

8-11-2015

# Problem Formatting, Domain Specificity, and Arithmetic Processing: The Promise of a Factor Analytic Framework

Katherine Rhodes

Follow this and additional works at: [https://scholarworks.gsu.edu/psych\\_diss](https://scholarworks.gsu.edu/psych_diss)

---

## Recommended Citation

Rhodes, Katherine, "Problem Formatting, Domain Specificity, and Arithmetic Processing: The Promise of a Factor Analytic Framework." Dissertation, Georgia State University, 2015.  
[https://scholarworks.gsu.edu/psych\\_diss/138](https://scholarworks.gsu.edu/psych_diss/138)

This Dissertation is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

PROBLEM FORMATTING, DOMAIN SPECIFICITY, AND ARITHMETIC PROCESSING:  
THE PROMISE OF A FACTOR ANALYTIC FRAMEWORK

by

KATHERINE T. RHODES

Under the direction of Julie A. Washington, Ph.D. and Lee Branum-Martin, Ph.D.

ABSTRACT

Leading theories of arithmetic cognition take a variety of positions regarding item formatting and its possible effects on encoding, retrieval, and calculation. The extent to which formats might require processing from domains other than mathematics (e.g., a language domain and/or an executive functioning domain) is unclear and an area in need of additional research. The purpose of the current study is to evaluate several leading theories of arithmetic cognition with attention to possible systematic measurement error associated with instrument formatting (method effects) and possible contributions of cognitive domains other than a quantitative domain that is specialized for numeric processing (trait effects). In order to simultaneously examine measurement methods and cognitive abilities, this research is approached from a multi-trait, multi-method factor analytic framework.

A sample of 1959 3rd grade students (age  $M=103.24$  months,  $SD=5.41$  months) were selected for the current study from the baseline time points of a larger, longitudinal study conducted in southeastern metropolitan school districts. Abstract Code Theory, Encoding

## Complex Theory, Triple Code Theory, and the Exact versus Approximate Calculations

Hypothesis (a specification of Triple Code Theory) were evaluated with confirmatory factor analysis, using 11 measures of arithmetic with symbolic problem formats (e.g., Arabic numeral and language-based formats) and various problem demands (e.g., requiring both exact and approximate calculations). In general, results from this study provided support for both Triple Code Theory and Encoding Complex Theory, and to some extent, Exact Versus Approximate Calculations Theory is also supported. As predicted by Triple Code Theory, arithmetic outcomes with language formatting, Arabic numeral formatting, and estimation demands across formats were related but distinct from one another. The relationship between problems that required exact calculations (across formats) also provided support for Exact Versus Approximate Calculations Theory's stipulation that exact calculation problems may draw from the same cognitive processes. As predicted by Encoding Complex Theory, executive function was a direct predictor of all arithmetic outcomes. Language was not a direct predictor of arithmetic outcomes; however, the relationship between *language* and *executive function* suggested that *language* may play a facilitative role in reasoning during numeric processing, particularly for language-formatted problems.

INDEX WORDS: Arithmetic cognition, Numeric processing, Format effects, Common method variance, Language, Executive functioning, Abstract Code Theory, Encoding Complex Theory, Triple Code Theory, Exact versus approximate calculations

PROBLEM FORMATTING, DOMAIN SPECIFICITY, AND ARITHMETIC PROCESSING:  
THE PROMISE OF A FACTOR ANALYTIC FRAMEWORK

by

KATHERINE T. RHODES

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

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2015

Copyright by  
Katherine T. Rhodes  
2015

PROBLEM FORMATTING, DOMAIN SPECIFICITY, AND ARITHMETIC PROCESSING:  
THE PROMISE OF A FACTOR ANALYTIC FRAMEWORK

by

KATHERINE T. RHODES

Committee Chairs: Julie A. Washington  
Lee Branum-Martin

Committee: Lynn Fuchs  
Rebecca Williamson

Electronic Version Approved:

Office of Graduate Studies  
College of Arts and Sciences  
Georgia State University  
August 2015

## **ACKNOWLEDGMENTS**

I would like to thank all of the children, families, teachers, and schools who participated in this study. Without their willingness to participate, this research would not be possible. I would also like to thank my dissertation chairs and committee members for their willingness to guide me and answer questions as I completed this research. Finally, I would like to thank my family, colleagues, and mentors for their unwavering support throughout my years of graduate school.

## TABLE OF CONTENTS

<b>ACKNOWLEDGMENTS .....</b>	<b>VI</b>
<b>LIST OF TABLES .....</b>	<b>X</b>
<b>LIST OF FIGURES .....</b>	<b>XII</b>
<b>1 CHAPTER 1: ARITHMETIC MASTERY, COGNITIVE DOMAINS, AND MEASUREMENT FORMATTING.....</b>	<b>1</b>
<b>1.1 Introduction .....</b>	<b>1</b>
<b>1.2 Arithmetic Mastery and Mathematics Achievement Difficulties.....</b>	<b>3</b>
<b>1.3 Interpreting Test Results: Math Content Versus Test Formatting .....</b>	<b>5</b>
<b>1.4 The Quantitative Domain and the Development of Arithmetic Cognition .....</b>	<b>7</b>
<i>1.4.1 Neural bases of arithmetic cognition.....</i>	<i>7</i>
<i>1.4.2 Other neural circuits in arithmetic cognition. ....</i>	<i>9</i>
<b>1.5 Problem Formatting and Common Method Variance.....</b>	<b>11</b>
<b>1.6 Cognitive Theories of Arithmetic.....</b>	<b>13</b>
<i>1.6.1 Abstract code theory. ....</i>	<i>13</i>
<i>1.6.2 Encoding complex theory.....</i>	<i>16</i>
<i>1.6.3 Triple code theory.....</i>	<i>19</i>
<i>1.6.4 Exact versus approximate calculations: An extension of triple code theory. ..</i>	<i>22</i>
<b>2 CHAPTER 2: MODELING ARITHMETIC PERFORMANCE AND THE ROLE OF PROBLEM FORMATTING .....</b>	<b>25</b>
<b>2.1 Modeling Leading Theories of Arithmetic Cognition .....</b>	<b>26</b>
<i>2.1.1 Abstract code theory. ....</i>	<i>27</i>
<i>2.1.2 Encoding complex theory.....</i>	<i>32</i>

2.1.3	<i>Triple code theory.....</i>	39
2.1.4	<i>Exact versus approximate calculations specification of triple code theory. ....</i>	45
2.2	Hypotheses.....	49
3	CHAPTER 3: METHODS .....	50
3.1	Participants .....	50
3.2	Procedures.....	54
3.3	Measures.....	54
3.3.1	<i>Mathematics achievement measures with language formatting. ....</i>	54
3.3.2	<i>Mathematics achievement measures with Arabic numeral formatting. ....</i>	56
3.3.3	<i>Mathematics achievement measures involving estimation or analog magnitude. ....</i>	59
3.3.4	<i>Language measures.....</i>	60
3.3.5	<i>Executive functioning measures.....</i>	61
3.4	Design.....	64
4	CHAPTER 4: RESULTS .....	70
4.1	Proposed Analyses Overview.....	70
4.2	Phase 1: Measurement Models for Arithmetic, Language, and Executive Functioning.....	70
4.2.1	<i>Abstract code model.....</i>	70
4.2.2	<i>Encoding complex model. ....</i>	72
4.2.3	<i>Triple code model. ....</i>	75
4.2.4	<i>Exact versus approximate model. ....</i>	77
4.2.5	<i>Language model. ....</i>	80

4.2.6	<i>Executive functioning model.</i>	82
4.3	<b>Phase 2: Full Measurement Models for Each Theory</b>	85
4.3.1	<i>Abstract code model.</i>	85
4.3.2	<i>Encoding complex model.</i>	88
4.3.3	<i>Triple code model.</i>	94
4.3.4	<i>Exact versus approximate model.</i>	98
4.3.5	<i>Post-hoc testing: Hybrid full measurement model.</i>	102
4.3.6	<i>Summary of Model Testing Results.</i>	109
5	<b>CHAPTER 5: DISCUSSION</b>	114
5.1	<b>Summary of Major Findings</b>	114
5.1.1	<i>The structure of arithmetic cognition.</i>	114
5.1.2	<i>Symbolic formatting and calculation demands.</i>	114
5.1.3	<i>Contributions from executive function and language.</i>	115
5.2	<b>Implications for Measuring Arithmetic.</b>	118
5.3	<b>Limitations and Future Directions.</b>	119
5.3.1	<i>Adapting theories toward specific measurement hypotheses.</i>	119
5.3.2	<i>Adapting theories toward developmental hypotheses.</i>	120
5.3.3	<i>Generalizability of symbolic formatting.</i>	122
5.3.4	<i>Overlap in features of item modality.</i>	123
5.3.5	<i>Measuring and modeling executive function.</i>	125
5.4	<b>Summary and Conclusions</b>	126
	<b>REFERENCES.</b>	128

## LIST OF TABLES

Table 1. Patterns of unavailable data on demographic variables of interest.....	53
Table 2. Cohort Measurement Information .....	65
Table 3. Most prevalent patterns of unavailable (planned missing) data on outcome measures of interest.....	66
Table 4. Math Measures Means and Correlations .....	67
Table 5. Language Measures Means and Correlations .....	68
Table 6. Executive Functioning Measures Means and Correlations.....	68
Table 7. Full Correlation Matrix for All Measures.....	69
Table 8. Abstract Code Theory Arithmetic Measurement Model CFA Results.....	71
Table 9. Encoding Complex Theory Arithmetic Measurement Model CFA Results.....	74
Table 10. Triple Code Theory Arithmetic Measurement Model CFA Results.....	76
Table 11. Triple Code Theory Arithmetic Latent Factor Correlations .....	77
Table 12. Exact Versus Approximate Calculations Arithmetic Measurement Model CFA Results .....	79
Table 13. Language Measurement Model CFA Results .....	81
Table 14. Executive Function Measurement Model CFA Results .....	84
Table 15. Abstract Code Theory Full Measurement Model CFA Results.....	86
Table 16. Abstract Code Theory Full Model Latent Factor Correlations.....	86
Table 17. Encoding Complex Theory Full Measurement Model Unstandardized CFA Results..	90
Table 18. Encoding Complex Theory Full Measurement Model Completely Standardized CFA Results.....	91
Table 19. Encoding Complex Theory Full Model Latent Factor Correlations.....	92

Table 20. Triple Code Theory Full Measurement Model CFA Result .....	96
Table 21. Triple Code Theory Full Model Latent Factor Correlations.....	96
Table 22. Exact V. Approximate Calculations Full Measurement Model CFA Results .....	100
Table 23. Exact V. Approximate Calculations Full Model Latent Factor Correlations .....	100
Table 24. Post Hoc Hybrid Full Measurement Model Completely Standardized CFA Results.	106
Table 25. Post Hoc Hybrid Full Model Latent Factor Correlations .....	107
Table 26. Summary of Model Testing Results .....	113

## LIST OF FIGURES

Figure 1. Abstract Code Theory: 3 Modules for Numeric Processing. ....	28
Figure 2. Abstract Code Theory: 1 Mental Representation Regardless of Stimulus Format. ....	29
Figure 3. Abstract Code Theory: Other Domains Do Not Predict Arithmetic Behavioral Outcomes. ....	30
Figure 4. Abstract Code Theory: General Factor Model. ....	32
Figure 5. Encoding Complex Theory: No Specialized Module for Numeric Processing → Becomes A Seemingly-Modular Encoding Complex with Practice. ....	34
Figure 6. Encoding Complex Theory: Language Formatting Effects As Language-based Common Method Variance. ....	35
Figure 7. Encoding Complex Theory: Other Domains Predict Arithmetic Behavioral Outcomes. ....	36
Figure 8. Encoding Complex Theory: General Factor Model. ....	39
Figure 9. Triple Code Theory: 3 Modules for Numeric Processing. ....	40
Figure 10. Triple Code Theory: Formatting Can Effect Mental Representation & Processing (inherent in the theory). ....	41
Figure 11. Triple Code Theory: Other Domains May Be Associated With 3 Modules in Predicting Arithmetic Behavioral Outcomes. ....	43
Figure 12. Triple Code Theory: General Factor Models. ....	45
Figure 13. Exact V. Approximate Theory: 2 Modules for Exact and Approximate Calculations. ....	46
Figure 14. Exact V. Approximate Theory: Problem Formatting Differences Not Specified. ....	47
Figure 15. Exact V. Approximate Theory: Other Cognitive Domains' Contributions Are Unclear. ....	49

Figure 16. Abstract Code Theory: Arithmetic Measurement Model.....	72
Figure 17. Encoding Complex Theory: Arithmetic Measurement Model.....	74
Figure 18. Triple Code Theory: Arithmetic Measurement Model.....	77
Figure 19. Exact versus Approximate Theory: Arithmetic Measurement Model. ....	80
Figure 20. Language Measurement Model. ....	82
Figure 21. Executive Function Measurement Model.....	84
Figure 22. Abstract Code Theory: Full Measurement Model.....	87
Figure 23. Encoding Complex Theory: Full Measurement Model.....	93
Figure 24. Triple Code Theory: Full Measurement Model.....	97
Figure 25. Exact versus Approximate Theory: Full Measurement Model. ....	101
Figure 26. Hybrid Model of Triple Code Theory Arithmetic and Encoding Complex Theory Structure as a Full Measurement Model .....	108

# **1 CHAPTER 1: ARITHMETIC MASTERY, COGNITIVE DOMAINS, AND MEASUREMENT FORMATTING**

## **1.1 Introduction**

Arithmetic is a process requiring both knowledge of what numbers are (numerosity) and knowledge of how to perform the most basic operations of addition, subtraction, multiplication, and division on numbers (operational or algorithmic knowledge; Brodinsky, 1977; Woodward, 2004). Arithmetic mastery is essential for successful daily living as well as for advanced-level participation in Science, Technology, Engineering, and Mathematics (STEM) disciplines (AAIDD, 2010; STEM Coalition, 2000). Despite decades of efforts toward mathematics education reform, children in the U.S. continue to struggle with math achievement, and this is true of both basic arithmetic skills and more advanced problem solving (National Center for Education Statistics, 2013; Woodward, 2004). Attempts to impact these difficulties have focused largely on a national shift in mathematics curriculum, changing the emphasis and delivery of mathematics content in efforts to impact children's mathematics achievement (Woodward, 2004). This project will explore the possibility that it is more than mathematical content that is influencing these difficulties in mathematics achievement. In particular, it has been suggested that problem formatting, the modality used to convey operands and operators in a mathematics problem during testing, may also be important to consider.

Unpacking the extent to which arithmetic achievement is a function of math content (e.g., accurate calculation) as opposed to test formatting (e.g., linguistic understanding for a word problem) is a crucial issue of valid test design and interpretation. Differentiating between content and test formatting effects requires mathematics researchers to separate individual traits that allow students to demonstrate math understanding from the possible effects of test formatting.

Unfortunately, extant theories of arithmetic cognition do not provide clear specifications of the traits of cognition that operate across various types of math problems.

The purpose of this research is to evaluate several leading theories of arithmetic cognition with special attention to possible systematic measurement error associated with item formatting and to possible contributions of cognitive abilities other than a quantitative domain that is specialized for numeric processing. Specifically, the following research questions and hypotheses guide the research described in this document:

**RQ<sub>1</sub>:** What is the cognitive structure of mathematics ability(s) involved in arithmetic cognition?

**RQ<sub>2</sub>:** Do problem formats (language versus Arabic numeral symbolic formats) affect the access of the mathematics domain being tested?

**RQ<sub>3</sub>:** Do problem demands (exact versus approximate calculation demands) affect the access of the mathematics domain being tested?

**RQ<sub>4</sub>:** Do language and executive functioning abilities contribute to arithmetic cognition, or does numeric processing appear to be a mathematics domain specific task?

Four potential hypotheses arise from each theory of arithmetic cognition (i.e., Abstract Code Theory, Encoding Complex Theory, Triple Code Theory, and Exact versus Approximate Specification of the Triple Code Theory) being examined in this project.

## **1.2 Arithmetic Mastery and Mathematics Achievement Difficulties**

Findings from national studies of mathematics achievement suggest that a significant portion of children (and likely adults) in the U.S. do not master basic arithmetic skills and among those who do master arithmetic, many cannot extend this basic knowledge to the more complex types of mathematical problem-solving they will encounter in everyday life. For example, results from the most recent National Assessment of Educational Progress (NAEP) found that 58% of 4<sup>th</sup> grade students were performing below grade level proficiency in mathematics, meaning that they were unable to consistently apply procedures and concepts in a variety of math content areas (e.g., number properties, operations with numbers) in everyday, applied mathematics problems (National Center for Education Statistics, 2013). Among those 4<sup>th</sup> grade students who were below grade level proficiency in math, 17% were unable to demonstrate even basic understanding of these math procedures and concepts (National Center for Education Statistics, 2013).

By 8<sup>th</sup> grade, 64% of NAEP students were performing below grade level proficiency in mathematics, meaning that they were unable to apply math procedures and concepts to more complex problems (National Center for Education Statistics, 2013). Among those 8<sup>th</sup> grade students who were below grade level proficiency in mathematics, 26% were unable to demonstrate understanding of basic arithmetic operation (including addition, subtraction, multiplication, division, and estimation) with problems involving whole numbers, fractions, decimals, and percentages (National Center for Education Statistics, 2013).

The mathematics achievement trends for high school students in the U.S. are similar. The most recent Program for International Student Assessment (PISA) study, which examines achievement profiles for 15 year old high school students in the U.S., found that 8% of

adolescents in the U.S. have not fully mastered basic arithmetic and are below the lowest level of mathematics proficiency, and of the 92% who had sufficient arithmetic knowledge for many tasks in everyday life (e.g., addition with whole numbers), 44% were not able to use this knowledge to solve more complex, multistep problems (i.e., students at levels 1 or 2 of mathematics proficiency; Kelly et al., 2013). These findings suggest that over 50% of American high school students struggle with everyday mathematics problems involving sequential decision-making, fractions, decimals, proportional relationships, and providing basic interpretations of their arithmetic reasoning (i.e., the arithmetic reasoning skills that are required for everyday tasks like deciding the order in which to pay bills and then projecting how much money will be left to pay remaining bills, deciding how much money is appropriate for a tip at a restaurant, or adjusting the ingredients in a recipe for a larger number of people).

