2018). Each achievement domain was analyzed separately from other achievement domains. For each achievement model, the achievement domain was modeled as a latent factor; each of the domains was measured by a single subtest (other than basic reading skills). In order to model a single-indicator factor, the variance of the contributing subtest was constrained to the value of the subtest's unreliability (i.e., one minus the reliability as reported in the technical manual (Breaux, 2010), multiplied by its variance). This technique has been previously used in studies of the WISC-V and WIAT-III that have measured cognitive effects on achievement outcomes (Caemmerer et al., 2018) and helps avoid model under-identification (Hayduk, 1987; Keith, 2019). Table 2 contains descriptions of each WIAT-III subtest and its corresponding achievement domain. For each achievement outcome, a known-class categorical latent variable was constructed to assess whether class membership influenced the loading from *g* onto the latent achievement outcome. An example of this model is shown in Figure 5 using the Reading Comprehension subtest (and its corresponding achievement domain).

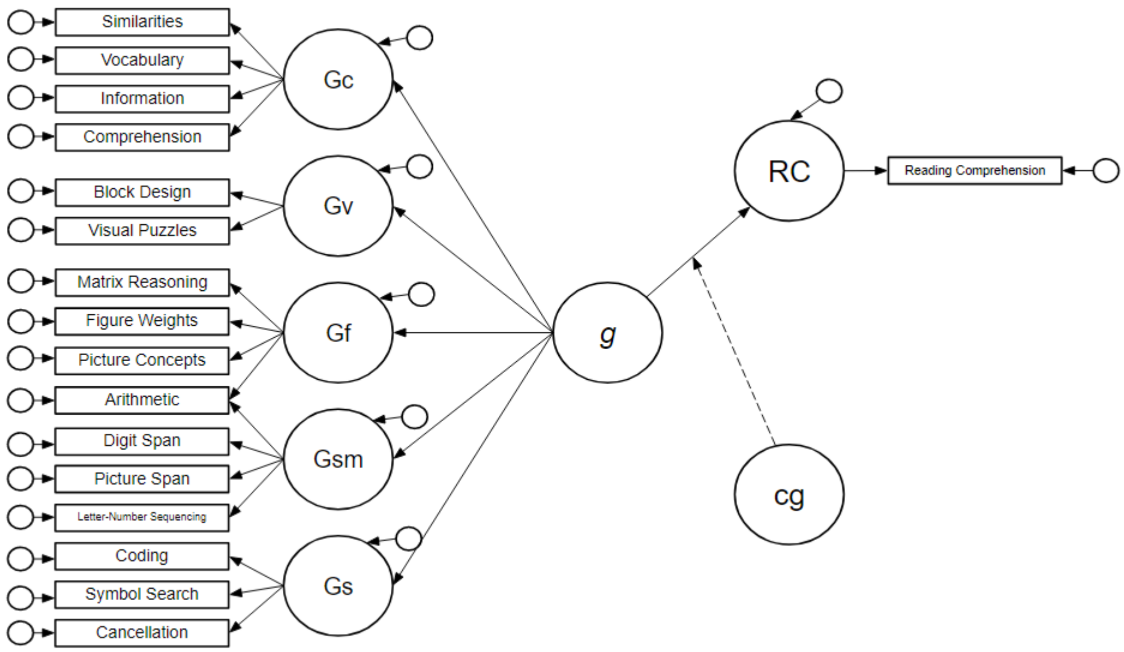



Figure 5. A cognitive-achievement model with a latent class categorical variable with known classes (cg)

For each of the achievement domains, a one-class validation model was created to measure the loading of *g* onto each achievement outcome. The two-class models were again compared to the one-class models using AIC, BIC, and aBIC, as well as the differences in the loading of *g* onto the achievement outcome. However, this step did not involve the selection of a better-fitting model.

# CHAPTER 4: RESULTS

## DESCRIPTIVE STATISTICS

Sample sizes, subtest minimum and maximum scores, means, standard deviations, skewness, and kurtosis estimates for the sample are presented in Table 4 for WISC-V and WIAT-III subtests. Skewness values were considered acceptable as all values fell within a range of -.50 to .50. Kurtosis values were also considered acceptable because none were 7.00 or larger; values this large can cause problems associated with nonnormality (Curran et al., 1996).

## WISC-V SCORING MODEL

The WISC-V scoring model (see Figure 3) was assessed for overall fit. This model fit the data well (RMSEA = .03, CFI = .99). Other fit indices also indicate adequate model fit and were used for comparison with latent class models ($\chi^2$ = 110.41, SRMR = .05, TLI = .98). Figure 4 illustrates the standardized first-order broad ability factor loadings on *g*. Fluid Reasoning demonstrated the strongest loading (.99), followed by Verbal Comprehension (.75). This pattern of results resembles findings from M. R. Reynolds and Keith (2017) in which the WISC-V standardization sample was used. This model appeared to reasonably explain the structure of the WISC-V in the present sample.