Each of these studies present a cross-sectional picture of mathematics achievement in the U.S., but they do not provide researchers with an understanding of how arithmetic competencies (or difficulties) develop. Understanding the developmental trajectory of arithmetic cognition is crucial for understanding the difficulties some children have with arithmetic cognition and identifying areas for intervention. It appears that many of the children who struggle with basic arithmetic in early elementary school are the same children (and possibly adolescents and adults) who later struggle with arithmetic and more advanced mathematical problem-solving (Geary, Hoard, Nugent, & Bailey, 2013); however, finding valid mathematics achievement instruments that can reliably identify young children with learning difficulties is challenging.

In the U.S., 3rd and 4th grade is a time at which many children are first identified as having significant learning difficulties, (particularly using an IQ-achievement discrepancy model for identifying learning difficulties; see for example Lyon et al., 2001). This identification of

learning difficulties in later elementary school years may be partially because it is difficult to use mathematics achievement patterns with very young children to identify and/or predict math learning difficulty (Gersten, Jordan, & Flojo, 2005). Poor mathematics achievement at one time point may not predict future learning difficulty, and identifying mathematics achievement instruments that capture skills essential for continued learning constitutes a major issue of measurement validity (Gersten et al., 2005). Indeed, research focused on explaining achievement discrepancies for students who would otherwise be predicted to achieve within typical ranges (e.g., students with learning difficulties, students from lower socio-economic backgrounds, or students who are English language learners) has been instrumental in raising questions about validity and reliability of mathematics achievement instruments.

### **1.3 Interpreting Test Results: Math Content Versus Test Formatting**

Understanding what assessment instruments are actually measuring is a necessary first step in understanding how their results should be interpreted, and some researchers (see for example Abedi & Lord, 2001) have argued that language-formatted mathematics assessment instruments like those used during NAEP testing may have inherent testing bias. The NAEP and PISA assessments attempted to measure children's understandings of real-world mathematics problems and often used "word problems" or language-formatted items to prompt students' responses. Although these patterns of arithmetic mastery and mathematics achievement are usually interpreted as indicating children's difficulties with math content, researchers have questioned the extent to which testing trends are the result of children's difficulties with the language of the mathematics achievement tests themselves (Abedi, Hofstetter, Baker, & Lord, 2001; Abedi, Lord, & Hofstetter, 1998; Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997; Martiniello, 2009; Rhodes, Branum-Martin, Morris, Ronski, & Sevcik, in press; Shaftel, Belton-

Kocher, Glasnapp, & Poggio, 2006; Terry, Hendrick, Evangelou, & Smith, 2010), particularly for students who are language minorities (e.g., English Language Learners, African American English dialect speakers, students with language disorders).

This issue of problem formatting and its effects on test validity (Messick, 1989, 1996) must be examined in order to determine the extent to which students are struggling with arithmetic as opposed to other, unintentionally accessed cognitive abilities (e.g., language, working memory). Mathematics test validity is perhaps best approached using psychometric analyses of mathematics achievement instruments in combination with theoretical models of arithmetic cognition. This approach allows direct testing of theoretical specifications of what cognitive abilities predict behavior (traits) as well as direct testing of problem formatting effects on behavioral outcomes (methods).

Theories of arithmetic cognition attempt to address the issue of arithmetic measurement by specifying aspects of the process of arithmetic: how we do arithmetic, what mental processes are involved in arithmetic, and why are we able (or unable) to successfully do arithmetic. Cognitive theories of arithmetic attempt to explain (1) how we encode numerical information and represent numerical information mentally, (2) how we retrieve math facts from memory, process the information, or operate upon numerical representations to achieve solutions to problems, (3) how we recode our mental, numerical representations of solutions into output and report our answers, and (4) which cognitive domains are involved in these activities. In general, these four facets compose the definition of the process of arithmetic for cognitive theories of arithmetic, and each of these facets of arithmetic are areas in which theories of arithmetic cognition may diverge from each other, sometimes irreconcilably. The consequence of this divergence has been that there is no consensus assessment of numeric processing.

## 1.4 The Quantitative Domain and the Development of Arithmetic Cognition

One area where dominant theories of arithmetic cognition largely agree is that there appears to be a *quantitative* domain of learning which is responsible for numeric processing tasks; it is believed that this is the domain which we are attempting to measure when we design and apply mathematics measurement instruments. However, the extent to which this domain changes over the course of human development, the extent to which its mental representations of number are influenced by measurement formatting, and the extent to which it relies upon other cognitive domains in order to accomplish numeric processing are unclear. Theorists disagree in their descriptions of these facets of arithmetic cognition, and their differing accounts present a challenge for psychometric evaluation of mathematics measurement. The following sections will review research on these issues and present four leading theories of arithmetic cognition.

### 1.4.1 *Neural bases of arithmetic cognition.*

Research indicates that there is a neural basis for numeric processing, supporting the notion that there is a cognitive domain largely responsible for recognizing quantity and processing arithmetic problems. The neural basis for this quantitative domain or “number sense” is well established in neuropsychological research with adults (S Dehaene, 2011); however, the neuropsychological changes in number sense across early childhood development are less well-understood. The few neuropsychological studies of quantity cognition with infants and young children indicate that this early, non-symbolic number sense has a neural basis in the intraparietal sulcus (IPS) which exists before children have formal educational experiences and is similar to the neural network utilized by adults. Research to date suggests that infants, young children, and adolescents display a pattern of IPS right hemisphere lateralization during numeric processing tasks that is different from the bilateral activations typically seen in adults (Ansari & Dhital,

2006; Cantlon, Brannon, Carter, & Pelphey, 2006; Izard, Dehaene-Lambertz, & Dehaene, 2008; Rivera, Reiss, Eckert, & Menon, 2005). Left lateralization of the IPS may increase with age, suggesting that the IPS becomes more specialized for numerical processing and more integrated with other neural circuits over the course of development and formal education (Ansari & Dhital, 2006; Cantlon, Brannon, Carter, & Pelphey, 2006; Izard, Dehaene-Lambertz, & Dehaene, 2008; Rivera, Reiss, Eckert, & Menon, 2005).

Behavioral studies of infant quantity cognition have found that human infants as young as 4 to 5 months of age display the ability to detect small quantities and perform comparisons of more or less (subitization or object file representation; e.g., Feigenson, Carey, & Hauser, 2002; Starkey & Cooper, 1980). At five months of age, infants also can apply this early number sense to perform more complex number operations (addition and subtraction) on small quantities (Wynn, 1992). When sets of quantities differ by large enough ratios and physical characteristics such as surface area and volume are controlled, six-month old infants are also able to discriminate between larger sets of quantities (analog magnitude representation; Xu & Spelke, 2000). However, the stimulus formats for these behavioral experiments all involve non-symbolic items (e.g., arrays of dots, tones, line drawings), and the extent to which this naïve number sense represents the sophisticated, symbolic numerical processing utilized by adults and older children (e.g., rapid numerical calculations with language-formatted or Arabic numeral symbols) is still being explored.

Research with neurotypical adults has consistently found that the horizontal (bilateral) segment of the intraparietal sulcus (HIPS) is activated during activities involving quantity recognition, number comparisons, approximate calculations, and exact calculations, suggesting that the HIPS region of the brain is essential for the mental representation of quantity (Dehaene,

Spelke, Pinel, Stanescu, & Tsivkin, 1999; Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, Piazza, Pinel, & Cohen, 2003). The HIPS activation during numeric processing tasks appears to be consistent across a variety of problem formats (e.g., language-based, Arabic numeral, visuospatial), suggesting that it is amodal and can be accessed regardless of stimulus format, unlike other brain regions supporting numeric processing (Dehaene et al., 1999; Dehaene et al., 2004).

Unfortunately, functional magnetic resonance imaging (fMRI; often considered to be a gold standard in neuro-imaging methodologies) requires participants to be very still, a methodological necessity that works for or adults, but which can be very difficult for young children. Consequently, most extant theories of arithmetic cognition pertain to the complex, integrated, and specialized arithmetic performed by skilled adults. However, research from a variety of methodological traditions supports the ideas that (1) some aspects of children's arithmetic abilities seem to be innate or at least present before exposure to formal education, and (2) there appears to be some developmental continuity between the naive arithmetic cognition of infants and young children, the developing arithmetic cognition of older children and adolescents, and the formalized arithmetic cognition of adults.

#### **1.4.2      *Other neural circuits in arithmetic cognition.***

Although there appears to be a unique neural circuit responsible for numeric processing across the course of human development, other regions of the brain (and other cognitive dimensions) may also be activated during arithmetic tasks. The extent to which other regions of the prefrontal cortex and the parietal lobe (most often the precentral sulcus, inferior frontal gyrus, and angular gyrus) are activated during numeric processing tasks depends on the problem formatting (e.g., language, Arabic numerals) and processing demands (e.g., quantity

comparisons, multi-step calculations) of the task. The precentral sulcus and inferior frontal gyrus have been consistently implicated in calculation activities, suggesting that they may contribute to arithmetic activities involving demands on the cognitive domains of working memory, sequencing, and planning (Stanescu-Cosson et al., 2000). During tasks involving complex language (e.g., spatial metaphors, phonemic awareness, word associations) as well as those involving quantity (e.g., digit naming, exact quantity calculations), the angular gyrus is also activated, suggesting that it may play a role in language-based fact retrieval (Dehaene et al., 1999; Dehaene et al., 2004).

Problem formatting also has an effect on behavioral indicators such as reaction time, error propagation, and accuracy, all of which play an important role in arithmetic performance, especially during achievement testing. Language formats in which number words, rather than Arabic numerals, are presented as operands increase both reaction time and error propagation by as much as 30% (J I Campbell, 1994). These formatting effects raise questions about the cognitive domain(s) responsible for conducting arithmetic. Difficulty with language-formatted problems may reflect problems with encoding (difficulty getting the input into the quantitative domain and mentally representing it there; McCloskey, Macaruso, & Whetstone, 1992). Alternately, difficulty with language-formatted problems may reflect problems with production (difficulty getting the output out of the quantitative domain due to problems of phonological interference with the output of the spoken answer; Noël, Fias, & Brysbaert, 1997). However, from a more comprehensive perspective, difficulty with language-formatted problems may be due to problems with interfering abilities during numeric processing in general (difficulty performing quantitative tasks due to interference from a language domain; Campbell & Clark, 1988).

## 1.5 Problem Formatting and Common Method Variance

The process of encoding and mentally representing arithmetic problems presented in various formats is not trivial because correct encoding is crucial for successful performance; however the extent to which problem formatting may affect subsequent processing and response generation is unclear. Previous studies provide some evidence that item features like language-formatting and problem size may make arithmetic problems more difficult, increasing reaction time and lowering the probability of correct solutions (Campbell, 1994) and that circuits of the brain involved in language processing may also aid in arithmetic processing (Dehaene et al., 1999; Dehaene et al., 2004). Research also indicates that language-formatting may be of particular consequence for linguistic minorities, who may struggle with language formatted math problems because of difficulty encoding in less familiar language formats (e.g., Abedi et al., 2001, 1998; Abedi & Lord, 2001; Abedi, Lord, & Plummer, 1997; Martiniello, 2009; Rhodes, Branum-Martin, Morris, Ronski, & Sevcik, in press; Shaftel et al., 2006; Terry et al., 2010).

From a psychometric perspective, the idea that problem formatting may affect the likelihood of generating a correct response is also an issue of common method variance. For example, children taking a vocabulary test may be more likely to correctly answer items about travel abroad and expensive leisure activities if they are of higher socioeconomic status. Items prompting these content areas may share common method variance for socioeconomic status above and beyond the variance they share with the rest of the vocabulary items. Common method variance, or variance that is due to measurement methods instead of the constructs under investigation, constitutes a serious threat to validity (Cote & Buckley, 1987, 1988; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). The idea that children may be more likely to correctly answer arithmetic problems which are formatted with Arabic numerals than problems which are

language-formatted would indicate that these language-formatted items share some common method variance above and beyond the variance they share with other items measuring the *quantitative* domain. Evaluating common method variance is a necessity for the theoretical evaluation of the extent to which problem formatting may affect arithmetic cognition.

The existence of common method variance is always evidence that some dimension other than the intended construct is being tested (Messick, 1989, 1996). Examining various arithmetic problem formats for mean differences in raw total scores, observing reaction time differences, describing differential patterns of error propagation, and identifying differences in neural network activation patterns are all methodologies that can support the idea that problem formatting may influence arithmetic cognition. However, these methodologies do not provide sufficient evidence for evaluating common method variance related to problem formatting because they cannot directly evaluate the fundamental question of whether a dimension other than *quantity* has been accessed by certain formats.

In order to examine the extent to which cognitive dimensions other than *quantitative* ability may be involved in generating responses to arithmetic problems, these other dimensions must be measured along with *quantity*, in a variety of formats, and included in statistical models of responses which evaluate not only mean structures, but also variance structures. This can be accomplished with a multitrait, multimethod methodology as well as statistical models capable of allowing for the modeling the possibility that multiple abilities may predict behaviors (see for example Cote & Buckley, 1987; Eid, Lischetzke, & Nussbeck, 2006; V. Marsh, Beard, & Bailey, 2002; Maul, 2013). These statistical models fall under the broad umbrella of factor analysis. Thus, evaluating theories of math cognition using a multitrait, multimethod, factor analytic

framework has substantive implications for scientific understanding of math cognition and math measurement (stimulus formatting).

## **1.6 Cognitive Theories of Arithmetic**

There appears to be a neural basis for quantity cognition, but it has nuance that remains unexplained by research to date. Questions remain about the extent to which cognitive domains responsible for language and executive functioning (including planning, sequencing, attention regulation, and working memory) may also play a role in certain types of numeric processing and the extent to which problem modality influences mental representation and subsequent operations upon quantity. Although a factor analytic examination of these issues could help to resolve some of these remaining questions about arithmetic cognition, hypotheses about the cognitive domains responsible for various arithmetic behaviors must be developed in order to guide modeling. Cognitive theories of arithmetic can help to specify model construction and hypothesis development. The following sections will present four of the leading theories of arithmetic cognition, considering their specifications for the process of arithmetic with special attention as to how they attempt to explain language-formatting effects and the roles that language and executive functioning domains may play in arithmetic performance.

### **1.6.1 *Abstract code theory.***

Abstract code theory stipulates that a single, abstract code is used to mentally represent all numeric information, regardless of input format (Arabic numeral, number words, operation procedures, arithmetic math facts; McCloskey, Caramazza, & Basili, 1985; McCloskey, 1992). Three domains are responsible for numeric processing in abstract code theory, the *comprehension*, *processing*, and *production* domains. The *comprehension* domain is responsible for recognizing numeric stimuli, encoding stimuli into abstract code for subsequent processing,

and activating procedural routines in the *processing* domain. The *processing* or *calculation* domain is responsible for calling on *comprehension* and *production* domains for various stages of needed input and output, retrieving arithmetic facts from memory, and executing calculations. Finally, the *production* domain is responsible for translating abstract codes into appropriate output formats. For McCloskey (1985; 1992), the processes of mentally representing stimulus input as abstract code or translating abstract codes to produce output is referred to as transcoding.

All input stimuli are encoded into an amodal, abstract code, which contains semantic information about quantity and is the basis for subsequent calculations and response productions (McCloskey et al., 1985; McCloskey, 1992). These abstract codes have both lexical and syntactic properties. The lexical properties of abstract codes are individual elements in the numeral, and the syntactic properties are relationships among elements in the numeral that facilitate comprehension of the numeral as a whole. For example, the input “13” has lexical properties {1} and {3} as well as syntactic properties  $10^1$  (tens) and  $10^0$  (ones) presented in a specific order, composing the abstract code  $\{1\}10^1, \{3\}10^0$ .

McCloskey (1992) illustrates abstract code theory’s numeric processing with the example of “64 x 59”. First, the system is presented with the stimulus input “64 x 59.” Next, the *comprehension* domain (also referred to as a "module" or dimension) recognizes the “x” symbol and activates a multiplication procedure in the *processing* domain. Next, the *processing* domain calls for input of the digits in the right (ones) column of the input, which the *comprehension* domain recognizes as Arabic numerals “4” and “9” and translates into abstract codes  $\{4\}10^0$  and  $\{9\}10^0$ . Then the *processing* domain retrieves the relevant arithmetic fact in abstract code,  $\{3\}10^1, \{6\}10^0$ . The *processing* domain then calls for the ones portion of the product to be written in Arabic numeral output, and the *production* domain translates this  $\{6\}10^0$  into the

Arabic numeral “6” to produce output in the ones column of the partial solution. The domains continue to use the multiplication procedure in this way, computing all partial products before calling for the addition procedure and finally producing solution output. This multistep, mechanistic account of numeric processing is argued to be a parsimonious cognitive model for solving arithmetic problems (McCloskey et al., 1985; McCloskey, 1992). These steps for calculation do not happen simultaneously, but are sequentially ordered and additive (not simultaneous or interactive; Campbell & Epp, 2005).

Empirical support for abstract code theory comes largely from case studies of adults with traumatic brain injuries in various regions of the brain, affecting language and arithmetic functioning. For example, McCloskey (1992) cites Benson and Denckla's (1969) case study of a man with left hemisphere trauma, who was able to comprehend numerals across various formats but could only produce correct arithmetic solutions given multiple choices, as evidence that numerical production and comprehension are distinct. Furthermore, McCloskey (1992) uses Singer and Low's (1933) case study of a man with brain trauma, who struggled with writing numeral greater than 2-digits using correct place value (place value being a syntactic property of number, why the numerals and their magnitudes are lexical properties), as evidence that the lexical and syntactic processes of *production* are distinct. In another example, Whalen, McCloskey, Lindemann, and Bouton (2002) reported on two patients with brain damage, who struggled with phonologically representing arithmetic information but were able to produce answers to arithmetic problems in Arabic numeral format, as evidence that the arithmetic facts used for numeric processing is language independent.

According to McCloskey (1992), because abstract, semantic codes are the object of numeric processing, formatting exerts no effect on numeric processing, save the time needed for

transcoding (encoding stimuli into abstract codes and recoding abstract code into output). All differences in reaction time seen with language-formatted arithmetic stimuli can be attributed to increased encoding time necessary for the *comprehension* domain to mentally represent the input (McCloskey et al., 1992). The extent to which a *language* domain may be involved in aiding the *comprehension* domain is unclear and not specified by the theory, but rather addressed as an area for future investigation (McCloskey, 1992). Similarly, the extent to which some *executive system of control* (regulation, attention, inhibition, working memory) is responsible for coordinating *comprehension*, *processing*, and *production* is not specified by the theory. Rather, as seen in McCloskey's (1992) example of arithmetic processing, abstract code theory tends to allow for the *processing* domain to facilitate the direction of other domains and the execution of arithmetic operations. McCloskey (1992) notes that the roles of general processing abilities (e.g., working memory) are issues for future investigation.

### **1.6.2            *Encoding complex theory.***

Encoding complex theory stipulates that the presentation of numerical stimuli activates an associative network of format-specific numerical “codes” or mental representations (Campbell, 1994; Campbell & Clark, 1988; Clark & Campbell, 1991). These format-specific mental representations are diverse. Mental representations of number can be verbal (e.g., articulatory, orthographic, motor-speech, and auditory mental representations of spoken or written number words, which are somewhat language specific and may be unique across populations of bilinguals and multilinguals) or nonverbal (e.g., visual, motor, analog magnitude, and combined visual-motor mental representations of digits, activities such as counting on fingers, and number lines; Campbell, 1994; Campbell & Clark, 1988; Clark & Campbell, 1991). The mental representations or “codes” are associatively connected within a complex network,

called the encoding complex, and as such, they are assumed to stimulate each other in complex patterns of activation without the use of a common, abstract code (Campbell & Clark, 1988; Clark & Campbell, 1991).

The notion of “transcoding,” or manipulation of one type of mental representation into another, is not applicable to encoding complex theory because multiple, format-specific codes are assumed to interact with each other in the encoding complex network. Similarly, the notion of “recoding,” or manipulation of mental representations into format-specific output, also is not applicable or necessary for encoding complex theory because format-specificity is inherent to these mental representations. Neither transcoding nor recoding is addressed in encoding complex theory (Campbell & Clark, 1988; Clark & Campbell, 1991).

Successful numeric processing (number comprehension, calculation, comparison, parity judgment) requires enhancing relevant association patterns and inhibiting interfering association patterns within the encoding complex network, and this is particularly true for calculation activities (Campbell & Clark, 1988; Clark & Campbell, 1991). The failure to inhibit associations that are irrelevant to the problem at hand ultimately results in difficulty achieving a correct response to the stimulus.

Campbell and Clark (1988; 1991) have drawn empirical support for encoding complex theory from a variety of methodologies; however, much of their own work has focused on behavioral studies of formatting effects on reaction time, accuracy, and quality of error propagation. In general, their findings support the ideas that (1) language formatting may increase reaction time and decrease accuracy, (2) problem size may increase reaction time and decrease accuracy, and (3) regardless of these main effects for certain characteristics of problems, format by operation by problem size interactions may occur (see for example,

Campbell, 1994). These problem-formatting effects are explained by interference from competing verbal codes and word stimuli, and ultimately, the system's failure to inhibit these competing responses (Clark & Campbell, 1991). The reason why these language-specific codes should be more susceptible to interferences and the role of the *language* domain in resolving interferences is unclear. Clark and Campbell (1991) have proposed that greater exposure to digit-formatted problems may increase system efficiency in resolving interferences for these types of mental representations.

Encoding complex theory is “integrative” (not modular) in that numerical processing is characterized by distinct domains that are specialized for numerical processing alone (Campbell & Clark, 1988; Clark & Campbell, 1991). Rather, the domains involved in numeric processing are assumed to contribute to a number of other cognitive activities. It is only when the system has enough practice to build “cognitive routines” for certain processes that inhibitory procedures might become automated enough to mimic modular cognitive architecture for quantity (Clark & Campbell, 1991). Thus, encoding complex theory does not specify a specific *quantitative* domain as being responsible for numeric processing. Instead, Campbell and Clark (1988; Clark & Campbell, 1991) have implicated a number of domain general cognitive capacities in resolving the complex network of associations of activated during numeric processing. These domains include *executive systems of control* (inhibition, problem-solving, attention, working memory, specifically, Baddeley and Hitch's 1974 model of working memory), the *motor* domain, the *language* domain, and the *visuo-spatial* domain.

It is also worth noting that culture, education, and individual differences can all impact the nature in which arithmetic is conducted in the encoding complex view of numeric processing. Encoding complex theory does not theorize a universal, human module for numeric processing.

Encoding complex allows for cultural variation in verbal and visuo-spatial procedures for calculation, and in fact anticipates that skilled calculators should be able to attempt problems using a variety of approaches (Clark & Campbell, 1991). There is no single mechanistic account for the process of calculation, estimation, comparison, or any other type of numeric processing task under encoding complex theory. Each of these processes is allowed to vary within individuals, across individuals, and across cultures.

### **1.6.3            *Triple code theory.***

Triple code theory stipulates that there are three, distinct, but interrelated domains responsible for encoding and mentally representing number and that these three domains are also responsible for numerical processing (mental arithmetic; (Dehaene & Cohen, 1995; Dehaene et al., 2003; Dehaene, 1992). According to triple code theory, (1) the *visual Arabic number form* domain is responsible for representing Arabic numeral input as visuo-spatial strings of digits, (2) the *verbal word frame* domain is responsible for representing spoken or written number words as sequences of words which are organized syntactically by place value, and correspond to the phonological and/or graphemic forms of words, and (3) the *analogical magnitude representation* domain is responsible for representing sets of visual or auditory objects as semantic mental representations of quantity, including the number's cardinality, its relationship to other quantities, its approximate or estimated value, and its position on an internal number line (which, following Weber's Law, becomes less precise as numbers increase in magnitude; Dehaene, 1992; Dehaene & Cohen, 1995). Importantly, triple code theory assumes that the semantic information for quantity is contained only in the analog magnitude domain of number representation (Dehaene & Cohen, 1995).

Each domain of triple code theory has anatomical correlates in the brain that support functioning, and these brain regions have been elaborated upon as triple code theory evolved. This research, although initially based in the case study reports of functional impairments in patients with brain trauma to various regions of the brain thought to be essential for number processing, has begun to consistently focus on the study of functional brain imaging of neurotypical adults exposed to various types of arithmetic stimuli (e.g., Dehaene & Cohen, 1995; Dehaene, Bossini, & Giraux, 1993). Currently, it appears that the *visual Arabic number form* is supported by the spatial attention network of superior, posterior parietal lobe, the *verbal word frame* is supported by the left angular gyrus and other left perisylvian areas, and the *analogical magnitude representation* is supported by the horizontal segment of the intraparietal sulcus (HIPS; Dehaene et al., 2003).

Triple code theory assumes that both transcoding and recoding occur for its domains. Transcoding is the process by which the three domains may share mental representations and quantity information. The semantic, quantity information for verbal or visual mental representation can be accessed from the *analogical magnitude representation* domain, and language-based or Arabic numeral representations for quantities can be accessed from the *verbal word frame* or *visual Arabic number form* domains (Dehaene, 1992; Dehaene & Cohen, 1995). Finally, the direct relationship between the *verbal word frame* and the *visual Arabic number form* allows for transcoding of word forms to visual forms and vice versa without processing semantic quantity representations (Dehaene, 1992; Dehaene & Cohen, 1995). Transcoding is necessary for numerical operations because it allows for processing of various input formats, accessing relevant verbal number facts, and accessing relevant semantic information about quantity. Recoding is the process by which “output routines” operate on mental representations

to produce stimulus output (e.g., written digits in the case of the *visual Arabic number form* and spoken or written words in the case of the *verbal word frame*; Dehaene & Cohen, 1995).

According to triple code theory, stimulus format does affect encoding and mental representation of number. The format in which number stimuli are presented will determine the type of mental representation encoded for them. Arabic numeral input is represented by the *visual Arabic number form*; language-based numeral input is represented by the *verbal word frame*; sets of objects are represented by the *analogical magnitude representation*. Although each of these domains is allowed to communicate directly with one another, problem demands influence the way in which numerical processing is conducted. Problems requiring comparisons, for example, require that semantic mental representations are accessed for both numerical inputs and answers are recoded into visual or linguistic output (Dehaene & Cohen, 1995). Problems requiring exact calculations, on the other hand, must be transcoded into *verbal word frame* in order for relevant number facts to be retrieved from verbal memory (Dehaene & Cohen, 1995). Under triple code theory, format-based differences in arithmetic performance are thus attributed to issues of efficiency in the transcoding process, and so transcoding may be considered at least somewhat additive (not simultaneous or interactive; Campbell & Epp, 2005).

The cognitive domains responsible for encoding and mentally representing numeric information are not the only domains involved in triple code theory's arithmetic. The language domain supports the recognition of spoken and written number input, the production of spoken and written number output, and the retrieval of number facts (e.g., two plus two equals four) from memory (Dehaene, 1992; Dehaene & Cohen, 1995). The role of executive systems in coordinating the functions of arithmetic is unclear in triple code theory. Although the three domains for the mental representation of number are assumed to cooperate with one another and

with the language domain in carrying out numeric processing, the extent to which their cooperation is self-directed as opposed to organized by a super ordinate system of attention, inhibition, working memory, and regulation is not specified by the theory. However, a visuo-spatial attentional circuit that appears to contribute to visuo-spatial attentional tasks (e.g., eye tracking, attention orienting, grasping, reaching, spatial working memory) and numerical processing tasks (e.g., comparison, estimation, subtraction, counting, multi-operation tasks) has been identified empirically (Dehaene, Piazza, Pinel, & Cohen, 2003). Dehaene has hypothesized that this region of the brain may aid in both the visual recognition of numbers and in the coordination of attention to quantities on the mental number line (Dehaene, Piazza, Pinel, & Cohen, 2003).

#### **1.6.4        *Exact versus approximate calculations: An extension of triple code theory.***

Unlike the other theories of arithmetic cognition reviewed thus far, exact versus approximate calculations theory is empirically generated and pertains specifically to the numeric processing task of calculations. It is an extension of triple code theory, supporting the idea that distinct neural networks contribute to (1) approximate calculation tasks involving semantic representations of quantity, comparison, and estimation versus (2) exact calculation tasks involving the retrieval of rote, verbal, numerical facts about quantity to compute exact arithmetic solutions (Dehaene et al., 1999; Stanescu-Cosson et al., 2000).

The *analogical magnitude representation* domain is hypothesized to be supported by the neural network for approximate calculations, and the *verbal word frame* domain is hypothesized to be supported by the neural network for exact calculations. These domains appear to be integrated, and they may both be recruited for difficult, exact calculation problems involving large quantities (Stanescu-Cosson et al., 2000). The visuo-spatial system implicated in the *visual*

*Arabic number form* domain of triple code theory is not a main focus of this extension of triple code theory; however, empirical evidence suggests that visuo-spatial networks involved in both numerical and non-numerical processing tasks may contribute to internal, mental representations of numbers during both approximate and exact calculation (Stanescu-Cosson et al., 2000).

Empirical support for the exact versus approximate extension of triple code theory are based mostly in research with adults who have verbal or quantity impairments as a result of traumatic brain injuries and brain imaging research with healthy adults performing various types of calculations. For example, Dehaene & Cohen (1991) reported a case study of man who had suffered severe head trauma to the right temporo-parieto-occipital region of his brain and associated acalculia and aphasia. Because this participant was able to correctly judge the correctness of approximate quantity calculations (e.g.,  $2 + 2 = 9$ ) but struggled with very simple exact calculations (e.g.,  $2 + 2 = 3$ ), Dehaene and Cohen (1991) hypothesized that there were two, distinct networks involved in calculation activities. Although the language-dependent exact calculation network was impaired, the specialized network for quantity approximation tasks remained intact. Lemer, Dehaene, Spelke, and Cohen (2003) reported similar results for two adults with traumatic brain injuries and associated aphasia and acalculia. The participant with left fronto-temporal atrophy and associated aphasia struggled with oral language comprehension and production, narration, word-finding, coherent speech, and exact calculations. The participant left intraparietal lesion and associated acalculia and apraxia struggled with visuo-spatial processing and approximate calculations.

Brain imaging studies, frequently relying on both fMRI and ERP methods have identified distinct neural networks and patterns of activation during exact versus approximate calculation tasks. Specifically, it appears that bilateral parietal and frontal regions of the brain, particularly

the intraparietal sulci, are consistently activated during both exact and approximate calculation tasks, but display higher levels of activation during approximate calculation tasks (Dehaene et al., 1999; Stanescu-Cosson et al., 2000). During exact calculation tasks, the left anterior inferior frontal regions of the brain, particularly the bilateral angular gyri, are consistently activated, suggesting that regions of the brain which are implicated in language processing tasks (e.g., word associations) also contribute to exact calculation tasks (Dehaene et al., 1999; Stanescu-Cosson et al., 2000).

Behavioral studies of reaction time and accuracy support the distinction between exact and approximate calculation activities and triple code theory's hypotheses about format effects on subsequent mental representation. For example, Dehaene and colleagues (1999) reported that Russian-English bilinguals who were taught 2-digit exact and approximate number facts in one of their languages (1) performed faster in the teaching language than in the untrained language for exact calculation facts, (2) performed equivalently in both languages for approximate calculation facts, and (3) performed similarly on trained facts and novel problems with operands of similar magnitudes when doing approximate calculations. These results were interpreted to support the ideas that (1) exact calculation facts were stored in language-specific codes and switching between languages resulted in a reaction time cost, (2) approximate calculation facts were stored in codes that were not language-specific and code switching between languages did not result in reaction time cost, and (3) approximate facts were stored in magnitude formats such that their information could generalize to novel problems involving similar magnitudes without reaction time costs.

Other assumptions of triple code theory, including the possible cognitive domains involved in numeric processing are generally not addressed in the empirical literature supporting

exact versus approximate calculations. The focus of this empirically generated theory is specifying the roles of the *analogical magnitude representation* domain and the *verbal word frame* domain on approximate and exact calculation activities.

## **2 CHAPTER 2: MODELING ARITHMETIC PERFORMANCE AND THE ROLE OF PROBLEM FORMATTING**

Although Abstract Code Theory, Encoding Complex Theory, Triple Code Theory, and the Exact versus Approximate Calculations specification of Triple Code Theory overlap in many areas, they also diverge in their explanations of mental representation of quantity and cognitive domains responsible for numeric processing. Encoding Complex Theory and Triple Code Theory both agree that stimulus formatting can largely influence both mental representation of quantity and subsequent numeric processing; however, Abstract Code Theory stipulates that regardless of stimulus format, mental representations are amodal abstract codes and subsequent numeric processing relies on these abstract codes. Triple Code Theory and Abstract Code Theory both agree that numeric processing relies on cognitive domains specialized for processing quantity; however, Encoding Complex Theory stipulates that numeric processing relies on cognitive domains which are not modular and not unique to processing quantity. In terms of specifying domains which may help to facilitate numeric processing, Abstract Code Theory is largely silent, but both Encoding Complex Theory and Triple Code Theory agree that executive domains (involving coordinating attention and inhibition) and the language domain (retrieving verbal information about number facts) may contribute. Clearly, encoding (forming mental representations) and cognitive dimensionality of numeric processing are major areas of departure for these theories. From a psychometric perspective, the issue of encoding is closely related to the issue of dimensionality because depending upon the theoretical perspective one takes,

encoding and mental representation may influence the cognitive domains involved in subsequent numeric processing. In the current chapter, each of these theories will be presented as confirmatory factor models in order to clarify the roles of format, problem demands, and cognitive abilities other than those in the quantitative domain.

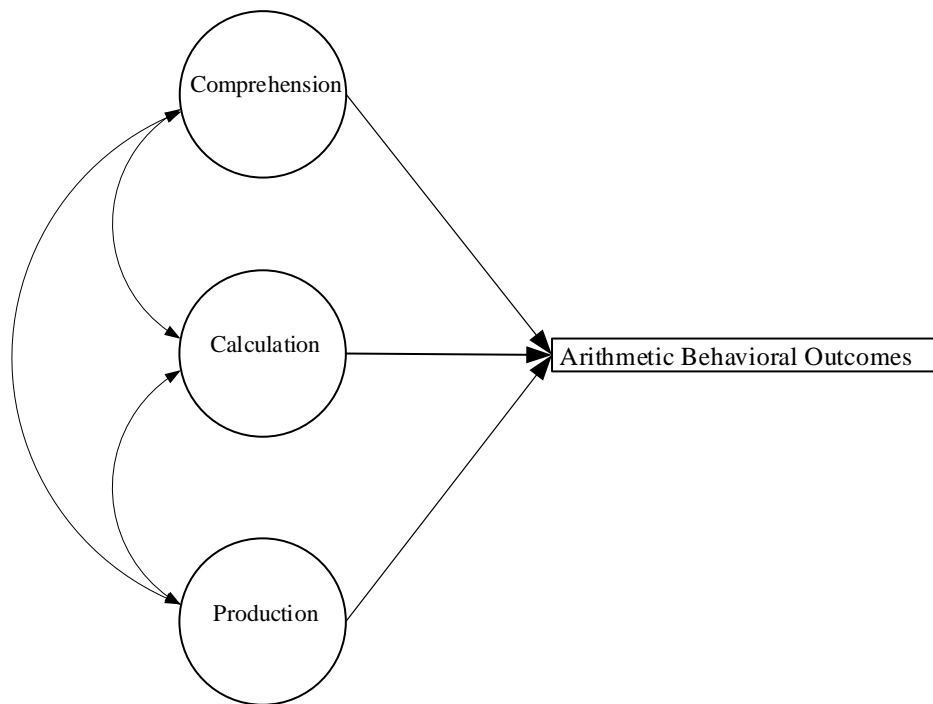
## **2.1 Modeling Leading Theories of Arithmetic Cognition**

The next sections will present factor analytic specifications of four leading theories of arithmetic cognition, (1) Abstract Code Theory, (2) Encoding Complex Theory, (3) Triple Code Theory, and (4) the Exact versus Approximate Calculations specification of Triple Code Theory. Each theory will be examined for its specifications of the role of problem formats in encoding and calculation, and possible factor structures to represent each theory's specifications will be provided.

The strength of the factor analytic framework lies in its specificity. Confirmatory factor analysis forces explicit statements about model parameters and the hypotheses they entail (Bollen, 1989; Brown, 2006; Kline, 2011; McDonald, 1999). It is a method that can reveal theoretical misspecifications by forcing explicit tests of relations (e.g., exact and approximate calculations are not predicted by the same cognitive architectures). It can also reveal areas in which theories have not provided hypotheses about possible relationships by forcing users to specify falsifiable relations (e.g., language does not relate to arithmetic behavioral outcomes). In sum, the specific, explicit nature of confirmatory factor analysis forces researchers to consider the testable dimensions of a theory. For example, if theoretical constructs do not have observable outcomes, they are not testable with factor analysis (and perhaps not with any other method).