| Subtest | N | Minimum | Maximum | Mean | Standard Deviation | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| WISC-V | | | | | | | |
| Similarities | 181 | 2.00 | 16.00 | 9.93 | 2.47 | -0.03 | 2.95 |
| Vocabulary | 181 | 1.00 | 17.00 | 9.98 | 2.64 | -0.10 | 3.00 |
| Information | 181 | 3.00 | 17.00 | 9.96 | 2.53 | 0.30 | 2.83 |
| Comprehension | 181 | 4.00 | 17.00 | 10.26 | 2.55 | 0.21 | 2.35 |
| Block Design | 181 | 4.00 | 17.00 | 10.12 | 2.70 | 0.01 | 2.55 |
| Visual Puzzles | 181 | 4.00 | 17.00 | 10.08 | 2.72 | 0.02 | 2.32 |
| Matrix Reas. | 181 | 3.00 | 17.00 | 10.13 | 2.71 | 0.22 | 2.55 |
| Figure Weights | 181 | 4.00 | 16.00 | 9.90 | 2.68 | -0.26 | 2.87 |
| Pic. Concepts | 181 | 1.00 | 17.00 | 10.18 | 3.08 | -0.10 | 2.74 |
| Arithmetic | 181 | 4.00 | 17.00 | 10.12 | 2.65 | 0.03 | 2.65 |
| Digit Span | 180 | 4.00 | 19.00 | 9.98 | 2.68 | 0.38 | 3.23 |
| Picture Span | 181 | 4.00 | 15.00 | 9.71 | 2.62 | -0.03 | 2.29 |
| L.-N. Seq. | 181 | 4.00 | 16.00 | 10.00 | 2.70 | 0.15 | 2.51 |
| Coding | 181 | 3.00 | 19.00 | 10.01 | 2.76 | 0.17 | 2.99 |
| Symbol Search | 181 | 1.00 | 19.00 | 9.99 | 3.12 | 0.14 | 3.58 |
| Cancellation | 181 | 3.00 | 18.00 | 10.27 | 2.86 | 0.29 | 3.06 |
| WIAT-III | | | | | | | |
| Ps. Decoding | 175 | 58.00 | 133.00 | 101.37 | 12.56 | -0.26 | 3.40 |
| Word Reading | 176 | 63.00 | 127.00 | 101.03 | 11.28 | -0.51 | 3.24 |
| Reading Comp. | 178 | 58.00 | 148.00 | 102.15 | 13.00 | 0.41 | 4.62 |
| Oral Read. Flu. | 177 | 59.00 | 137.00 | 102.90 | 12.10 | -0.01 | 3.70 |
| Num. Oper. | 181 | 59.00 | 130.00 | 100.78 | 12.18 | -0.07 | 2.89 |
| M. Prob. Solv. | 181 | 70.00 | 133.00 | 100.76 | 12.04 | 0.18 | 2.56 |
| Math Fluency | 177 | 71.00 | 133.00 | 100.11 | 13.23 | 0.23 | 2.41 |
| Spelling | 181 | 65.00 | 127.00 | 100.20 | 12.90 | 0.00 | 2.39 |
| Sent. Comp. | 179 | 68.00 | 140.00 | 99.59 | 12.86 | -0.12 | 2.88 |
| Essay Comp. | 151 | 57.00 | 141.00 | 100.76 | 15.77 | -0.16 | 2.51 |

Table 4. Descriptive statistics for subtests in the overall sample.

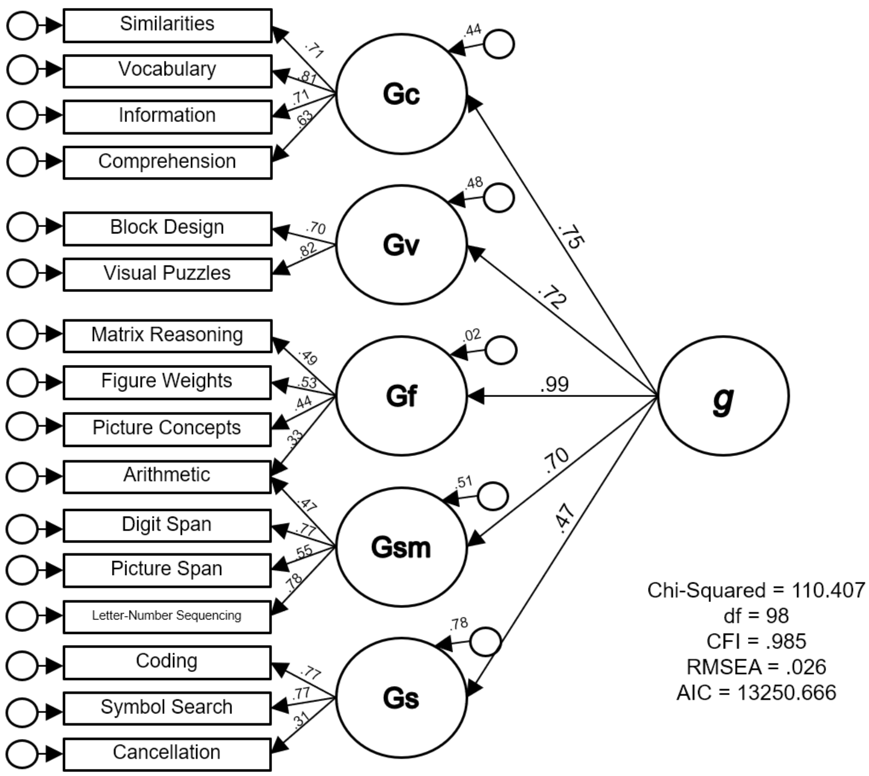



Figure 6. Standardized results of the WISC-V validation model

## Latent class models

Mixture CFA models were assessed with two-, three-, and four-class solutions. Table 3 describes each model used in the study. To evaluate these model solutions, information criteria were compared. Table 5 presents values for AIC, BIC, and aBIC for each of these models. Lower values are associated with better fitting models. In the event that information criteria provided contradictory conclusions, preference was given to the aBIC as it tends to perform better in latent variable mixture models (Henson et al., 2007). Furthermore, the AIC and BIC tend to perform less favorably in latent class models (Lin & Dayton, 1997). The AIC favors models with more latent classes while the BIC favors models with fewer latent classes (Yang, 1998). Table 5 also provides measures of aLRT and entropy.

**Two-class models**

Overall, two-class models fit the data better than did three- and four-class solutions across AIC, BIC, and aBIC. The AIC and aBIC favored M4, which freed constraints on *g* means and variances, as well as subtest variances; this model also best illustrated SLODR given its model constraints. None of the latent class models demonstrated a significant *p*-value for aLRT, indicating that no latent class model fit better than the same one with one fewer class. Furthermore, entropy values indicated that models with more classes tended to do a better job of differentiating individuals. As described below, though, M10 failed to separate individuals at all, leading to an entropy value of 1.00.

| **Model** | **# of classes** | **AIC** | **BIC** | **aBIC** | **aLRT (*p*)** | **Entropy** |
|---|---|---|---|---|---|---|
| M1 | 1 | 13,250.66 | **13,423.39** | 13,252.36 | N/A | N/A |
| M2 | 2 | 13,248.00 | 13,439.91 | 13,249.89 | .67 | .46 |
| M3 | 2 | 13,250.11 | 13,448.42 | 13,252.06 | .64 | .53 |
| M4 | 2 | **13,231.68** | 13,481.16 | **13,234.13** | .68 | .63 |
| M5 | 3 | 13,248.94 | 13,463.24 | 13,251.05 | .55 | .62 |
| M6 | 3 | 13,246.17 | 13,466.87 | 13,248.34 | .69 | .65 |
| M7 | 3 | 13,259.59 | 13,531.46 | 13,262.26 | .97 | .62 |
| M8 | 4 | 13,255.67 | 13,492.36 | 13,257.99 | .67 | .69 |
| M9 | 4 | 13,260.46 | 13,503.54 | 13,262.85 | .46 | .68 |
| M10 | 4 | 13,326.67 | 13,620.93 | 13,329.56 | .50 | 1.00 |

*Note*: Bolded values indicate the best-fitting model for that information criterion.