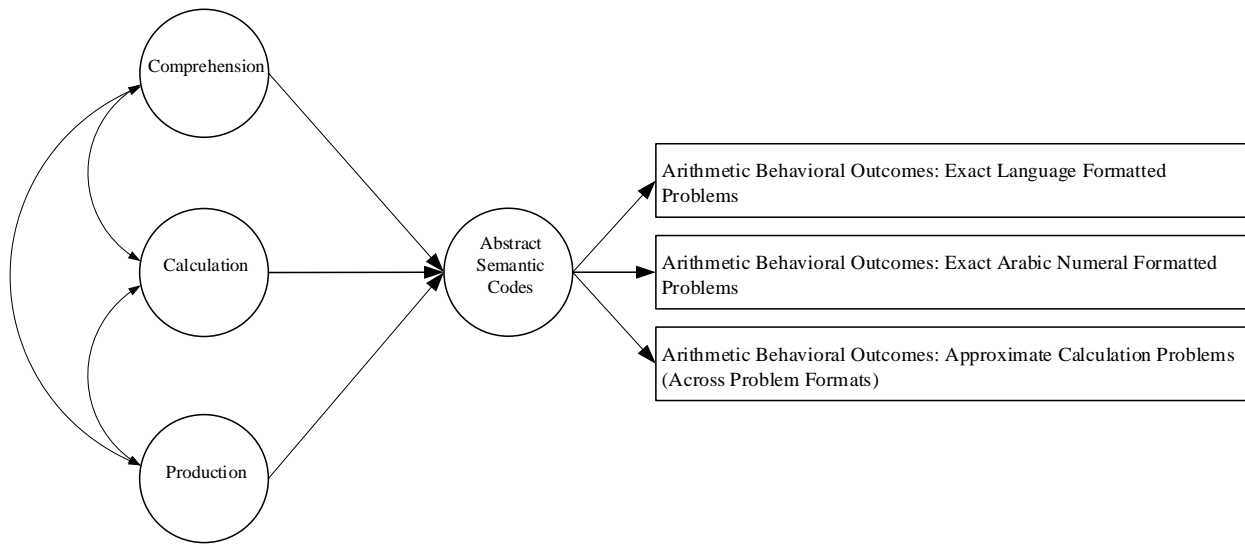
### 2.1.1 *Abstract code theory.*

Within the framework of Abstract Code Theory, three cognitive systems are responsible for processing numerical information (McCloskey, 1992). A *comprehension* system encodes stimuli into *abstract semantic representations*. A *calculation* system accesses arithmetic facts, rules, and complex procedures using those *abstract semantic representations*. Finally, a *production* system recodes *abstract semantic representations* back into verbal or written output. These three domains communicate and work together collaboratively to execute numeric processing. In a factor model, each of these domains could be represented by latent variables, and correlations between the three latent variables could represent their communication as shared variance in predicting arithmetic behavioral outcomes. As a schematic factor diagram, the latent variables are represented by circles; the latent variable shared variance or communication, by curved arrows; the observed outcomes, by rectangles (in this case one rectangle is used to represent all possible behavioral outcomes); and the assumption that latent variables predict observed outcomes, by straight arrows.



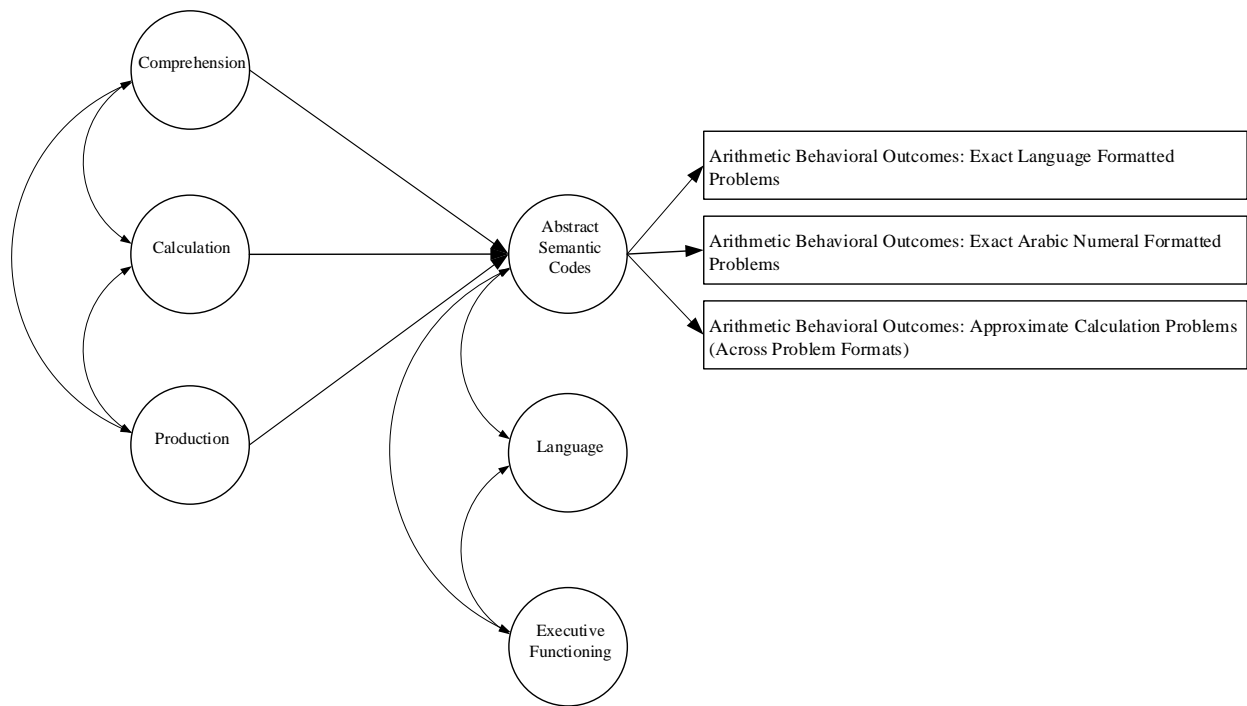
**Figure 1. Abstract Code Theory: 3 Modules for Numeric Processing.**

Furthermore, Abstract Code Theory postulates that each of these three dimensions for numeric processing rely on a single, amodal, mental representation of number. Regardless of stimuli formatting and problem demands (e.g., exact versus approximate calculations), the fact, rule, and procedure mechanisms at work during the calculation stage of cognitive processing are reliant upon *abstract semantic representations* independent of encoding and recoding processes. Stimuli format should not affect *calculation*. The specification that one, latent form of mental representations predicts arithmetic behavioral outcomes across a variety of stimulus formats and problem demands can be represented as a factor model in which various stimulus formats have no distinct common method variance and instead are predicted by one, latent dimension.



**Figure 2. Abstract Code Theory: 1 Mental Representation Regardless of Stimulus Format.**

Abstract Code Theory specifies that the quantitative domain outlined above is specialized for numeric processing. The roles of other domains in helping with language processing, language-based fact retrieval, or coordinating the activities of numeric processing are not specified. From a factor analytic framework, the roles of a language and executive functioning domain in Abstract Code Theory could be modeled as separate latent variables which are allowed to correlate with the numeric processing domain but are not involved in predicting arithmetic behavioral outcomes. The extent to which other cognitive domains may or may not correlate with various facets of the numeric processing domain is not addressed by Abstract Code Theory; however, because *abstract semantic representations* are the common form of mental representation upon which all three modules of numeric processing operate, one would expect that this latent variable, at a minimum, should be allowed to correlate with other cognitive domains.

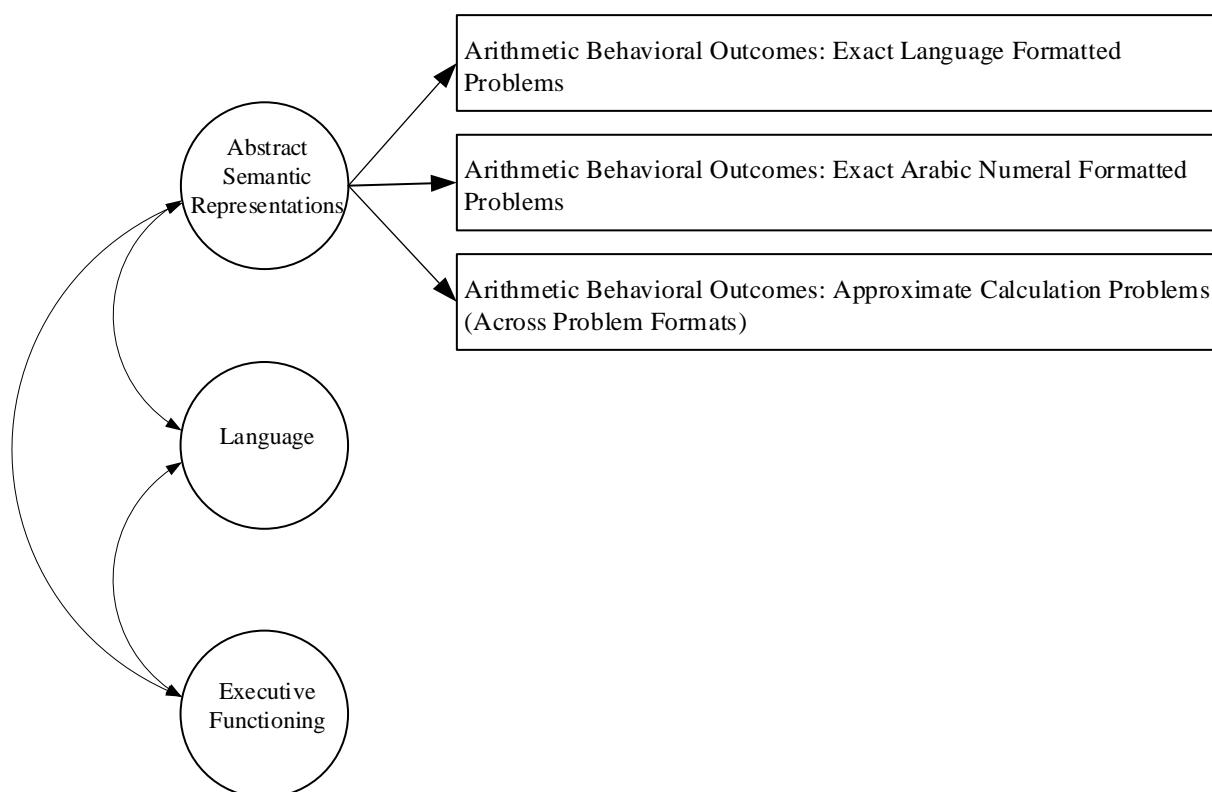


**Figure 3. Abstract Code Theory: Other Domains Do Not Predict Arithmetic Behavioral Outcomes.**

It is important to note that Abstract Code Theory specifies arithmetic problem processing as occurring mechanistically in many, unobservable stages of mathematical cognition. The production of oral, written, perhaps even gestured responses from *abstract semantic codes* may be the only stage of this theory to produce observable behaviors in practice. The theory's additional internal stages are not directly testable using a methodology which relies upon behavioral observations (or perhaps any currently available methodology). However, the hypothesis that all stimuli are encoded, operated upon, and responded to as *abstract semantic codes*, regardless of original stimulus format, can be modeled and tested from a factor analytic measurement standpoint. One would expect that across all stimulus formats, no common method variance effects for formatting should be observed.

Also worth noting is the fact that Abstract Code Theory, like all of the theories of arithmetic cognition considered thus far, pertains to the skilled arithmetic cognition of adults. The extent to which children or other novice numeric processors may differ in the structure(s) of their quantitative domain is not specified by Abstract Code Theory. Without those developmental specifications, one must assume that individuals who are developing numeric processing (i.e., children, persons without access to formal education, unskilled adults) have the same cognitive architecture as skilled adults, an assumption which is perhaps untenable.

Thus, with these caveats in mind, Abstract Code Theory may be best represented with a one factor model of *abstract semantic representation*, which at a minimum, is allowed to correlate with other cognitive domains (e.g., *language*, *executive functioning*). Here, *language* and *executive functioning* are not allowed to predict arithmetic behavioral outcomes, and so, their predictions are fixed at zero (and not drawn) across formats and problem demands. The *comprehension*, *calculation*, and *production* modules may operate on and with *abstract semantic codes*; however, they are not formulated to predict unique variance in specified behavioral outcomes, and therefore, their dimensionality separate from *abstract semantic codes* is not testable.



**Figure 4. Abstract Code Theory: General Factor Model.**

### 2.1.2 *Encoding complex theory.*

Encoding Complex Theory stipulates that when mathematics problems are presented, a diverse network of mental representations called an *encoding complex*, is activated by numerical stimuli (Campbell, 1994). The *encoding complex* associations can involve number reading, fact retrieval, procedural operations, comparison, estimation, and the elimination of similar but irrelevant semantic representations of math facts. The mental representations of the *encoding complex* can be verbal or nonverbal, and they can stimulate and interact with each other. This *encoding complex* is not modular, and it is not specialized for numeric processing; however, practice with a given format can reinforce associations in the *encoding complex*, increasing

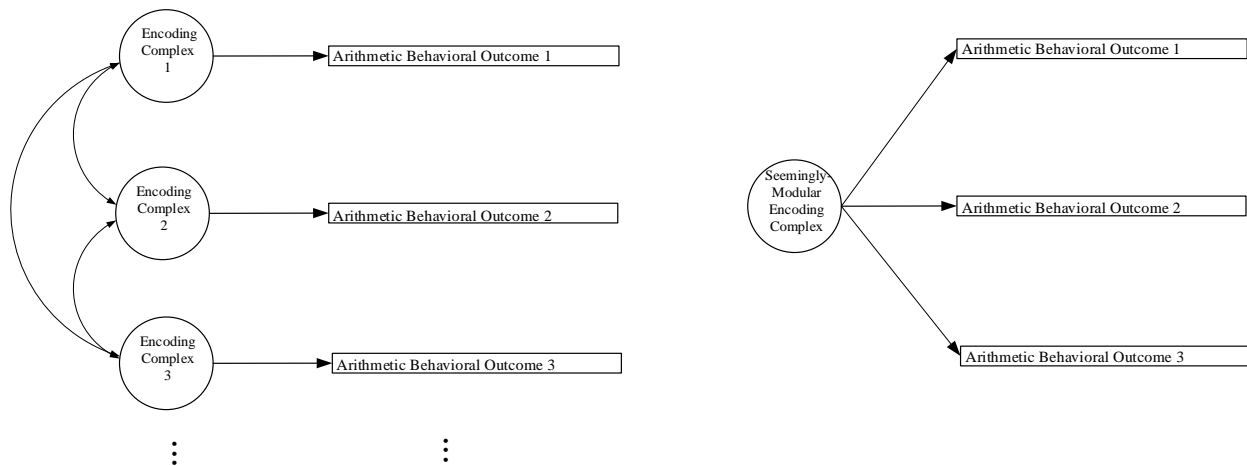
skilled problem-solving and allowing the *encoding complex* to mimic a specialized numeric processing module.

Encoding Complex Theory, like all of the theories of arithmetic cognition reviewed thus far, is specified to model highly-skilled, adult arithmetic cognition. Although Encoding Complex Theory is one of the few arithmetic cognition theories to address changes in the system as a function of practice, the extent to which the *encoding complex* begins to appear modular over the course of development (and various types of formal or informal practice with numeric stimuli) is not specified by the theory, and this issue is crucial for applications of the theory to developing children.

One might imagine two extremes for the question of modularity in Encoding Complex Theory, (1) in individuals who have relatively little practice with numeric stimuli such as infants, the *encoding complex* may be different for every problem stimulus they encounter, as opposed to (2) in individuals who have some unspecified amount of practice with numeric stimuli such as children and adults with formal schooling experience, the *encoding complex* may have become seemingly-modular for every problem they encounter. The gradients of the seemingly-modular *encoding complex* as a function of "practice" are unclear.

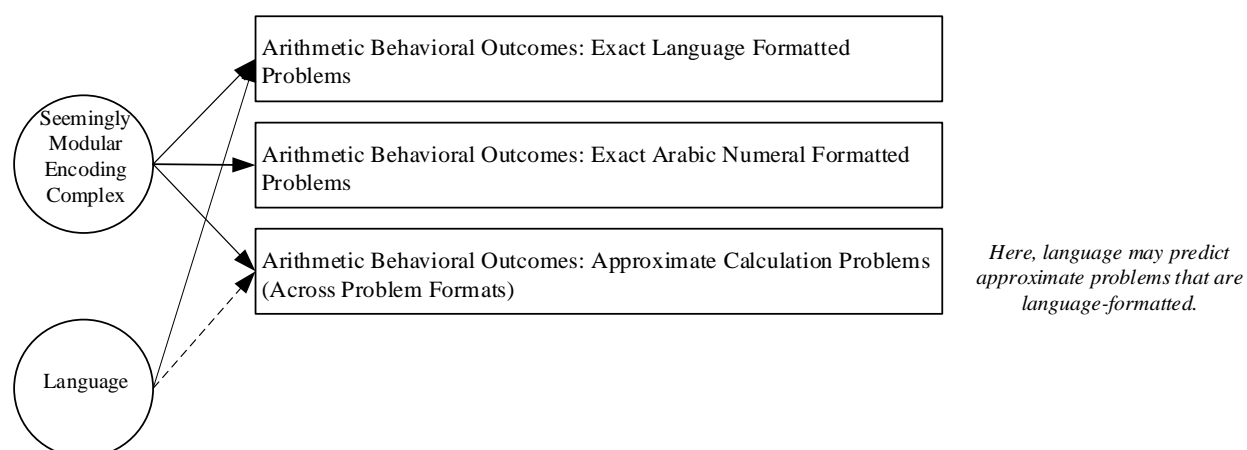
From a factor analytic framework, the first extreme could be represented by a model in which several, distinct latent variables (the *encoding complexes*) predict arithmetic behavioral outcomes for every numeric stimulus and no, one latent variable shares predictive value across arithmetic problems. These *encoding complexes* are allowed to covary (or not covary), and to the extent that covariances increase over time (and with practice) these separate *encoding complexes* may begin to converge. The second extreme could be represented by a model in which one latent variable (the "practiced" and seemingly-modular encoding complex) predicts arithmetic

behavioral outcomes across various arithmetic problems. One would expect the second extreme to apply to individuals who have at least some amount of practice with arithmetic problem-solving.



**Figure 5. Encoding Complex Theory: No Specialized Module for Numeric Processing → Becomes A Seemingly-Modular Encoding Complex with Practice.**

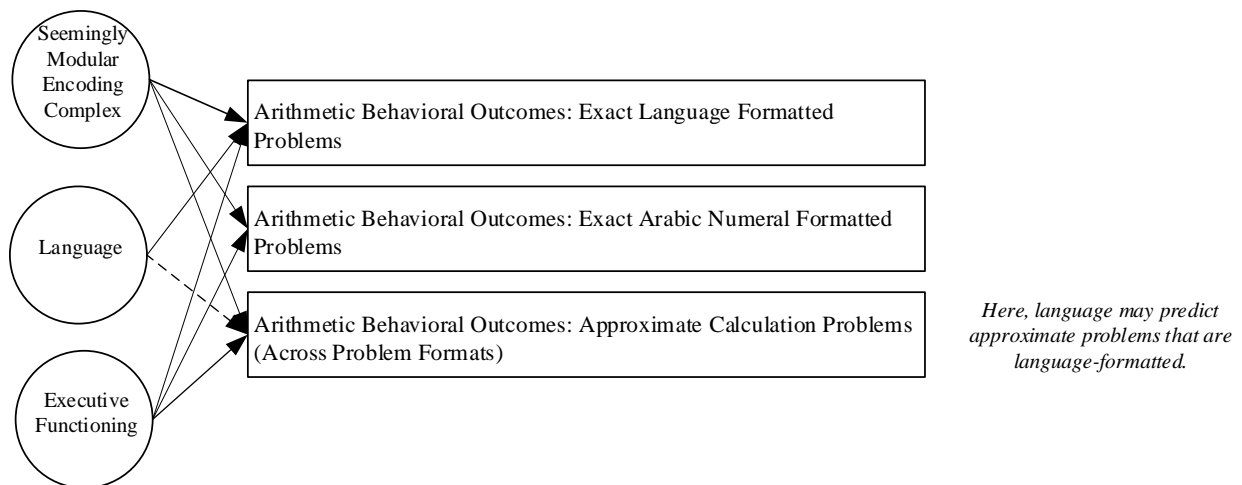
Furthermore, Encoding Complex Theory specifies that format interactions (e.g., language-based format by number size interactions) exist and can result in longer reaction times for correct responses and differential patterns of error production (Campbell, 1994). Format is not independent of calculation efficiency. Format can affect both mental representation of numbers and subsequent numeric processing, and this is especially true for language-formatted problems. The specification that language-formatted problems may display some common method variance can be represented as a factor model in which language-formatted items are predicted by a separate, format-specific, latent dimension (in this case, the *language* domain).



**Figure 6. Encoding Complex Theory: Language Formatting Effects As Language-based Common Method Variance.**

Encoding Complex Theory specifies that the quantitative domain outlined above is not specialized for numeric processing and is not modular (meaning that it is not a self-contained, separate cognitive domain or dimension of ability) although it may appear to be modular with practice. The competing and sometimes interfering responses to the stimuli must be sorted for relevance, and any interference must be overcome in order for successful performance to occur. The task of arithmetic is largely to inhibit competing and irrelevant signals activated in the encoding complex and to enhance signals that are relevant to the problem. Failure to successfully perform arithmetic constitutes a failure of the system to inhibit. Cognitive domains involved in this process are assumed to contribute to other cognitive activities, and Campbell and Clark (1988; Clark & Campbell, 1991) have suggested that domains such as *executive systems of control* (working memory, inhibition, attention) and the *language* domain (among others) may help to resolve the conflicting signals activated in the encoding complex. Although Encoding Complex Theory has also suggested that the motor and visuo-spatial domains may also predict

arithmetic behavioral outcomes, the extent to which these domains predict outcomes across various formats and the extent to which they relate to other cognitive domains involved in arithmetic processes is unclear and not specified by the theory. However, from a factor analytic framework, the roles of *language* and *executive functioning* could be modeled as separate latent variables which are allowed to predict arithmetic outcomes along with the *seemingly modular encoding complex for arithmetic*. As previously outlined, the *language* domain is expected to contribute to language-formatted problems. The *executive functioning* domain is expected to contribute to arithmetic behavioral outcomes regardless of problem formatting. Allowing these domains to correlate with the *seemingly modular encoding complex* for arithmetic in addition to predicting arithmetic outcomes would constitute an over-specification of the model; however, to the extent that these cognitive domains do not predict arithmetic outcomes, they could also be allowed to correlate with the *encoding complex*.



**Figure 7. Encoding Complex Theory: Other Domains Predict Arithmetic Behavioral Outcomes.**

Importantly, the stipulation that the diverse network of mental representations (which constitute the encoding complex) are allowed to stimulate and interact with each other is not modeled. The entire network of possible numerical associations cannot be measured behaviorally without additional specifications about exactly which kinds of mental representations would be expected for various arithmetic problems and how these mental representations relate to arithmetic behavioral outcomes (i.e., what exactly is activated for various types of numeric stimuli).

Also not modeled is the stipulation that item by person interactions may occur in the encoding complex such that encoding complexes are unique across items, within individuals, across individuals, and across cultures. Although factor analysis allows for latent variables to vary within and across individuals and for measurement models to be compared across groups of individuals (e.g., cultural groups), it would be exceedingly difficult to model a cognitive system which is structurally different for all individuals in a population using factor analysis.

Similarly, the extent to which this encoding complex becomes seemingly modular at various points of development is not specified by the theory, and modeling latent dimensions which are unique for every arithmetic behavioral outcomes is untenable. At most, we might assume that for all individuals who have some unspecified amount of practice with arithmetic problems, the seemingly modular encoding complex architecture is in place and is a single latent factor along which each person might have a unique value.

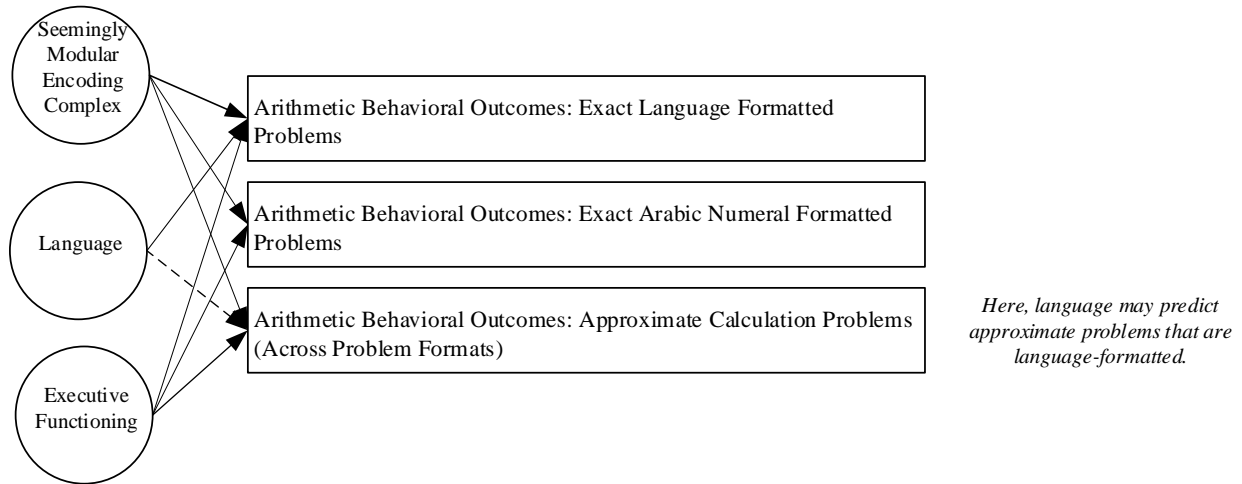
Thus, with these caveats in mind, the hypotheses that (1) numerical stimuli are represented in a single, *seemingly modular encoding complex*, (2) formatting may affect performance on arithmetic items, particularly for language-formatted items, and (3) additional cognitive domains of *language* and *executive functioning* may predict performance, can be tested

with behavioral responses to arithmetic problems. From a factor analytic measurement standpoint, Encoding Complex Theory may be best represented by a single mathematics *encoding* factor which indicates the semantic memory associations involved in numerical processing. Additional factors representing *language ability* and *executive functioning* may also impact the performance on mathematics problems, and these factors are allowed to predict arithmetic behavioral outcomes across various formats and problem demands. Mathematics problems may also demonstrate difficulty in predictable patterns such that items with language-based formatting are more difficult than items formatted with Arabic numerals and items with larger numbers are more difficult than items with smaller.

Importantly, in this model (presented below as a schematic) the *seemingly modular encoding complex* predicts no unique behavioral outcomes. Given that Clark and Campbell (1991) have specified that the *quantitative* domain is not a modular domain and that the task of successful arithmetic performance is successful inhibition of signals and responses irrelevant to solving the problem, the inclusion of *executive functioning* in the model may leave no unique variance for a *seemingly modular encoding complex*. In other words, to the extent that arithmetic performance simply constitutes successful executive control (inhibition, attention, working memory), including a *quantitative domain* which is responsible for explaining the majority of shared variance across arithmetic behavioral outcomes may be of little utility.

Because each of the predictive relationships between latent factors and behavioral outcomes is falsifiable, the role of *executive functioning* as opposed to a *seemingly modular encoding complex* (quantitative semantic representations activated by arithmetic problems) can be examined by first allowing for the possibility of a *seemingly modular encoding complex* to have predictive value above and beyond *executive functioning*. However, to the extent that

*executive functioning* (and possibly *language*) is the major predictor of these behavioral outcomes, the *seemingly modular encoding complex* (and possibly *language*) may disappear from the model.



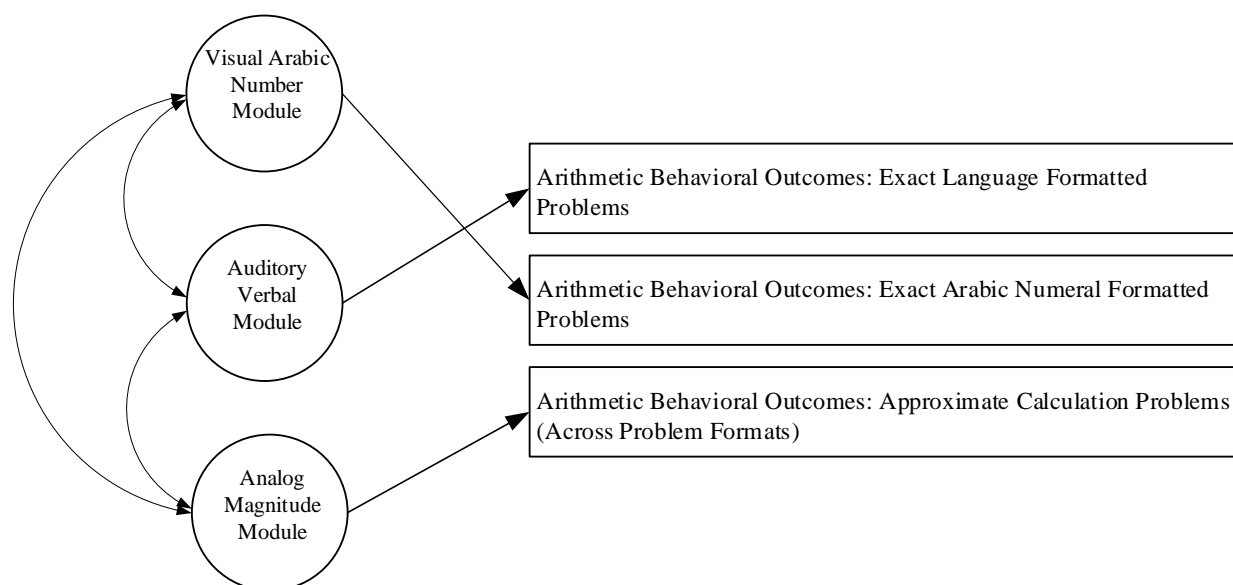
**Figure 8. Encoding Complex Theory: General Factor Model.**

### 2.1.3 *Triple code theory.*

Within the framework of Triple Code Theory, three cognitive modules are responsible for encoding, retrieving and processing mathematical tasks (Dehaene, 1992; Dehaene & Cohen, 1995). The *visual Arabic* module processes digital input and output and multi-digit operations. The *auditory verbal* module processes simple arithmetic facts, written and spoken input and output, and language-based memory of numbers. The *analog magnitude representation* module processes semantic numeric content, comparison, estimation, approximate calculation, and subitizing tasks. Measurement stimuli are encoded in the appropriate numeric module, where processing and calculation largely occur; however, problem demands may necessitate that a module calls upon another module of Triple Code Theory in order to complete numeric

processing. This communication or collaboration between modules is accomplished by transcoding, and each of the three domains of Triple Code Theory are allowed to communicate with each other directly and without the need for common abstract codes.

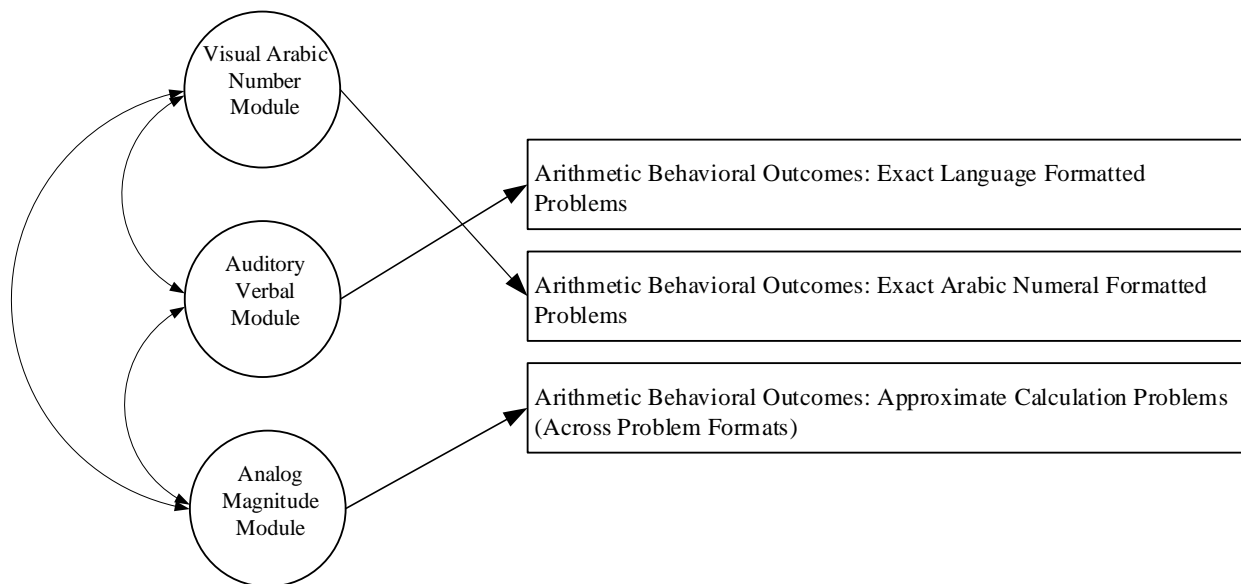
From a factor analytic framework, Triple Code Theory can be represented with a three factor model of arithmetic cognition in which (1) a visual Arabic factor is largely responsible for Arabic numeral formatted problems, (2) an auditory verbal factor is largely responsible for language-formatted problems, and (3) an analog magnitude factor is largely responsible for approximate calculations across formats. The communication between these factors, transcoding, can be represented with factor correlations.



**Figure 9. Triple Code Theory: 3 Modules for Numeric Processing.**

As mentioned above, Triple Code Theory allows for format to influence encoding, processing, and recoding of mental representations. Both formatting and problem demands may influence numeric processing, and formatting effects may be especially evident in the efficiency

of encoding and recoding. From a factor analytic framework, common method variance may be the result of problem format, problem demands (exact versus approximate), or both, and the predicted patterns of common method variance are outlined in Triple Code Theory's specifications about which domains should be largely responsible for which tasks.

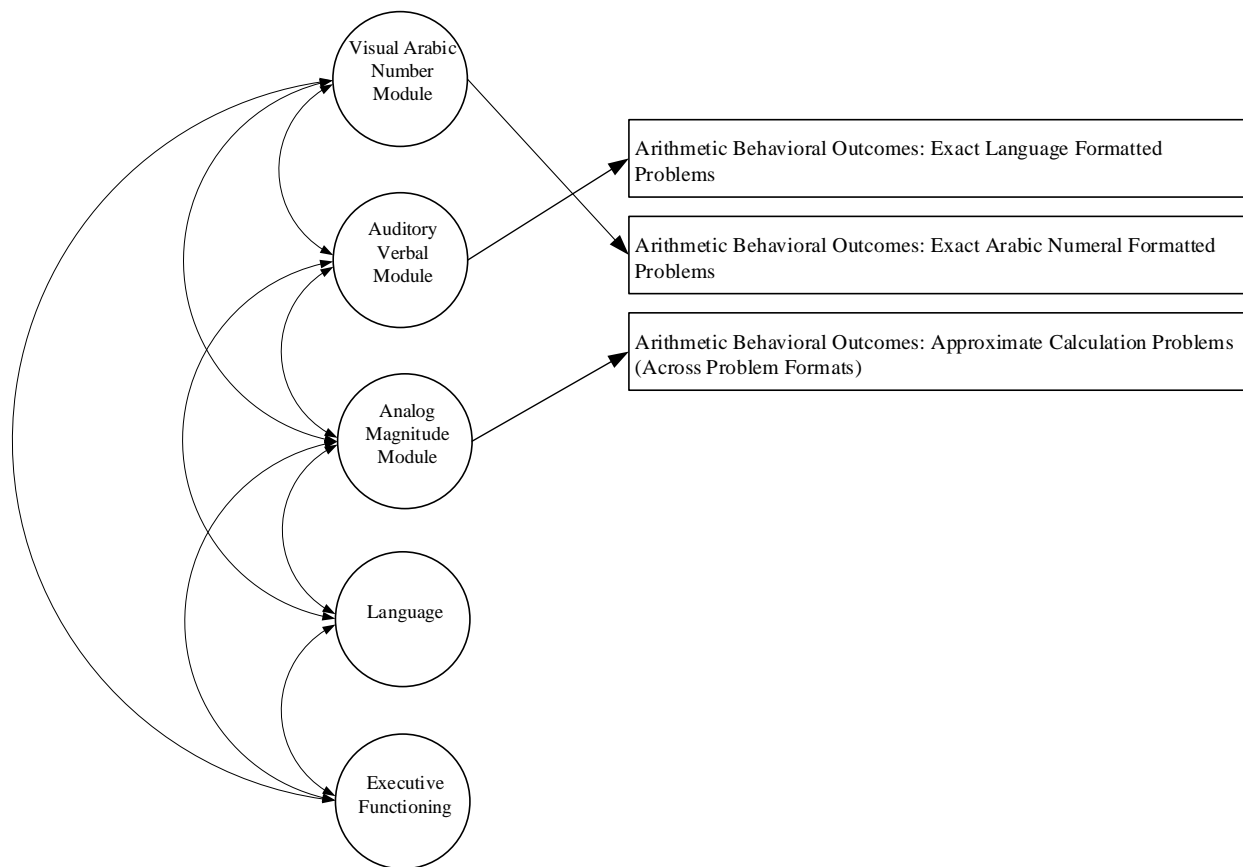


**Figure 10. Triple Code Theory: Formatting Can Effect Mental Representation & Processing (inherent in the theory).**

Triple Code Theory allows for domains other than the quantitative domain to facilitate numeric processing. The *language* domain is allowed to inform the *quantitative* domain by providing linguistically stored math facts. Although the *auditory verbal* module is responsible for mentally representing written (graphemes) or spoken (phonemes) numbers syntactically, by place value, the extent to which the *language* domain may or may not overlap with the *auditory verbal* module of Triple Code Theory is unclear.