Table 5. Fit indices for two-, three-, and four-class cognitive models.

Sample characteristics for the two-class models (M2-M4) are presented in Table 6. In each of the models, *g* means were significantly higher ($p < .05$) in one class over another; these classes are henceforth referred to as “high ability.” Additionally, *g* variance was higher in the low ability groups in Models 3 and 4; *g* variance was constrained to be equal in M2.

| | M2 | | M3 | | M4 | |
|---|---|---|---|---|---|---|
| | Low | High | Low | High | Low | High |
| *N* | 108 | 73 | 105 | 76 | 55 | 126 |
| *g* variance | -- | -- | 1.60 | 1.31 | 2.36 | 1.32 |
| *g* mean | .00 | .82 | .00 | .63 | .00 | .72 |

Table 6. Descriptive statistics for the two-class model solutions.

**Three-class models**

Sample characteristics for the three-class models (M5-M7) are presented in Table 7. The patterns established in the two-class models were not as consistent in the three-solution. Although significantly different groups emerged based on *g* means, these models did not provide substantially improved fit based on information criteria such as aBIC. Furthermore, aLRT values suggested that a three-class model was not significantly better than an equivalent two-class model for any of the models tested.

| | M5 | | | M6 | | | M7 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Low | Mid | High | Low | Mid | High | Low | Mid | High |
| *N* | 22 | 62 | 97 | 22 | 69 | 90 | 79 | 36 | 66 |
| *g* variance | -- | -- | -- | 1.35 | 1.65 | 1.35 | 1.14 | 1.14 | 1.43 |
| *g* mean | .00 | .71 | 1.66 | .00 | .56 | 1.08 | .00 | .96 | 1.38 |

Table 7. Descriptive statistics for the three-class model solutions.

**Four-class models**

When four-class models were evaluated, several problems arose. In Model 8, the *g* means model, one class only had three individuals in it. In Model 9, which allowed *g* means and variances to freely vary, one class had a *g* mean below zero; this likely indicates that the zero-*g* class was not the lowest ability class. Finally, Model 10, which allowed for subtest variance to freely vary, all individuals ended up in the same class (which is why entropy for this model was 1.00). These observations, along with the fact that information criteria and other measures demonstrated poorer model fit for the four-class models, indicated that a four-class solution was not tenable.

Given that three- and four-class solutions had consistently worse fit than two-class solutions, the best two-class solution was chosen to create known classes for the next stage of analysis. Model 4, which allowed for *g* means, *g* variances, and subtest variances to vary across classes, was selected as it demonstrated the best model fit per the AIC and the aBIC. More importantly, this model most closely resembled SLODR (M. R. Reynolds, 2008) and demonstrated differences between high- and low-ability classes in both *g* means and *g* variances.

### ACHIEVEMENT MODELS

Fit statistics for each one-class achievement validation model are presented in Table 8. In each case, the model demonstrated adequate fit; that is, CFI and TLI were close to or greater than .95 and RMSEA and SRMR were close to or less than .05 (Schermelleh-Engel et al., 2003).

| **Achievement outcome** | $\chi^2$ | *df* | *p* | **CFI** | **TLI** | **RMSEA** | **SRMR** |
|---|---|---|---|---|---|---|---|
| Basic reading skills | 1,198.13 | 153 | <.001 | .97 | .96 | .04 | .06 |
| Reading comprehension | 1,020.80 | 136 | <.001 | .98 | .98 | .03 | .05 |
| Reading fluency | 1,007.59 | 136 | <.001 | .96 | .96 | .04 | .06 |
| Basic math skills | 1,015.87 | 136 | <.001 | .98 | .98 | .03 | .05 |
| Math problem solving | 1,066.52 | 136 | <.001 | .97 | .96 | .04 | .05 |
| Math fluency | 1,047.47 | 136 | <.001 | .94 | .92 | .05 | .06 |
| Basic writing skills | 1,045.43 | 136 | <.001 | .97 | .96 | .04 | .05 |
| Written expression (I) | 998.40 | 136 | <.001 | .98 | .98 | .03 | .05 |
| Written expression (II) | 973.83 | 136 | <.001 | .99 | .98 | .02 | .05 |

Table 8. Fit statistics for one-class achievement validation models.

Table 9 shows the comparative fit statistics of the one- and two-class achievement models using AIC, BIC, and aBIC. For every achievement outcome, all three criteria indicated better model fit for the one-class validation model. However, this step of analysis did not involve the selection of a better-fitting model.

| Achievement outcome | # of classes | AIC | BIC | aBIC |
|---|---|---|---|---|
| Basic reading skills | 1 | 15,794.49 | 15,989.60 | 15,796.41 |
| | 2 | 15,958.33 | 16,182.23 | 15,960.53 |
| Reading comprehension | 1 | 14,615.52 | 14,797.84 | 14,617.84 |
| | 2 | 14,784.87 | 14,995.97 | 14,786.95 |
| Reading fluency | 1 | 14,608.27 | 14,790.59 | 14,610.06 |
| | 2 | 14,779.16 | 14,990.26 | 14,781.24 |
| Basic math skills | 1 | 14,619.56 | 14,801.88 | 14,621.36 |
| | 2 | 14,789.76 | 15,000.86 | 14,791.83 |
| Math problem solving | 1 | 14,574.95 | 14,757.26 | 14,576.74 |
| | 2 | 14,746.89 | 14,958.00 | 14,748.97 |
| Math fluency | 1 | 14,627.37 | 14,809.69 | 14,629.16 |
| | 2 | 14,800.69 | 15,011.79 | 14,802.76 |
| Basic writing skills | 1 | 14,624.49 | 14,806.81 | 14,626.29 |
| | 2 | 14,795.35 | 15,006.45 | 14,797.43 |
| Written expression (I) | 1 | 14,636.91 | 14,819.22 | 14,638.70 |
| | 2 | 14,808.39 | 15,019.49 | 14,810.46 |
| Written expression (II) | 1 | 14,498.22 | 14,680.54 | 14,500.01 |
| | 2 | 14,666.61 | 14,877.71 | 14,668.68 |

Table 9. Comparative fit statistics for the one- and two-class achievement models.