Similarly, Triple Code Theory is a bit vague in its specification of which cognitive domains help to coordinate numeric processing and to control the complex sub-processes of numeric processing (like transcoding). The extent to which these subprocesses may be self-directed is unclear; however, an *attentional control* domain is allowed to coordinate visuo-spatial attention to numbers on the internal number line. The extent to which this *attentional control* domain helps to coordinate the working memory, inhibition, and planning required to complete numeric processing is not specified.

From a factor analytic framework, a latent *language* factor and an *executive control* factor could be added to the previously specified model. Because these domains may communicate with the three modules of Triple Code Theory's *quantitative* domain, at a minimum these additional domains may correlate with the numeric processing domains of Triple Code Theory. To the extent that the *auditory verbal module* and the *language* domain correlate, they may not be separate domains (i.e., if they correlate highly or at unity). To the extent that the *executive functioning domain* correlates with the modules of Triple Code Theory, it may not be helping to facilitate numeric processing by coordinating control (i.e., if it does not correlate significantly).



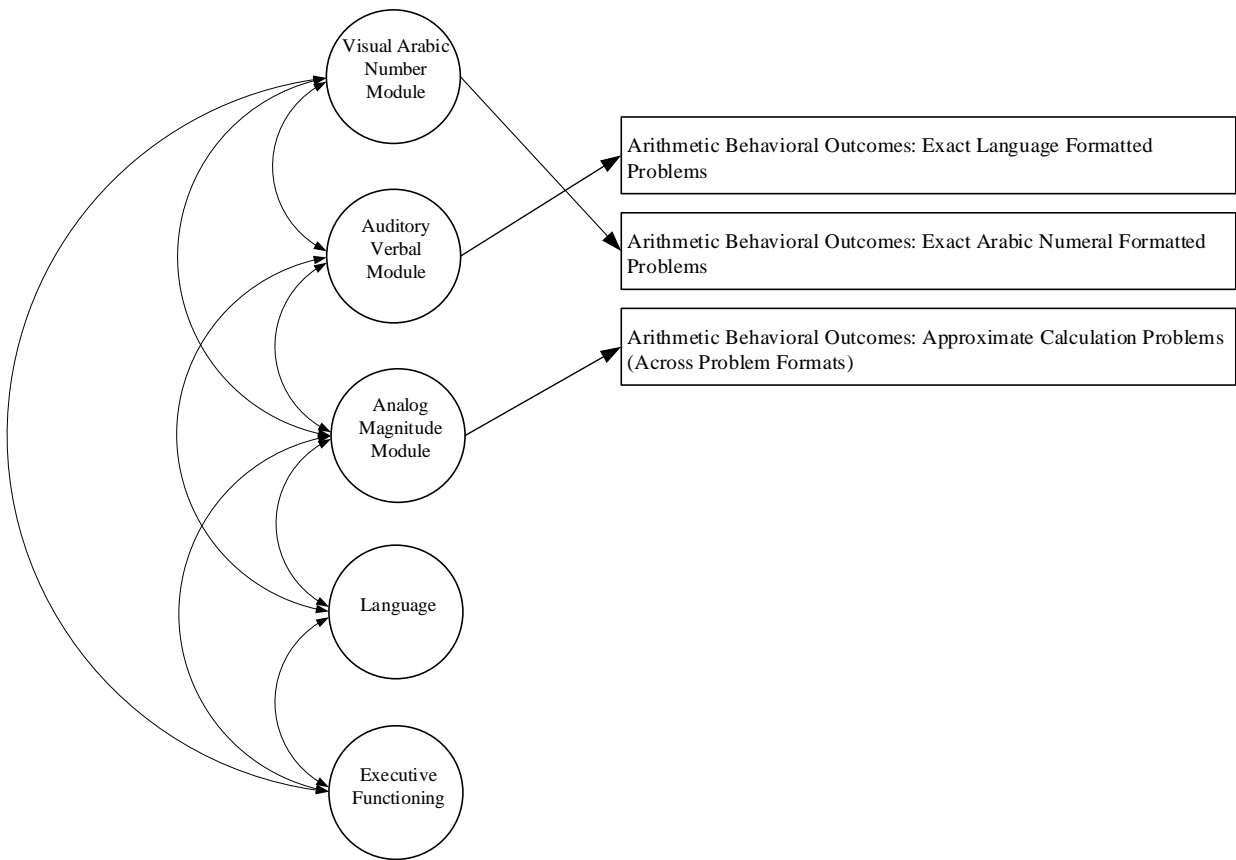
**Figure 11. Triple Code Theory: Other Domains May Be Associated With 3 Modules in Predicting Arithmetic Behavioral Outcomes.**

It should be noted that the specification of transcoding in the suggested factor model for Triple Code Theory is tenuous. The extent to which the three modules of Triple Code Theory may contribute to numeric processing depending upon problem demands is unclear. From one extreme, we might expect that the three domains remain relatively separate in predicting their various arithmetic outcomes and communicate only via transcoding. This notion of transcoding is represented in the proposed factor models with factor correlations. However, from another extreme, we might expect that although one domain is primarily responsible for certain tasks, other domains of Triple Code Theory may also directly predict outcomes (i.e., transcoding may

be best represented with latent factor loadings and not latent factor correlations). For example, although the *auditory verbal* module may be largely responsible for processing language-formatted problems, the *analog magnitude* module may also predict these outcomes. Because these specifications have not been made *a priori* by Triple Code Theory, they are not hypothesized here. Given that the latent correlation model of Triple Code Theory fails, this post hoc model of transcoding may need to be explored.

Also noteworthy is that Triple Code Theory does not specify developmental effects for numeric processing. Research suggests that hemispheric lateralization may become more uniform with age and that the process of arithmetic may be more integrated for adults, though largely relying on the same circuits involved in arithmetic during childhood; however, Triple Code Theory does not postulate these developmental effects. Like the other theories of arithmetic cognition considered here, Triple Code Theory pertains to adult arithmetic cognition, and the pathway(s) from childhood to this model are not considered.

With these caveats, Triple Code Theory can be represented as a factor model in which three, separate but related domains (*visual Arabic number*, *auditory verbal*, and *analog magnitude* modules) predict format-specific mental representation and calculation, and *language* and *executive functioning* domains may also facilitate arithmetic.

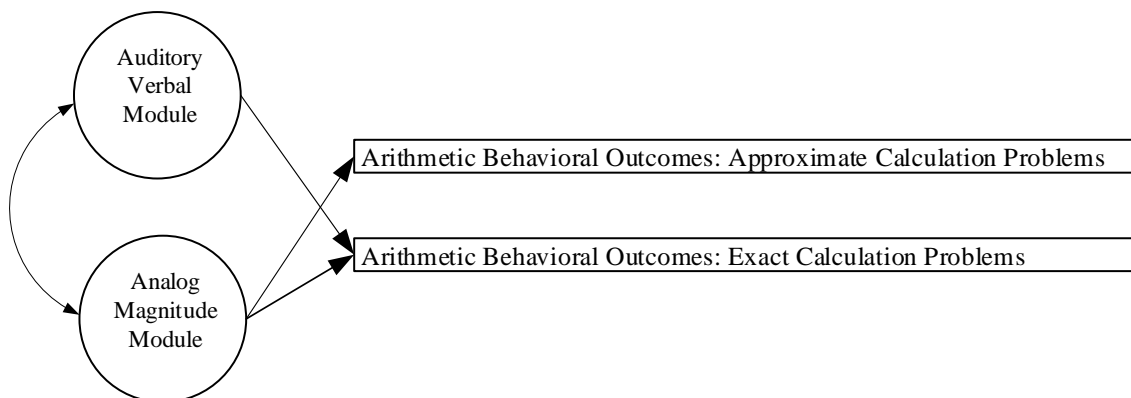


**Figure 12. Triple Code Theory: General Factor Models.**

#### **2.1.4**      *Exact versus approximate calculations specification of triple code theory.*

The Exact versus Approximate Calculations hypothesis is an extension of Triple Code Theory which stipulates that stimulus demands may differentially affect subsequent numeric processing. Problems which require exact calculations may call on both an *analogical magnitude* module, responsible for representing semantic information about quantity and the *auditory/verbal* module, contributing verbally stored information about number facts. Exact calculation problems would be expected to call on the *analog magnitude* module when they are not stored as facts in the *auditory verbal* module (e.g., if the facts have not yet been learned or involve numbers and operations which are not commonly executed). Problems which require

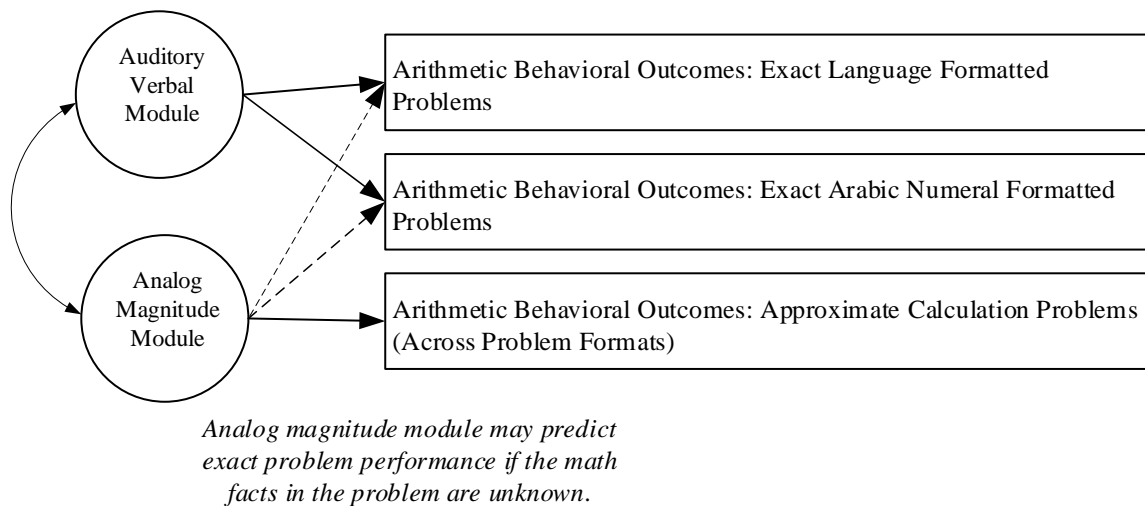
estimation or approximate calculations, on the other hand, may call only on the *analogical magnitude* module. The *visual Arabic number form* module is largely absent from this specification of Triple Code Theory; however, spatial attention networks, possibly representing some of the predictive power of the visual Arabic number form module and possibly representing some form of executive control for attention, may contribute to coordinating both types of task. The core premise of Exact versus Approximate Calculations Theory can be represented with a factor model in which two latent factors (representing the *analog magnitude* module and the *auditory verbal* module) predict arithmetic behavioral outcomes for exact and approximate problem demands. At a minimum, these latent factors can be allowed to correlate and communicate with one another.



**Figure 13. Exact V. Approximate Theory: 2 Modules for Exact and Approximate Calculations.**

Importantly, Exact Versus Approximate Calculations Theory does not specify predictions for problem formatting. As an extension of Triple Code Theory, it is mainly concerned with specifying domains responsible for exact and approximate problem demands. The extent to which the *analog magnitude* and *auditory verbal* modules differentially predict performance on

various formats of items is unclear. Without further specification of the theory, it would appear that Exact Versus Approximate Calculations Theory would predict that across various problem formats, problems requiring exact calculations will be largely predicted by both the *analog magnitude* module (to the extent that they require number facts that are not linguistically stored) and the *auditory verbal* module, and problems requiring approximate calculations will be largely predicted by the *analog magnitude* module.



**Figure 14. Exact V. Approximate Theory: Problem Formatting Differences Not Specified.**

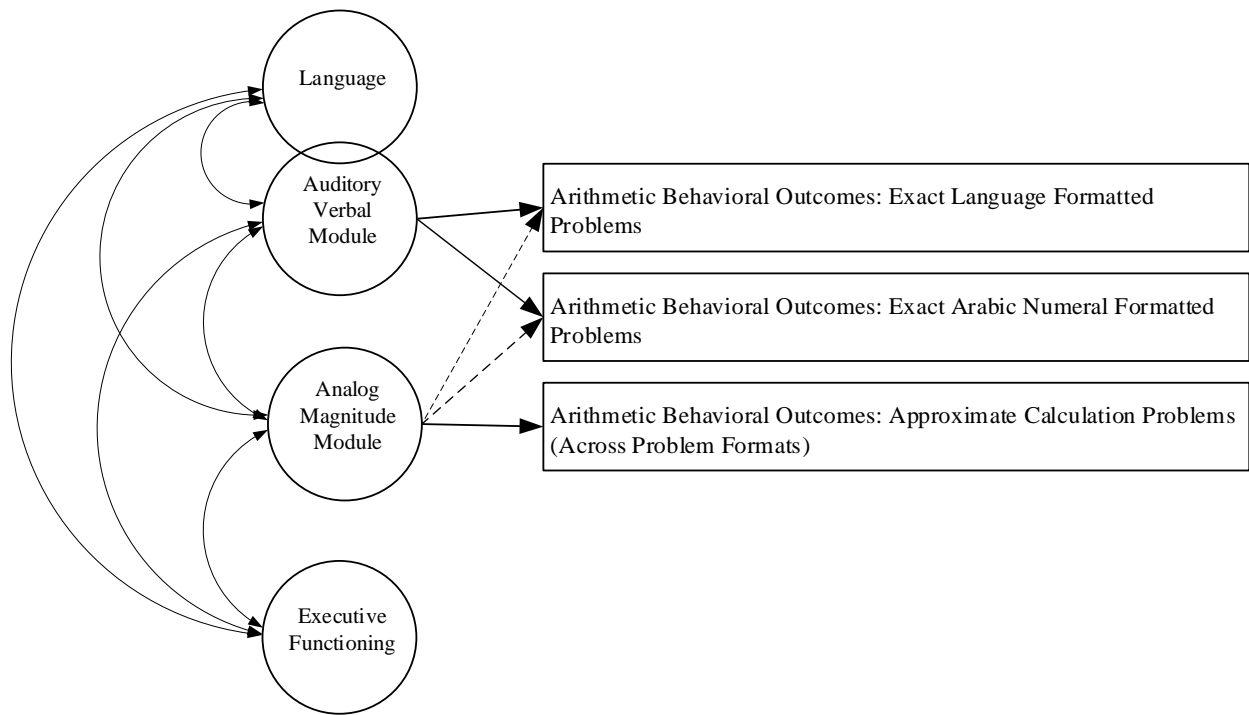
The contributions of domains other than the *analog magnitude* module and the *auditory verbal* module are also unclear. Language networks which also contribute to purely linguistic tasks are implicated in exact calculations, but the extent to which the *language* domain represents a unique contribution beyond the *auditory verbal* module is unclear. They may overlap so much that they do not appear to be separate domains, which would indicate that the *auditory verbal* module is in fact the *language* domain and does not have unique predictive

power in numeric processing. From a factor analytic framework, this can be represented with a latent *language* factor, which is allowed to correlate with the *auditory verbal module*. An extremely high correlation would indicate that they are not distinct factors. Without a compelling reason to restrict associations with other aspects of the model, this *language* domain can also be allowed to correlate with the analog magnitude module.

Similarly, the role of executive control in coordinating processing and facilitating spatial attention is unclear. Without more information on the contributions of "spatial attention networks", this domain can be tentatively represented with a latent factor for *executive control* which is allowed to associate with the *analog magnitude* and *auditory verbal* modules as well as with the *language* domain.

Like Triple Code Theory, Exact Versus Approximate Calculations Theory does not make specifications for developmental effects. This theory pertains to skilled adult arithmetic cognition, and the extent to which the model may apply to young calculators and may change with development and experience is unclear and an area in need to additional research.

With these caveats in mind, the Exact versus Approximate Calculations hypothesis of Triple Code Theory may be best represented as a four factor model in which both the *analogical magnitude representation* domain and an *executive* domain coordinating attention contribute to all numeric processing tasks, but the *language* domain and possibly a unique *auditory verbal* module contribute only to tasks requiring exact calculations.



*Analog magnitude module may predict exact problem performance if the math facts in the problem are unknown.*

**Figure 15. Exact V. Approximate Theory: Other Cognitive Domains' Contributions Are Unclear.**

## 2.2 Hypotheses

We might expect four potential hypotheses which can be tested against each other as well as examined for language-based and/or executive functioning-based common method variance.

**H<sub>1</sub>:** Mathematics performance is best represented by the Abstract Code Theory of arithmetic cognition. One factor, *abstract semantic representation*, predicts performance across a variety of arithmetic problem formats, and this factor may or may not correlate with additional cognitive domains (i.e., language, executive functioning).

**H<sub>2</sub>:** Mathematics performance is best represented by the Encoding Complex Theory of arithmetic cognition. Processing of mathematical tasks occurs in a network of numerical

associations. Performance on arithmetic problems is also predicted by the cognitive domains of executive functioning and language ability.

**H<sub>3</sub>:** Mathematics performance is best represented by the Triple Code Theory of arithmetic cognition. Processing of mathematical tasks occurs in one of three domains, which are separate but related via transcoding. These domains may or may not correlate with additional cognitive abilities (i.e., language, executive functioning).