Table 10 presents standardized *g* loadings for each achievement outcome in a one-class validation model, as well as the standardized *g* loadings for the high- and low-ability groups created by M4. Results from the one-class validation model indicate that *g* loadings for each of the achievement outcomes were between .40 and .73. Known classes generated by M4 demonstrated differences in their *g* loadings. In seven of the nine achievement models, the high-ability group had lower *g* loadings than the low-ability group; this finding supports the hypothesis that SLODR differentially predicts achievement based on ability within those domains. The two achievement outcomes that did not align with this hypothesis were reading fluency (measured by the Oral Reading Fluency subtest) and the first written expression measure (measured by the Sentence Composition subtest). For all models, *g* demonstrated statistically significant and large effects on the achievement domain (per the criteria in Keith, 2019), even in the high-ability group.

| **Achievement outcome** | **Standardized *g* loading in one-class model** | **Standardized *g* loading for high-ability group** | **Standardized *g* loading for low-ability group** |
|---|---|---|---|
| Basic reading skills | .61 | .44 | .64 |
| Reading comprehension | .61 | .54 | .63 |
| Reading fluency | .49 | .49 | .34 |
| Basic math skills | .58 | .54 | .62 |
| Math problem solving | .73 | .69 | .73 |
| Math fluency | .56 | .46 | .59 |
| Basic writing skills | .66 | .56 | .67 |
| Written expression (I) | .52 | .49 | .48 |
| Written expression (II) | .40 | .30 | .49 |

Table 10. Standardized *g* loadings for each achievement outcome in the one- and two-class achievement models.

**Reading**

In the one-class validation model, *g* demonstrated statistically significant and large effects on all three reading skills: basic reading skills, reading comprehension, and reading fluency. Basic reading skills, which was indicated by two subtests (Pseudoword Decoding and Word Reading), had a standardized *g* loading of .61 ($b = 4.19$; $\beta = 0.61$, $SE = 0.07$, $p < 0.001$), meaning that each standard deviation increase in *g* resulted in an increase of .61 standard deviations in basic reading skills after controlling for everything else. Reading comprehension, which was measured by the Reading Comprehension subtest, had the same standardized *g* loading as basic reading skills ($b = 5.38$; $\beta = 0.61$, $SE = 0.06$, $p < 0.001$). Reading fluency, which was measured by the

Oral Reading Fluency subtest, had a lower loading ($b = 4.25$; $\beta = 0.49$, $SE = 0.07$, $p < 0.001$).

In the two-class model, the SLODR hypothesis was supported for two of the three areas of reading. For basic reading skills, the high-ability group ($b = 2.84$; $\beta = 0.44$, $SE = 0.13$, $p < 0.01$) demonstrated lower loadings than the low-ability group ($b = 4.91$; $\beta = 0.64$, $SE = 0.07$, $p < 0.001$). The same finding was true for reading comprehension, where the high-ability group ($b = 4.88$; $\beta = 0.54$, $SE = 0.07$, $p < 0.001$) was less influenced by *g* than the low-ability group ($b = 6.09$; $\beta = 0.63$, $SE = 0.10$, $p < 0.001$). However, the opposite pattern was found for reading fluency, where the high-ability group ($b = 4.47$; $\beta = 0.49$, $SE = 0.10$, $p < 0.001$) was more influenced by *g* than the low-ability group ($b = 2.91$; $\beta = 0.34$, $SE = 0.13$, $p < 0.01$).

**Mathematics**

All three math achievement domains demonstrated statistically significant and large effects of *g* in the one-class validation model. Basic math skills, which was measured by the Numerical Operations subtest, had a standardized *g* loading of .58 ($b = 5.30$; $\beta = 0.58$, $SE = 0.06$, $p < 0.001$). Math problem solving, which was indicated by the Math Problem Solving subtest, had the largest *g* loading of any math domain ($b = 6.86$; $\beta = 0.73$, $SE = 0.05$, $p < 0.001$). Math fluency, measured by the Math Fluency subtest, demonstrated similar *g* effects as the basic math skills domain ($b = 6.13$; $\beta = 0.56$, $SE = 0.06$, $p < 0.001$).

In the two-class model, the low-ability group demonstrated higher standardized *g* loadings on each achievement domain than the high-ability group. For basic math skills, the high-ability group ($b = 4.81$; $\beta = 0.54$, $SE = 0.09$, $p < 0.001$) was less influenced by *g* than the low-ability group ($b = 5.90$; $\beta = 0.62$, $SE = 0.09$, $p < 0.001$). Math problem solving had the smallest differences between the *g* loadings of the high-ability ($b = 6.32$; $\beta = 0.69$, $SE = 0.06$, $p < 0.001$) and low-ability group ($b = 7.22$; $\beta = 0.73$, $SE = 0.06$, $p < 0.001$) within the math domains.

**Writing**

Just as with reading and math, the three writing domains all demonstrated large, statistically significant standardized *g* loadings onto each area. Basic writing skills, which was indicated by the Spelling subtest, was the most *g*-loaded writing domain ($b = 6.10$; $\beta = 0.66$, $SE = 0.05$, $p < 0.001$). The area of written expression was measured twice by two different subtests, Sentence Composition and Essay Composition, which have not been shown to correlate strongly with one another (Caemmerer et al., 2018). In the first written expression domain (Sentence Composition), the average *g* loading was higher ($b = 4.95$; $\beta = 0.52$, $SE = 0.07$, $p < 0.001$) than the second written expression domain, measured by Essay Composition ($b = 4.72$; $\beta = 0.40$, $SE = 0.08$, $p < 0.001$).

In the two-class model, high-ability standardized *g* loadings onto achievement domains were lower than the low-ability group in two out of the three measured areas. In basic writing skills, the high-ability group demonstrated a lower *g* loading ($b = 5.07$; $\beta = 0.56$, $SE = 0.09$, $p < 0.001$) than the low-ability group ($b = 6.75$; $\beta = 0.67$, $SE = 0.07$,

$p < 0.001$). The first written expression domain, measured by Sentence Composition, showed the opposite pattern, with the high-ability group's *g* loading being slightly higher ($b = 4.75$; $\beta = 0.49$, $SE = 0.09$, $p < 0.001$) than the low-ability group ($b = 4.64$; $\beta = 0.48$, $SE = 0.10$, $p < 0.001$). The hypothesized pattern was present in the second written expression domain, measured by Essay Composition, where the high-ability group demonstrated a lower *g* loading ($b = 3.61$; $\beta = 0.30$, $SE = 0.11$, $p < 0.01$) than the low-ability group ($b = 6.40$; $\beta = 0.49$, $SE = 0.09$, $p < 0.001$).

# CHAPTER 5: DISCUSSION

The purpose of the present study was to address whether a general intelligence factor (*g*) has smaller effects on achievement domains at higher ability levels. Latent mixture modeling was used to create latent ability classes using the full cognitive model of the WISC-V. A two-class solution was the best-fitting model and appeared to support the presence of SLODR. These known classes were then used to model differences in the relations between *g* and various achievement domains. The working hypothesis of this study was that the high-ability class (that is, the latent class with the higher *g* mean) would have smaller factor loadings from each achievement domain onto *g*. Indeed, in almost all cases, the standardized *g* loading was higher for the low-ability group as compared to the high-ability group. The average standardized *g* loading was also lower in the high-ability group (.50) as compared to the low-ability group (.58).

This section will discuss the implications of the present study across theoretical, methodological, and, where possible, practical domains. First, the implications for the study of Spearman's Law of Diminishing Returns will be examined, especially as this study built on the work of similar methodologies. Then, the cognitive-achievement relations examined in this study will be discussed across each achievement domain. Although this study was primarily theoretically and methodologically oriented, suggestions will then be made regarding how these findings may influence practice, especially within the school psychology domain. Finally, suggestions for future research and a discussion of the present study's limitations will be presented.

## IMPLICATIONS FOR SLODR

The present study served, in part, as a replication of the work of M. R. Reynolds et al. (2010), which tested whether SLODR could be effectively modeled using factor mixture modeling (FMM) approaches. While the present study used a sample based on the Wechsler tests, the Reynolds et al. study used the standardization sample from the KABC-II to test iterative latent class models.

M. R. Reynolds et al. (2010) outlined a few disadvantages of a priori group division when examining SLODR, the most notable of which is the determination of a cut-point that may not align with a particular model or theory. Since there is no theory that states how many ability classes exist within the continuous *g* distribution, such a division would necessarily be arbitrary. In addition, this approach also incurs the potential for range restriction due to the dichotomization of a continuous variable (Cohen, 1983; MacCallum et al., 2002) and the possibility that groups may not be factor invariant (Nesselroade & Thompson, 1995).

The M. R. Reynolds et al. study addressed these limitations by introducing a factor mixture modeling approach to the study of SLODR. This approach is advantageous as it is divides classes probabilistically; that is, FMM provides the probability of each individual belonging to a given class. These classes are derived from the overall model. In doing so, the interpretation is not that there are discrete classes within the overall sample, but rather that the latent variables may be distributed non-normally.

Of course, by allowing the FMM approach to select groups rather than by doing so manually, it becomes possible for individuals with higher FSIQ to be in the low-ability group and vice versa. This observation was made in by M. R. Reynolds et al. (2010), and the same was true in the present study. The most likely explanation of this oddity, as described by M. R. Reynolds et al. (2010), is that there are high-ability individuals who rely more on *g* and low-ability individuals who rely more on specific abilities when given items on an intelligence assessment.

M. R. Reynolds et al. (2010) posited a few key predictions regarding the detection of SLODR using FMM: First, it was predicted that a single class model would have worse model fit than a model with two latent classes. In the Reynolds study, a two-class model fit better than a one-class model. In the present study, a two-class model also demonstrated better fit according to the AIC and aBIC. This finding was true in each iterative two-class model (M2, M3, and M4). The best fitting model per the aBIC (M4) allowed *g* means, *g* variances, and subtest variances to differ across classes.

Another prediction of M. R. Reynolds et al. (2010) stated that classes should differ in their latent *g* means and variances, and that the class with the highest *g* mean should show smaller *g* variance. This hypothesis was indeed the case in the Reynolds study. In the present study, the two-class model solutions in which *g* variances were allowed to differ followed this pattern. In M3 and M4, the higher-ability class (i.e., the class with a non-zero *g* mean) had lower *g* variances than the lower-ability classes. However, this finding was not replicated in the three-class solutions where *g* variance could differ (M6 and M7). As *g* means increased across low-, mid- and high-ability groups, *g* variances either remained constant or increased. Given that the presence of SLODR is contingent upon the inverse relationship between *g* means and variances, these models are likely not a meaningful model of the phenomenon.

The final prediction of M. R. Reynolds et al. (2010) was that residual variances should be larger in the high-ability group since less subtest variance would be explained by *g*. This prediction was supported in the Reynolds study, as well as in the present study. Table 11 provides a comparison of subtest residuals between the high- and low-ability classes in the standardized M4 solution. Apart from three subtests (Arithmetic, Coding, and Symbol Search), the high-ability class demonstrated higher subtest residual variances than the low-ability class.

| | **Residual variances** | |
|---|---|---|
| **Subtest** | **High-ability** | **Low-ability** |
| Similarities | 0.57 | 0.30 |
| Vocabulary | 0.45 | 0.16 |
| Information | 0.61 | 0.36 |
| Comprehension | 0.65 | 0.61 |
| Block Design | 0.52 | 0.23 |
| Visual Puzzles | 0.49 | 0.13 |
| Matrix Reasoning | 0.85 | 0.68 |
| Figure Weights | 0.75 | 0.33 |
| Picture Concepts | 0.89 | 0.73 |
| Arithmetic | 0.42 | 0.50 |
| Digit Span | 0.51 | 0.23 |
| Picture Span | 0.73 | 0.67 |
| Letter-Number Sequencing | 0.51 | 0.15 |
| Coding | 0.36 | 0.45 |
| Symbol Search | 0.21 | 0.58 |
| Cancellation | 0.91 | 0.90 |

Table 11. Standardized subtest residual variances for each WISC-V subtest in the high- and low-ability classes (M4)

In M. R. Reynolds et al. (2010) and the present study, the presence of SLODR was detected. This finding further suggests that a general factor alone may not be equally meaningful across the cognitive ability spectrum. Considering this finding, researchers such as Kamphaus (2009) have suggested that a three-stratum model of intelligence should be used when making interpretations about an individual's test performance. More specifically, overall IQ scores such as FSIQ or GIA may be more meaningful and interpretable for lower ability individuals, while broad ability composites may explain test performance more accurately than a general factor score for individuals with higher ability. The presence of SLODR in the present study, as well as in prior studies, necessitates the study of how other domains may be predicted and explained differently between high- and low-ability classes.

### IMPLICATIONS FOR COGNITIVE-ACHIEVEMENT RELATIONS

A general intelligence factor (*g*) is a strong predictor of academic achievement across a variety of domains (Caemmerer et al., 2018; S. B. Kaufman et al., 2012). The present study sought to examine cognitive-achievement relations for each achievement domain measured by the WIAT-III (as determined by Caemmerer et al., 2018). In Caemmerer et al. (2018), models of the direct effects of *g* on various achievement outcomes found that *g*-loadings on reading domains were between .51 and .61; for math domains, loadings were between .58 and .76; for writing domains, loadings were between .42 and .68. The present study found similar standardized *g*-loadings in a one-class model: On reading domains, loadings were between .49 and .61; on math domains, loadings were between .56 and .73; for writing domains, loadings were between .40 and .66 (see Table 10).