**H<sub>4</sub>:** Arithmetic performance is best represented by the Exact versus Approximate specification of the Triple Code Theory of arithmetic cognition. Processing of mathematical tasks occurs largely in the analog magnitude domain, but executive domains may coordinate attention on all tasks. For tasks requiring exact calculations, the auditory verbal module and/or the language domain may also contribute to processing. For tasks requiring approximate calculations, the analog magnitude domain may show larger contributions than it does on exact calculations (represented by factor loadings). This pattern of dimensionality is a function of problem demands and is expected across various problem formats.

### **3 CHAPTER 3: METHODS**

The participants of this study were drawn from the baseline data of a six year, prospective, longitudinal study designed to test the effectiveness of an experimental instructional program for mathematics problem solving and to examine the cognitive development and predictors of mathematics problem solving (see for example Fuchs et al., 2008 ).

#### **3.1 Participants**

Participants were enrolled in public schools in Southeastern metropolitan school districts. Upon entering third grade during the fall of each school year, those students who assented to participation and whose parents consented to participate in the study were included in assessment

(and instructional intervention for the purposes of the parent study; see for example (Fuchs et al., 2008). An initial 2,023 students across 120 classrooms had consent to participate in the parent study. A subset of  $N=1320$  children were randomly selected for full participation in the parent study. These participants received the full testing battery (including screening measures, the full mathematics battery, cognitive measures, and demographic reports from teachers); however, for the purpose of the current study, all students with some data on mathematics, language, and/or executive function measures were included as participants. Because participants in the parent study were a randomly sampled subset of students, some data were unavailable on the full battery for students not included in the parent study. Thus, for all measures included in the current study, certain percentages of data were unavailable because the parent study, by design, collected data on a smaller sample than were included in the current study. Because these data were unavailable by design, they were considered to be "planned missing" and "missing at random". Implications for modeling data that are missing at random are considered in the analysis section.

A final sample of 1959 children was selected for the current study from the baseline time points of the Grade 3 Mathematics Problem Solving Study (MPS3). Of this total sample, approximately 67% of participants ( $N=1312$ ) had available demographic data on measures considered in the present study (age, gender, race/ethnicity, eligibility for free or reduced price lunch, and special education status/category), and according to the parent study's design, demographic data were not collected on the remaining students. Table 1 presents information for patterns of unavailable data on demographic measures of interest to the current study.

Based on the students for whom demographic data were collected (See Table 1), the current sample had a mean age of 103.24 months ( $SD=5.41$ , range = 89 – 142), was

approximately 50% female (N=660 females, N=652 males), and was ethnically and racially diverse (43% African American, 40% White, 10% Hispanic, 1% Kurdish, 4% other not specified, and 1% missing). Approximately 56% of the children in the sample qualified for free or reduced lunch. Teachers reported that approximately 5% of the children in the parent study sample were receiving special education services. Of those 67 children whose teachers reported receiving special education services, most were receiving services for learning disabilities (N=22), speech/hearing/language (N=21), ADHD (N=7), or giftedness (N=4).

**Table 1. Patterns of unavailable data on demographic variables of interest**

General Pattern	Specific Pattern	N	% of Data
Full Data Coverage 44.77%	All data present on demographic measures	877	44.77%
1 Measure Unavailable 41.14%	Special education category unavailable	325	16.59%
	Free and reduced price lunch unavailable	97	4.95%
2 Measures Unavailable 13.48%	Both free and reduced price lunch AND special education category unavailable	13	.66%
No Available Data .31%	No available data on demographic measures (students not included in parent study)	647	33.03%

## 3.2 Procedures

During September and October of each year of the study, (1) a demographic questionnaire was completed by teachers, (2) students' mathematical skills were assessed in three sessions lasting 30-60 minutes each, and (3) students' cognitive abilities were assessed in two sessions lasting 45 minutes each. Total testing span from first assessment to last was approximately one month.

The mathematics battery was administered to students using a whole classroom assessment methodology. Students received individual stimulus papers and pencils. Trained assessment professionals read questions aloud while students followed along on their own paper copies. Students were given time to respond to each question, and the next question was not administered until all students or all but two students had put their pencils down. Students were not permitted to communicate answers or disrupt the testing of the whole class.

The cognitive battery (which includes the measures of language and executive functioning) was administered using an individual assessment methodology. Trained assessment professionals administered items to students in one-on-one interactions in quiet testing locations within their schools.

## 3.3 Measures

### 3.3.1 *Mathematics achievement measures with language formatting.*

The current research study used a variety of measures of mathematics achievement, each designed to capture various types of formatting for arithmetic problems. Next, three measures of mathematics which used language-formatting will be reviewed.

### 3.3.1.1 *WJ III Applied Problems.*

This measure consisted of 60 orally presented word problems designed to represent every day, practical math problems (McGrew & Woodcock, 2001). Items required examinees to count, perform simple arithmetic operations, tell time, tell temperature, or problem-solve by eliminating extraneous information from the prompt. The test was not timed and was discontinued after examinees reached a ceiling of six, consecutive incorrect items. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Test developers report a one year test-retest reliability of .85 to .86 and a split half reliability of .88 to .95 for ages 2-18 years (McGrew & Woodcock, 2001). The WJ-III Broad Math Cluster, which includes the Applied Problems subtest, correlates well with other measures of mathematics achievement (Wechsler Individual Achievement Test, WIAT, at  $r=.70$  and the Kaufman Test of Educational Achievement, KTEA, at  $r=.66$ ; McGrew & Woodcock, 2001). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.1.2 *Single Digit Story Problems.*

This experimental measure consisted of 14 items (developed from Jordan & Hanich, 2000). Students were presented with the written word problems, which were read aloud by examiners. Each item could be solved in one step and involved combining, comparing, changing, and equalizing relationships with sums or minuends of 9 or less. Students were required to provide a correct response within 30 seconds of the oral prompt in order to receive credit for a correct answer; however, students were permitted to ask for re-readings of items as needed and without penalty to their timed responses. All students were administered all 14 items. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of

correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### *3.3.1.3 Vanderbilt Story Problems.*

This experimental measure consisted of 18 items read aloud to students while they followed along on their own written copies (Fuchs, Hamlett, & Powell, 2003). Students were not timed and were permitted to ask for re-readings of items as needed. Each item involved one to four steps for solution and could be solved by using step-up functions, adding multiple quantities of items with different prices, calculating money remaining after a purchase, finding half of a quantity, or summing quantities derived from pictographs in which quantities were also presented verbally. Nine of these items were more complex and required students to eliminate extraneous information from the problem, solve problems involving novel contexts using real-world information and their own problem-solving experiences, and apply information and solutions generated in previous complex problems on the assessment. Students could earn a total of 2 points per item, 1 point for correctly calculating intermediate steps in the problem, and 1 point for correctly labeling the final answer. Raw scores were total number of points achieved per item. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### ***3.3.2 Mathematics achievement measures with Arabic numeral formatting.***

#### *3.3.2.1 Basic Facts Addition.*

This experimentally designed measure consisted of 25 addition fact items, which were Arabic numeral formatted and delivered in written form to students (Fuchs et al., 2003). Each item involved addends of 9 or less and sums of 12 or less. Students were provided with the stimulus paper and a pencil and were permitted one minute to complete as many items as

possible. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.2.2 *Basic Facts Subtraction.*

This experimentally designed measure consisted of 25 subtraction fact items, which were Arabic numeral formatted and delivered in written form to students (Fuchs et al., 2003). Each item involved minuends of 18 or less and answers of 12 or less. Students were provided with the stimulus paper and a pencil and were permitted one minute to complete as many items as possible. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.2.3 *WRAT Written Arithmetic.*

The WRAT-3 Written Arithmetic subtest (Blue form; Wilkinson, 1993) consisted of 40, Arabic numeral formatted computation problems. Items were presented in written format, and students were provided a pencil and asked to produced written responses to as many items as possible within 15 minutes. Items contained a variety of arithmetic content ranging from basic facts (basic addition, subtraction, multiplication, and division), to performing arithmetic operations involving multiple operands, to performing arithmetic operations with percentages and fractions, to reducing and evaluating algebraic expressions. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Test developers reported WRAT Arithmetic coefficient alpha reliabilities ranging from .80 to .89 for individuals ages 6 to 16 years and test-retest reliability of .94 for individuals ages 6 to 16 years

(Wilkinson, 1993). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### *3.3.2.4 2nd grade Computational Fluency.*

This experimental measure consisted of 25, Arabic numeral formatted items and was designed for second grade addition, subtraction, number combinations, and procedural computation problems (Fuchs, Hamlett, & Fuchs, 1990). Examinees were given 3 minutes to complete as many problems as possible. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### *3.3.2.5 Double Digit Addition.*

This experimentally designed measure consisted of 20, Arabic numeral formatted, 2-digit by 2-digit addition items with and without regrouping (Fuchs et al., 2003). Students were provided with a written protocol and pencil, and given 5 minutes to complete as many problems as possible. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### *3.3.2.6 Double Digit Subtraction.*

This experimentally designed measure consisted of 20, Arabic numeral formatted, 2-digit by 2-digit subtraction items with and without regrouping (Fuchs et al., 2003). Students were provided with a written protocol and pencil, and given 5 minutes to complete as many problems as possible. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### ***3.3.3 Mathematics achievement measures involving estimation or analog magnitude.***

#### ***3.3.3.1 Double Digit Estimation Addition.***

This experimentally designed measure consisted of 20, Arabic numeral formatted, 2-digit by 2-digit addition items in which students were instructed to estimate answers to the nearest ten (Fuchs et al., 2003). Examiners completed a sample problem in order to demonstrate estimation and to remind students that they would not be computing exact answers to problems. Students were provided with a written protocol and pencil, and given 5 minutes to complete as many problems as possible. Correct items were scored "1," and incorrect items were scored "0." Because this was an estimation task, exact calculated answers were scored as incorrect. Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### ***3.3.3.2 Double Digit Estimation Subtraction.***

This experimentally designed measure consisted of 20, Arabic numeral formatted, 2-digit by 2-digit subtraction items in which students were instructed to estimate answers to the nearest ten (Fuchs et al., 2003). Examiners completed a sample problem in order to demonstrate estimation and to remind students that they would not be computing exact answers to problems. Students were provided with a written protocol and pencil, and given 5 minutes to complete as many problems as possible. Correct items were scored "1," and incorrect items were scored "0." Because this was an estimation task, exact calculated answers were scored as incorrect. Total raw score was the number of correct items. Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### **3.3.4 *Language measures.***

#### **3.3.4.1 *WASI Vocabulary.***

The Vocabulary subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) consisted of 42 items, assessing expressive vocabulary. The initial four items required students to view a picture display and provide a verbal label for the object in each picture. Remaining items required students to provide definitions for vocabulary prompts given by examiners. Responses to all items were scored "0" if incorrect, "1" if partially correct, or "2" if the targeted response was present. The test was not timed and was discontinued after examinees reached a ceiling of five, consecutive incorrect items. Total raw score was the number of correct items. Test developers report a split half reliability of .86 and .88 for ages 8 to 9 years and test-retest reliability of .85 for ages 6-16 years (Wechsler, 1999). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### **3.3.4.2 *WDRB Listening Comprehension.***

The Listening Comprehension subtest of the Woodcock Diagnostic Reading Battery (WDRB; Woodcock, 1997) consisted of 38 sentences or passages, read aloud to examinees who were then prompted to supply the missing word at the end of each prompt. Initial items required students to complete simple verbal analogies and word associations, and as the test continued, items became more complex and required students to discern implications of the passages they had just heard. The test was not timed and was discontinued after examinees reached a ceiling of six, consecutive incorrect items. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Test developers report a reliability of .80 for ages 5-18 years (Woodcock, 1997). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.4.3 *TOLD Grammatical Closure.*

The Grammatical Closure subtest of the Test of Language Development (TOLD-Revised edition; Newcomer & Hammill, 1988) consisted of 30 sentences, assessing ability to recognize, understand, and express English morphology. Students are prompted with a sentence that is missing a word and respond verbally to supply the missing word and complete the sentence. The test was not timed and was discontinued after examinees reached a ceiling of six, consecutive incorrect items. Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Test developers report a reliability of .88 for age 8 years (Newcomer & Hammill, 1988). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.5 *Executive functioning measures.*

#### 3.3.5.1 *SWAN.*

The SWAN (Swanson & Beebe-Frankenberger, 2004) is a teacher survey with 18 items measuring attention, inhibition, and self-regulation. This instrument was originally designed to measure the inattentive behavior, distractibility, impulsivity, and hyperactivity characteristic of Attention-Deficit/Hyperactivity Disorder (ADHD) while also capturing the normal distribution of non-clinical behavior. The first nine items of the SWAN prompted teachers to rate students for various types of inattentive behavior and distractibility, and the next nine items prompted teachers to rate students for various types of impulsive and hyperactive behaviors. Each item asked teachers to rate a student's behaviors on a seven point Likert-type scale (ranging from 7 "far above average," 6 "above average," 5 "slightly above average," 4 "average," 3 "slightly below average," 2 "below average," 1 "far below average." Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.5.2 *WMTB Listening Recall.*

The Listening Recall subtest of the Working Memory Test Battery for Children (WMTB-C; Pickering & Gathercole, 2001) consisted of sequences of sentences, assessing verbal working memory. Examiners read aloud a series of short sentences to students. After listening to each sentence, the student evaluates the sentence as true or false. Finally, after evaluating all of the sentences in a trial, the student is asked to recall, in order, the last word of each sentence in the trial. The test was not timed and was discontinued after examinees reached a ceiling of three or more errors in any block of items. Each sequence of final words recalled correctly and in the correct order was scored "1". A sequence in which either final words were not recalled correctly or were not recalled in the correct order was scored "0". Total raw score was the number of correct sequences recalled. Test developers report a test-retest reliability of .93 for ages 5 to 15 years (Pickering & Gathercole, 2001). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### 3.3.5.3 *WJ III Numbers Reversed.*

The Numbers Reversed subtest of the WJ-III (Test of Cognitive Abilities; Woodcock, McGrew, & Mather, 2001) consisted of 30 items, assessing working memory. On each item, students listened to orally presented, random spans of digits, and upon completion of the span, students were prompted to orally list the digits they had just heard in reversed order. As students progressed through the test, digit spans increased, ranging from two to eight digits. The test was not timed and was discontinued after examinees reached a ceiling of three errors in a block of items (note that blocks vary in the number of items they contain; each block ends with a possible stopping point that is pre-determined by the test developer). Correct items were scored "1," and incorrect items were scored "0." Total raw score was the number of correct items. Test

developers report a split half reliability of .84 to .93 for ages 2 to 18 years (McGrew & Woodcock, 2001). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### 3.3.5.4 WASI Matrix Reasoning.

The Matrix Reasoning subtest of the WASI is designed to measure nonverbal problem-solving or induction (Wechsler, 1999). This assessment requires examinees to view visual displays of matrices from which a section is missing and to use pattern completion, classification, analogy, and serial reasoning to induct the rule in the matrix and predict the next item in the sequence. Examinees complete the matrix using one of five possible response choices from a multiple choice array beneath the matrix prompt. Responses could be identified verbally or with pointing. Testing is discontinued after examinees make four errors within a set of five consecutive items. Correct responses are recorded as "1" and incorrect responses are recorded as "0". Test developers report a split half reliability of .94 and .93 for ages 8 and 9 years and test-retest reliability of .77 for ages 6-16 years (Wechsler, 1999). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

#### 3.3.5.5 WJ III Concept Formation.

The Concept Formation subtest of the WJ-III (Test of Cognitive Abilities; Woodcock et al., 2001) consisted of 40 items, assessing fluid intelligence and induction. On each item, students were shown illustrations which demonstrated instances and non-instances of a concept and were asked to identify the rules for concepts by inducting or inferring the rules. The test was not timed and was discontinued after examinees reached one of four cut-off points that were predetermined by the test developer (e.g., 2 or fewer correct among items 1 - 5, or 5 or fewer correct among items 1 - 11). Correctly identified rules were scored "1," and incorrectly identified rules

were scored "0." Total raw score was the number of correct items. Test developers report a split half reliability of .75 to .96 for ages 2 to 18 years (McGrew & Woodcock, 2001). Model-based reliability, in the form of  $R^2$ , will be considered and reported for this study.

### **3.4 Design**

The parent study was designed to sample four cohorts of 3<sup>rd</sup> grade students, following each cohort for three academic years spanning from the fall of 3<sup>rd</sup> grade until the spring of 5<sup>th</sup> grade. The current study, however, is focused on the baseline time points of testing for each of these four cohorts of students. Table 2 displays the cohort sampling information.

Only a randomly selected subset of children in the parent study received the full measurement battery (including mathematics, language, executive function, and demographic measures). Those children who were not selected to receive the full measurement battery (approximately 33% of the total sample) have consistently unavailable data (planned missing) on several outcomes of interest.

Furthermore, the full mathematics assessment battery involved 11 measures total, and therefore, the mathematics assessments also were delivered using a planned missing design such that not all measures were administered to the random subset of children selected to receive the full battery every year of the study (for more information on planned missing designs, see for example Graham, Hofer, & MacKinnon, 1996). Because of the planned missingness inherent in this design, cohorts which have unavailable data on certain measures are assumed to have data that are missing completely at random, or MCAR. Table 3 lists patterns of unavailable data in the outcome measures of the testing battery received by each cohort. Tables 4, 5, and 6 list descriptive information for mathematics, language, and executive function measures respectively. Table 7 presents a full correlation matrix for all measures in the study.