These cognitive-achievement relations were compared between high- and low-ability classes generated in Model 4 as this model best captured the presence of SLODR; the underlying theory of SLODR would predict that the *g*-loading of a high-ability group would be lower for each achievement domain than a lower-ability group.

Few studies have addressed this premise, though they each suffer from potential methodological limitations. Coyle et al. (2011) used the SAT as a cognitive indicator and college grade-point average (GPA) as an achievement outcome, both of which are potentially more confounded by other factors such as practice effects or socioeconomic status than the measures used in the present study. Furthermore, factorial invariance between groups was unexamined. In McGill (2015), ability classes were created using a priori group formation: The WJ-III COG descriptive ranges (e.g., Below Average, Average, Above Average) were used to determine ability levels. Additionally, a higher-order factor analytic model of intelligence was not used; rather, the study relied on observed scores. The present study improves upon these prior examinations by measuring SLODR in a methodologically and theoretically advantageous way and by using well-researched measures of intelligence and achievement.

The standardized *g*-loadings on basic reading skills and reading comprehension were both lower in the high-ability group than the low-ability group, thus supporting the SLODR hypothesis. The opposite pattern was found for reading fluency, however, where the high-ability group was more influenced by *g* than the low-ability group. Less is known about the broad ability relations with reading fluency than the other two reading domains; both verbal comprehension (Gc) and processing speed (Gs) appear to have significant effects on reading fluency (Benson, 2007; Cormier et al., 2017; Niileksela et al., 2016). The finding that reading fluency did not conform to the SLODR hypothesis may be attributable to reading fluency's relation to processing speed, which is the least *g*-loaded broad ability. Alternately, or perhaps relatedly, it may simply be the case that reading fluency is immune to the effects of SLODR; it would therefore follow that reading fluency is more accurately predicted by *g* for higher-ability individuals than for lower-ability individuals. Indeed, some research has indicated that decoding matters more than comprehension for readers who are more dysfluent (e.g., Jenkins et al., 2003).

Across all three math domains (basic math skills, math problem solving, and math fluency), the SLODR hypothesis was generally supported, with all *g*-loadings being higher in the low-ability group. These differences were smallest in math problem solving and largest in math fluency. This finding runs counter to the notion that fluency measures do not conform to SLODR because of their relation to Gs (as was the case with reading fluency); Gs exerts a significant effect on math fluency, even when accounting for the effects of basic math skills (Niileksela et al., 2016). This discrepancy between math and reading fluency warrants additional consideration examining their patterns of cognitive-achievement relations.

The basic writing skills domain conformed to the SLODR hypothesis as predicted, but only one of the two written expression measurements followed the same pattern. While the domain measured by the Essay Composition subtest demonstrated a large difference between the *g*-loadings of the high- and low-ability classes, the domain measured by the Sentence Composition subtest had remarkably similar *g*-loadings. This discrepancy could be attributable to any number of differences between what the two subtests measure; for example, writing is influenced by short-term memory in the Wechsler tests (Caemmerer et al., 2018). In theory, an essay-oriented task would require a greater ability to hold information in mind than a task that only requires the individual to write a sentence containing a given word. However, surprisingly little research has examined cognitive-achievement relations related to writing, especially at smaller domain levels such as basic writing skills and written expression. Furthermore, differences in test batteries are most apparent on writing domains' relations with intelligence as compared to other domains (Caemmerer, 2017). This finding is likely because writing measures vary greatly between tests while reading and math measures are more similar.

Although the SLODR hypothesis was supported across most achievement domains, there is no theoretical guideline for what constitutes a "large" difference in *g*-loadings between high- and low-ability classes. Considering the novelty of this examination, it would be inappropriate to make over-generalizations regarding expected differences in high- and low-ability groups for each achievement domain. Rather, this study's findings can be more accurately stated in summary: The predictive nature of intelligence on achievement appears to decrease as ability increases, as predicted by SLODR.

## Implications for Practice

Although the present study is important from a theoretical and methodological perspective, there are potential practical implications, as well. First, the study served as a replication of previous SLODR examinations such as M. R. Reynolds et al. (2015) and supported the SLODR hypothesis. This finding alone has important implications on the practice of psychoeducational assessment. The Individuals with Disabilities Education Act (IDEA; 2004) dictates that cognitive factors must be considered in the context of a complete psychoeducational assessment; the notion that cognitive ability may be interpreted differently at different levels of ability could have an impact on special education eligibility determinations. Instead of relying solely on a general factor, this study, in conjunction with others that have found evidence of SLODR, would suggest that a three-stratum theory of intelligence should be used to interpret intelligence test findings. Generally, a general factor score such as FSIQ is more meaningful and descriptive at lower ability levels and broad ability composites are more meaningful and descriptive at higher ability levels; however, there remains no clear guidance regarding the exact point at which a general factor becomes more or less meaningful for interpretation.

The study further suggested that intelligence has variable effects on achievement based on overall ability. As such, increases in intelligence make broad ability scores less meaningful or predictive. Again, a single study of this topic using one set of tests is not sufficient for making broad determinations regarding the exact nature of these differences by domain. Nevertheless, these findings are suggestive of a more complex connection between cognition and achievement than previously hypothesized: Just as a general factor of intelligence likely varies across the ability spectrum, so does the relation between intelligence and achievement.

In school settings, psychologists are often asked to make comparisons between intelligence and achievement and assess the discrepancies or patterns observed (Flanagan et al., 2011; Naglieri, 2011). These comparisons are often the primary basis upon which special education eligibility decisions are made. These models of comparison are generally based upon a hierarchical model of intelligence in which broad ability scores are paired with associated achievement domains. If, as was suggested in the present study, achievement domains are generally more predictable by general intelligence at lower ability levels, then these approaches may be less applicable or appropriate for lower-ability students. Further studies may also extend the present findings to the broad ability level (i.e., whether SLODR also predicts changes in the broad ability-achievement relation). Additionally, researchers may also consider whether these patterns exist for gifted or twice-exceptional students.

### LIMITATIONS AND FUTURE RESEARCH

Of course, the present study was not without its limitations. A primary limitation of this study was the novelty of this approach for assessing cognitive-achievement relations. Factor mixture modeling, though established as way to test SLODR through extant research (e.g., M. R. Reynolds et al., 2015), remains a relatively unstudied topic. Not much is known regarding the performance of these models when latent categorical variables and latent continuous variables are modeled simultaneously (Bauer & Curran, 2003).

Further, the selection of an appropriate model of SLODR (and therefore the selection of classes) was both exploratory and confirmatory; despite the selection of a two-class model (M4), fit indices differed in their ability to identify a better fitting two-class model between AIC, BIC, and aBIC. More research is still needed that evaluates how various fit indices perform in complex mixture models. The present study, however, built on the existing factor structure of the WISC-V (Wechsler, 2014); this factor structure is based on CHC theory, a well-established theory of intelligence. Therefore, despite the exploratory nature of class selection (in that the number and nature of classes was unknown prior to analysis), the approach was improved by the confirmatory nature of the within-class model.

As is typically the case in research involving cognitive-achievement relations, the present study was limited to children and adolescents; therefore, these results may not be generalizable across the lifespan. Furthermore, the linking sample used was rather small ($n = 181$). This small sample size likely made the investigations of model fit in models with many parameters more prone to error. Future research of this topic should seek to find or create linking samples that are both larger and more representative of the complete age span.

The present study was guided by a hierarchical, higher-order approach to CHC theory as opposed to a bifactor, nested approach (Gustafsson & Balke, 1993; Mulaik & Quartetti, 1997). There remains some debate among researchers regarding which model is more appropriate when assessing intelligence; some researchers assert that a bifactor model does not actually align with the theoretical framework of CHC theory (Keith & Reynolds, 2012; Murray & Johnson, 2013). Nevertheless, the present study is limited in the sense that it only considered one theoretical approach to the measurement of intelligence. This consideration was not simply a theoretical decision, though: The WISC-V, while commonly used by psychologists, especially in school settings (Benson et al., 2019), is also informed by a hierarchical approach to CHC theory. Therefore, the theoretical approach undertaken in the present study is in alignment with the measures used. Future studies may consider the use of other test measures to generalize these findings across batteries.

## SUMMARY

The present study provided support to the use of factor mixture modeling in investigations of Spearman's Law of Diminishing Returns. This method appears to be an appropriate way to model high- and low-ability classes. Furthermore, the use of these known classes also appeared to be effective in modeling the effects of SLODR on various achievement domains. Of course, further research is needed to determine whether these findings are consistent across test batteries and among a larger age range. These findings may help psychologists better understand the relations between intelligence and achievement, as well as how *g* operates differently based on cognitive ability. The correlation between intelligence and achievement is well-documented, but is often understood to be consistent across the ability span. This assumption likely stems from the use of a global ability composite alone to make decisions about intelligence. The present study's findings suggest that appropriate psychoeducational evaluations should include more than just a broad intelligence composite score when examining cognitive ability, especially for students who demonstrate higher overall cognitive ability. More broadly, this study supports the use of a three-stratum theory of intelligence wherein the effects of broad abilities are meaningful above and beyond the effects of a general intelligence factor alone, especially in higher-ability individuals.

## References

Abad, F. J., Colom, R., Juan-Espinosa, M., & Garcıa, L. F. (2003). Intelligence differentiation in adult samples. *Intelligence, 31*(2), 157–166.

Ackerman, P. L. (2003). Cognitive ability and non-ability trait determinants of expertise. *Educational Researcher, 32*(8), 15–20.

Akaike, H. (1987). Factor analysis and AIC. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of Hirotugu Akaike* (pp. 371–386). Springer.

Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Blackwell Publishing.

Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies*. Cambridge University Press.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(4), 513–535.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338.

Benson, N. F. (2007). Cattell–Horn–Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment, 26*(1), 27–41.

Benson, N. F. (2017). Review of the Wechsler Intelligence Scale of Children. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurements yearbook* (5th ed.).

Breaux, K. C. (2010). *WIAT-III technical manual with adult norms*.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *Educational Psychology*, *19*, 176–199.

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287–321.

Caemmerer, J. M. (2017). *Beyond individual tests: The effects of children's and adolescents' cognitive abilities on their achievement* [PhD Thesis]. The University of Texas at Austin.

Caemmerer, J. M., Maddocks, D. L. S., Keith, T. Z., & Reynolds, M. R. (2018). Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-V and WIAT-III. *Intelligence*, *68*, 6–20.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31.

Carlstedt, B. (2001). Differentiation of cognitive abilities as a function of level of general intelligence: A latent variable approach. *Multivariate Behavioral Research*, *36*, 589–609.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153–193.

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology*, *33*(3), 300–317.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement*, *7*(3), 249–253.

Cormier, D. C., Bulut, O., McGrew, K. S., & Frison, J. (2016). The role of Cattell–Horn–Carroll (CHC) cognitive abilities in predicting writing achievement during the school-age year. *Psychology in the Schools*, *53*(8), 787–803.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment*, *35*(8), 731–754.

Coyle, T., Snyder, A., Pillow, D., & Kochunov, P. (2011). SAT predicts GPA better for high ability subjects: Implications for Spearman's Law of Diminishing Returns. *Personality and Individual Differences*, *50*(4), 470–474.

Cudeck, R., & MacCallum, R. C. (2007). *Factor analysis at 100: Historical developments and future directions*. Lawrence Erlbaum Associates.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*(2), 105–132.

Deary, I. J., & Pagliari, C. (1991). The strength of g at different levels of ability: Have Detterman and Daniel rediscovered Spearman's "law of diminishing returns"? *Intelligence*, *15*, 247–250.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21.

Detterman, D. K. (1991). Reply to Deary and Pagliari: Is g intelligence or stupidity? *Intelligence*, *15*, 251–255.

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence*, *13*(4), 349–359.

Dolan, V. C., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*(3), 227–253.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*, 246–262.

Evans, M. G. (1999). On the asymmetry of g. *Psychological Reports*, *85*(3_suppl), 1059–1069.

Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence*, *32*(4), 391–401.

Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2011). A CHC-based operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 233–298). John Wiley.

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell-Horn-Carroll model? Some evidence from a joint factor analysis of the Delis-Kaplan Executive Function System and the Woodcock-Johnson III tests of cognitive abilities. *Psychology in the Schools*, *47*(7).

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155–171.

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell-Horn-Carroll cognitive abilities and their effects on reading decoding skills: G has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly*, *22*(2), 200–233.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools*, *45*(2), 132–144.

Floyd, R. G., Meisinger, E., Gregg, N., & Keith, T. Z. (2012). An explanation of reading comprehension across development using models from Cattell-Horn-Carroll theory: Support for integrative models of reading. *Psychology in the Schools*, *49*(8), 725–743.

Galton, F. (1889). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, *45*(273–279), 135–145.

Garcia, G. M., & Stafford, M. E. (2000). Prediction of reading by Ga and Gc specific cognitive abilities for low-SES White and Hispanic English-speaking children. *Psychology in the Schools*, *37*(3), 227–235.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences*, *9*(1), 6–10.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, *40*(2), 103–114.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132.

Guilford, J. P. (1959). Three faces of intellect. *American Psychologist*, *14*(8), 469–479.

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*(4), 407–434.

Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review*, *43*(4), 385–406.

Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children*, *70*(4), 453–468.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances*. Jhu Press.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 202–226.

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Reshaping of American life by differences in intelligence*. Free Press.

Holzinger, K. J., & Harman, H. H. (1938). Comparison of two factorial analyses. *Psychometrika*, *3*(1), 45–60.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). The Guilford Press.

*Individuals with Disabilities Education Act, 20 U.S.C. § 1400*. (2004).

Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of educational psychology*, *95*(4), 719.

Jensen, A. R. (1992). Commentary: Vehicles of g. *Psychological Science*, *3*(5), 275–278.

Jensen, A. R. (2003). Regularities in Spearman's Law of Diminishing Returns. *Intelligence*, *31*(2), 95–105.

Jewsbury, P. A., Bowden, S. C., & Strauss, M. E. (2016). Integrating the switching, inhibition, and updating model of executive function with the Cattell—Horn—Carroll model. *Journal of Experimental Psychology: General*, *145*(2), 220–245.

Kamphaus, R. W. (2009). Assessment of intelligence and achievement. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 230–246). Wiley.

Kane, H. D., & Brand, C. R. (2006). The variable importance of general intelligence (g) in the cognitive abilities of children and adolescents. *Educational Psychology*, *26*(6), 751–767.

Kane, H. D., Oakland, T., & Brand, C. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *The Journal of Genetic Psychology*, *167*(3), 327–341.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence*, *40*(2), 123–138.

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly*, *14*(3), 239–262.

Keith, T. Z. (2017). Review of the Wechsler Intelligence Scale of Children. In J. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), *The twentieth mental measurements yearbook* (5th ed.).

Keith, T. Z. (2019). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (3rd ed.). Routledge.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7).

Keith, T. Z., & Reynolds, M. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 758–799). Guilford Press.

Kelley, T. L. (1928). *Crossroads in the mind of man: A study of differentiable mental abilities*. Stanford University Press.

Legree, P. J., Pifer, M. E., & Grafton, F. C. (1996). Correlations among cognitive abilities are lower for higher ability groups. *Intelligence*, *23*(1), 45–57.

Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for non-nested latent class models. *Journal of Educational and Behavioral Statistics*, *22*(3), 249–264.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767–778.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*(1), 21–39.

Lubke, G. H., & Spies, J. R. (2008). Choosing a "correct" factor mixture model: Power, limitations, and graphical data exploration. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent variable mixture models* (pp. 343–362). Information Age.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19–40.

McGill, R. J. (2015). Spearman's Law of Diminishing Returns (SLODR): Examining effects at the level of prediction. *Journal of Psychology and Behavioral Science*, *3*(1), 24–36.

McGrew, K. S. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and reading achievement across the lifespan. *Journal of Psychoeducational Assessment*, 39–53.

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. L. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review*, *26*(2), 189–210.

McGrew, K. S., & Hessler, G. L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment*, *13*, 21–38.

McGrew, K. S., & Knopik, S. N. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. *School Psychology Review*, *22*(4), 687–695.

McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, *47*(7), 651–675.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Mulaik, S. A., & Quartetti, D. A. (1997). First order or higher order general factor? *Structural Equation Modeling*, *4*(3), 193–211.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*(5), 407–422.

Muthén, L. K., & Muthén, B. (2012). *Mplus User's Guide, Seventh Edition*.

Naglieri, J. A. (2011). The discrepancy/consistency approach to SLD identification using the PASS theory. In D. P. Flanagan & V. C. Alonso (Eds.), *Essentials of specific learning disability identification* (pp. 145–172). John Wiley.

Nesselroade, J. R., & Thompson, W. W. (1995). Selection and related threats to group comparisons: An example comparing factorial structures of higher and lower ability groups of adult twins. *Psychological Bulletin*, *117*(2), 271–284.

Niileksela, C. R., Reynolds, M. R., Keith, T. Z., & McGrew, K. S. (2016). A special validity study of the WJ IV: Acting on evidence for specific abilities. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 65–106). Elsevier.

Nylund, K. L., Asparouhov, T., & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*(4), 535–569.

Oh, H. J., Glutting, J. J., Watkins, M. W., Youngstrom, E. A., & McDermott, P. A. (2004). Correct interpretation of latent versus observed abilities: Implications from structural equation modeling applied to the WISC-III and WIAT linking sample. *The Journal of Special Education*, *38*(3), 159–173.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, *86*, 159–175.

R Core Team. (2020). *R: A language and environment for statistical computing*.

Reynolds, M. R. (2008). *The use of factor mixture modeling to investigate population heterogeneity in hierarchical models of intelligence* [PhD Thesis]. The University of Texas at Austin.

Reynolds, M. R., & Hadorn, M. (2016). Our WISC-V review. In A. S. Kaufman, S. Raiford, & D. Coalson (Eds.), *Intelligence testing with the WISC-V* (pp. 615–636). John Wiley.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*, *35*(3), 267–281.

Reynolds, M. R., & Keith, T. Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition: What does it measure? *Intelligence*, *62*, 31–47.

Reynolds, M. R., Keith, T. Z., & Beretvas, S. N. (2010). Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence*, *38*(2), 231–241.

Reynolds, M. R., & Turek, J. J. (2012). A dynamic developmental link between verbal comprehension-knowledge (Gc) and reading comprehension: Verbal comprehension-knowledge drives positive change in reading comprehension. *Journal of School Psychology*, *50*(841–863).

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–17.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23–74.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99–144). The Guilford Press.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, *52*(3), 333–343.

Spearman, C. (1904). “General intelligence,” objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. MacMillan.

Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, *23*(2), 187–198.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*(4), 1097–1118.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences*, *13*, 159–188.

Wechsler, D. (2009). *Wechsler Individual Achievement Test, Third Edition*.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children, Fifth Edition*.

Wickham, H., Averick, M., Bryan, J., Chang, W., D’Agostino McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Milton Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wolfle, D. (1940). Factor analysis to 1940. *Psychometric Monographs*, *3*, 69.

Yang, C.-C. (1998). *Finite mixture model selection with psychometric applications*. University of California, Los Angeles.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*(3), 297–330.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113–128.