**Table 2. Cohort Measurement Information**

Measures	Cohort 1 Received	Cohort 2 Received	Cohort 3 Received	Cohort 4 Received
Mathematics Measures:				
WJ-III Applied Problems	X	X	X	X
Single Digit Story Problems	X	X	X	X
Vanderbilt Complex Story Problems				X
Basic Facts Addition	X	X	X	X
Basic Facts Subtraction	X	X	X	X
Test of Computational Fluency	X	X	X	X
WRAT Written Arithmetic		X	X	X
Double Digit Addition	X			
Double Digit Subtraction	X			
Double Digit Addition Estimation	X			
Double Digit Subtraction Estimation	X			
Language Measures:				
WASI Vocabulary	X	X	X	X
WDRB Listening Comprehension	X	X	X	X
TOLD Grammatic Closure	X	X	X	X
Executive Function Measures:				
SWAN teacher survey	X	X	X	X
WMTB Listening Recall	X	X	X	X
WJ-III Numbers Reversed	X	X	X	X
WASI Matrix Reasoning	X	X	X	X
WJ-III Concept Formation	X	X	X	X
Cohort Sampling Information	N=491 stud. N=30 class. N=7 school.	N=485 stud. N=30 class. N=8 school.	N=452 stud. N=29 class. N=8 school.	N=531 stud. N=31 class. N=9 school.
Total Sample for the Current Study	N=1959 students N=120 classrooms (classrooms do not overlap) N=16 schools (schools do overlap across cohorts)			

**Table 3. Most prevalent patterns of unavailable (planned missing) data on outcome measures of interest**

General Pattern	Sample Subset of Parent Study	Specific Pattern of Unavailable Data (Assumed MCAR)	N	% Of Data
Cohort 1	Selected for Full Battery	Planned Missing VSP and WRAT only	312	15.93%
	Not Selected for Full Battery	Planned Missing All Except Screen: SDS, BFA, BFS, CBM, DD	120	6.13%
	Various	Other patterns of coverage	59	3.01%
Cohort 2	Selected for Full Battery	Planned Missing VSP and DD	309	15.77%
	Not Selected for Full Battery	Planned Missing All Except Screen: SDS, BFA, BFS, CBM, WRAT	146	7.45%
	Various	Other patterns of coverage	30	1.53%
Cohort 3	Selected for Full Battery	Planned Missing VSP and DD	302	15.42%
	Not Selected for Full Battery	Planned Missing All Except Screen: SDS, BFA, BFS, CBM, WRAT	130	6.64%
	Various	Other patterns of coverage	20	1.02%
Cohort 4	Selected for Full Battery	Planned Missing DD	300	15.31%
	Not Selected for Full Battery	Planned Missing All Except Screen: SDS, BFA, BFS, CBM, WRAT, VSP	200	10.21%
	Various	Other patterns of coverage	31	1.58%

Note: Taken together, Tables 2-7 help to explain the patterns of unavailable data in outcome measures. Table 2 gives information about the measures administered to each cohort. Table 3 gives information about the patterns of unavailable data amongst outcome measures. Tables 4-7 give information about the sample size, correlations, means, standard deviations, and ranges for all outcome measures of interest in the current study. Patterns of planned missing data represent (1) children who were selected for participation in the parent study and had complete data on all planned study measures during their years of participation, approximately 62% of the current sample, (2) children who were not selected for participation in the parent study and had complete data on all planned screening measures during their years of participation, approximately 30% of the current sample, and (3) children who were missing data as a result of unplanned issues during data collection (e.g., 6 children in cohort 1 for whom a teacher did not complete the SWAN survey), an additional approximate 7% of participants.

**Table 4. Math Measures Means and Correlations**

	1	2	3	4	5	6	7	8	9	10	11
1. WJ App. Prb.	1.00										
2. Story Prb.	.58	1.00									
3. VU Story Prb.	.53	.50	1.00								
4. Basic Add.	.40	.36	.37	1.00							
5. Basic Sub.	.42	.39	.37	.58	1.00						
6. WRAT Arth.	.56	.51	.48	.48	.49	1.00					
7. Comp Fluency	.49	.45	.43	.68	.65	.57	1.00				
8. DD Add	.34	.38	.30	.41	.26	.31	.43	1.00			
9. DD Sub	.40	.42	.34	.33	.39	.34	.42	.47	1.00		
10. DD Add. Est.	.49	.49	.39	.41	.43	.39	.46	.38	.45	1.00	
11. DD Sub. Est.	.44	.44	.35	.36	.45	.35	.41	.36	.50	.73	1.00
N	1303	1949	530	1950	1950	1464	1940	467	467	468	466
Mean	29.15	9.89	8.22	11.80	6.85	23.63	11.97	17.38	11.64	9.16	7.01
(SD)	(4.32)	(3.48)	(5.95)	(4.89)	(4.80)	(2.55)	(5.75)	(3.99)	(5.51)	(6.93)	(5.83)
Range:	2-48	0-14	0-34	0-25	0-25	15-38	0-25	0-20	0-20	0-20	0-20
Min-Max											

\*Note: All correlations were significant at the  $p < .001$  level

**Table 5. Language Measures Means and Correlations**

	1	2	3
1. WASI Vocab.	1.00		
2. WDRB List. Comp.	.53	1.00	
3. TOLD Gram. Clos.	.52	.52	1.00
N	1314	1302	1303
Mean	27.35	21.12	18.78
(SD)	(6.45)	(4.29)	(6.60)
Range: Min-Max	5-51	0-33	0-30

\*Note: All correlations were significant at the  $p < .001$  level

**Table 6. Executive Functioning Measures Means and Correlations**

	1	2	3	4	5
1. SWAN Teach Suv.	1.00				
2. WMTB List. Rec.	.25	1.00			
3. WJ Num Rev.	.28	.31	1.00		
4. WASI Mat. Rea.	.29	.23	.32	1.00	
5. WJ Con. Form.	.37	.37	.30	.40	1.00
N	1276	1302	1302	1314	1302
Mean	75.48	9.97	9.37	15.51	15.64
(SD)	(23.52)	(3.58)	(2.85)	(6.45)	(7.07)
Range:	18-126	0-63	1-26	0-30	1-39
Min-Max					

\*Note: All correlations were significant at the  $p < .001$  level

**Table 7. Full Correlation Matrix for All Measures**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. App. Prb.	1.00																	
2. Story Prb.	.58	1.00																
3. VU Story Prb.	.53	.50	1.00															
4. Basic Add.	.40	.36	.37	1.00														
5. Basic Sub.	.42	.39	.37	.58	1.00													
6. WRAT Arth.	.56	.51	.48	.48	.49	1.00												
7. Comp Fluency	.49	.45	.43	.68	.65	.57	1.00											
8. DD Add.	.34	.38	.30	.41	.26	.31	.43	1.00										
9. DD Sub.	.40	.42	.34	.33	.39	.34	.42	.47	1.00									
10. DD Add. Est.	.49	.49	.39	.41	.43	.39	.46	.38	.45	1.00								
11. DD Sub. Est.	.44	.44	.35	.36	.45	.35	.41	.36	.50	.73	1.00							
12. Vocab.	.45	.45	.38	.18	.22	.35	.29	.19	.32	.36	.30	1.00						
13. List. Comp.	.40	.44	.38	.14	.17	.28	.20	.22	.24	.37	.25	.53	1.00					
14. Gram. Clos.	.41	.44	.33	.16	.18	.26	.22	.22	.25	.26	.18	.52	.52	1.00				
15. SWAN	.38	.44	.45	.27	.29	.43	.37	.37	.45	.40	.41	.38	.31	.32	1.00			
16. List. Rec.	.31	.32	.25	.17	.18	.22	.20	.25	.27	.30	.30	.33	.32	.39	.25	1.00		
17. Num Rev.	.35	.35	.37	.22	.24	.30	.23	.25	.31	.32	.31	.26	.21	.27	.28	.31	1.00	
18. Mat. Rea.	.45	.42	.35	.21	.24	.37	.27	.28	.35	.42	.37	.31	.28	.30	.29	.23	.32	1.00
19. Con. Form.	.46	.49	.41	.25	.26	.37	.32	.28	.39	.43	.38	.44	.40	.41	.37	.37	.30	.40

## 4 CHAPTER 4: RESULTS

### 4.1 Proposed Analyses Overview

Planned analyses were executed in two phases of testing. Phase one began by examining measurement models for mathematics measures using confirmatory factor analysis with maximum likelihood estimation in MPlus 7 (Muthén & Muthén, 2012). Next measurement models for language and executive functioning were examined using confirmatory factor analysis with maximum likelihood estimation in MPlus 7 (Muthén & Muthén, 2012). Phase two examined full measurement models, incorporating all constructs of interest to the current study (mathematics, language, and executive functioning as outlined in the hypotheses). Missing data were estimated using full information maximum likelihood estimation (see for example Enders & Bandalos, 2001) in MPlus 7 (Muthén & Muthén, 2012).

### 4.2 Phase 1: Measurement Models for Arithmetic, Language, and Executive Functioning

#### 4.2.1 *Abstract code model.*

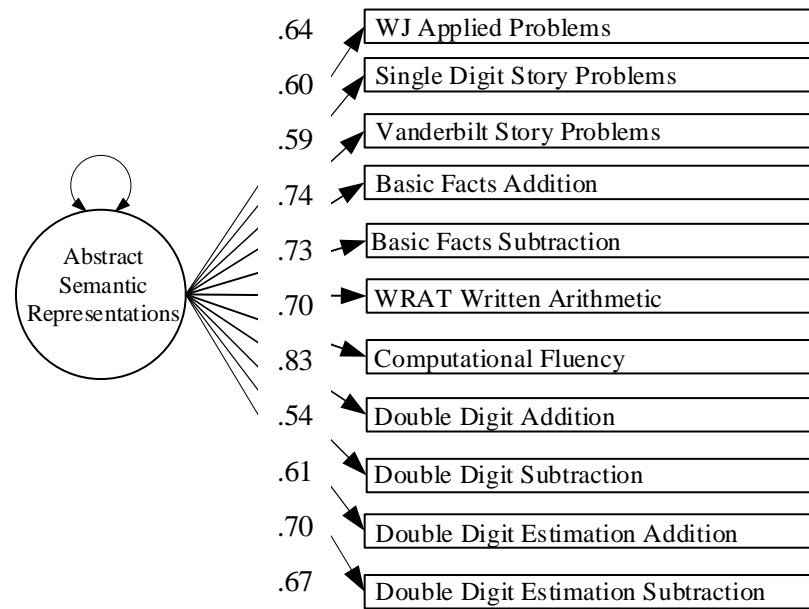
Under Abstract Code Theory, *abstract semantic representations* are the common form of mental representation upon which all modules of numeric processing operate. Arithmetic behavioral outcomes are predicted by one, latent form of mental representation across a variety of stimulus formats and problem demands. Thus, the arithmetic portion of the Abstract Code Theory measurement model is represented as a factor model in which various stimulus formats have no distinct common method variance and instead are predicted by one, latent dimension.

The *abstract semantic representations* measurement model tests the extent to which the 11 mathematics indicators measure a unitary, underlying, *abstract semantic representation* in predicting mathematics outcomes. Global fit statistics indicated that this factor model was not a good fit for the data,  $(\chi^2(36) = 705.68, p < .001, RMSEA = .10, CFI = .88)$ ; for a discussion of fit,

see Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004). Local fit statistics indicated that although most factor loadings were adequate (both significant and salient), several indicator residuals were undesirably high. Completely standardized factor loadings ranged from .54 to .83, indicator residual variances ranged from .31 to .71, and model  $R^2$  ranged from .29 to .69. Taken together, these results indicate that although the 11 mathematics measures share some underlying commonality, they are also predicted by complexities not modeled in the Abstract Code Theory measurement model, which predicts that despite format differences, abstract semantic representations should underlie arithmetic cognition. Table 8 presents standardized and unstandardized results for the Abstract Code Theory Arithmetic measurement model. Figure 16 displays a model schematic.

**Table 8. Abstract Code Theory Arithmetic Measurement Model CFA Results**

Indicator	Intercept		Factor Loadings		Residual Variance	$R^2$
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
App. Prb.	6.84 (.13)	29.08 (.11)	.64 (.02)	2.73 (.11)	.59	.41
Story Prb.	2.85 (.05)	9.90 (.08)	.60 (.02)	2.07 (.08)	.65	.35
VU Story Prb.	1.40 (.06)	8.41 (.23)	.59 (.03)	3.55 (.25)	.65	.35
Basic Add.	2.41 (.05)	11.79 (.11)	.74 (.01)	3.60 (.10)	.46	.54
Basic Sub.	1.43 (.03)	6.85 (.11)	.73 (.01)	3.51 (.10)	.47	.54
WRAT Arth.	9.31 (.17)	23.73 (.06)	.70 (.02)	1.80 (.06)	.50	.50
Comp Fluency	2.08 (.04)	11.97 (.13)	.83 (.01)	4.76 (.11)	.31	.69
DD Add.	4.23 (.15)	17.00 (.17)	.54 (.04)	2.16 (.18)	.71	.29
DD Sub.	1.99 (.08)	11.04 (.23)	.61 (.03)	3.37 (.24)	.63	.37
DD Add. Est.	1.18 (.06)	8.31 (.27)	.70 (.03)	4.93 (.29)	.51	.49
DD Sub. Est.	1.07 (.06)	6.32 (.24)	.67 (.03)	3.94 (.25)	.55	.45



$$\chi^2(36) = 705.68, p < .001, RMSEA = .097, CFI = .880$$

**Figure 16. Abstract Code Theory: Arithmetic Measurement Model.**

#### 4.2.2 Encoding complex model.

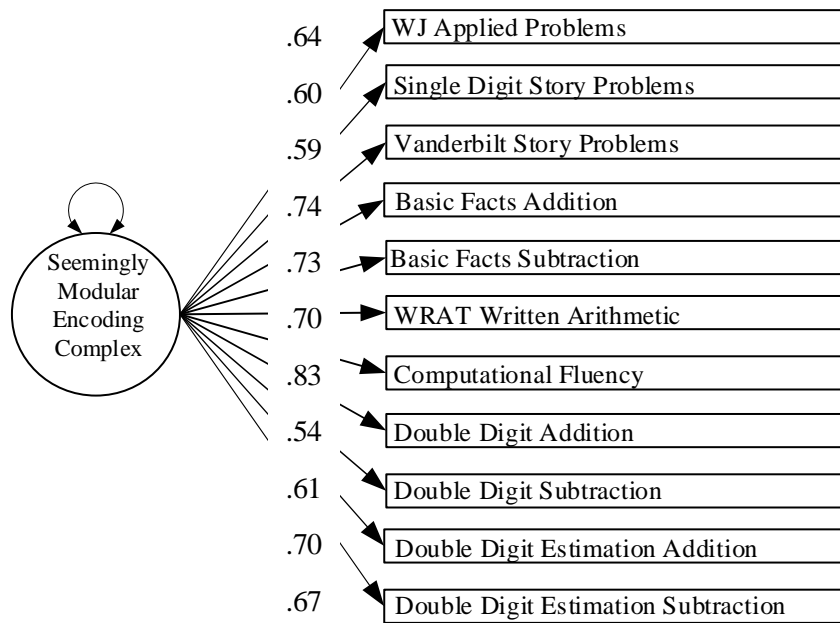
Under Encoding Complex Theory, the *quantitative* domain is not specialized for numeric processing and is not modular (meaning that it is not a self-contained, separate cognitive domain or dimension of ability), although it may appear to be modular with practice. The gradients of the seemingly-modular *encoding complex* as a function of "practice" are unclear; however, for individuals who have at least some amount of practice with arithmetic problem-solving, the arithmetic measurement portion of an Encoding Complex Theory model can be represented with a model in which one latent variable (the "practiced" and seemingly-modular encoding complex) predicts arithmetic behavioral outcomes across various arithmetic problems. To the extent that an encoding complex becomes seemingly-modular with practice, one would expect to see

commonality and overlap among arithmetic outcomes, indicating that they are predicted by the same cognitive trait.

The *seemingly-modular encoding complex* model tests the extent to which 11 arithmetic indicators measure a unitary, underlying, encoding complex factor, which appears to be modular with practice. It should be noted that this factor is being called "*seemingly-modular encoding complex*" here, but in actuality is the same measurement model as the *abstract semantic representations* measurement model. Thus, the *seemingly-modular encoding complex* model evidenced the same model fit problems as the *abstract semantic representations model*. Again, global fit statistics indicated that this factor model was not a good fit for the data, ( $\chi^2(36) = 705.68, p < .001, RMSEA = .10, CFI = .88$ ), and local fit statistics indicated that although most factor loadings were adequate, indicator residuals were undesirably high. Completely standardized factor loadings ranged from .54 to .83, indicator residual variances ranged from .31 to .71, and model  $R^2$  ranged from .29 to .69. Although the 11 mathematics measures share some underlying commonality, they are also predicted by complexities not modeled in the Encoding Complex Theory measurement model, which would seem to indicate that their overlap may not be best explained by a *seemingly-modular encoding complex*. Table 9 presents standardized and unstandardized results for the Encoding Complex Theory Arithmetic measurement model. Figure 17 displays a model schematic.

**Table 9. Encoding Complex Theory Arithmetic Measurement Model CFA Results**

Indicator	Intercept		Factor Loadings		Residual Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
App. Prb.	6.84 (.13)	29.08 (.11)	.64 (.02)	2.73 (.11)	.59	.41
Story Prb.	2.85 (.05)	9.90 (.08)	.60 (.02)	2.07 (.08)	.65	.35
VU Story Prb.	1.40 (.06)	8.41 (.23)	.59 (.03)	3.55 (.25)	.65	.35
Basic Add.	2.41 (.05)	11.79 (.11)	.74 (.01)	3.60 (.10)	.46	.54
Basic Sub.	1.43 (.03)	6.85 (.11)	.73 (.01)	3.51 (.10)	.47	.54
WRAT Arth.	9.31 (.17)	23.73 (.06)	.70 (.02)	1.80 (.06)	.50	.50
Comp Fluency	2.08 (.04)	11.97 (.13)	.83 (.01)	4.76 (.11)	.31	.69
DD Add.	4.23 (.15)	17.00 (.17)	.54 (.04)	2.16 (.18)	.71	.29
DD Sub.	1.99 (.08)	11.04 (.23)	.61 (.03)	3.37 (.24)	.63	.37
DD Add. Est.	1.18 (.06)	8.31 (.27)	.70 (.03)	4.93 (.29)	.51	.49
DD Sub. Est.	1.07 (.06)	6.32 (.24)	.67 (.03)	3.94 (.25)	.55	.45



$$\chi^2(36) = 705.68, p < .001, RMSEA = .097, CFI = .880$$

**Figure 17. Encoding Complex Theory: Arithmetic Measurement Model.**

### 4.2.3 *Triple code model.*

Under Triple Code Theory, arithmetic behavioral outcomes are predicted by three modules of a latent *quantitative* domain. The *visual Arabic* module processes digital input and output and multi-digit operations. The *auditory verbal* module processes simple arithmetic facts, written and spoken input and output, and language-based memory of numbers. The *analog magnitude representation* module processes semantic numeric content, comparison, estimation, approximate calculation, and subitizing tasks. The process of transcoding allows each of the three domains of Triple Code Theory are allowed to communicate with each other directly and without the need for common abstract codes. From a factor analytic framework, Triple Code Theory can be represented with a three factor model of arithmetic cognition in which (1) a visual Arabic factor is largely responsible for Arabic numeral formatted problems, (2) an auditory verbal factor is largely responsible for language-formatted problems, and (3) an analog magnitude factor is largely responsible for approximate calculations across formats. The communication between these factors, transcoding, can be represented with factor correlations.

The Triple Code Theory measurement model tests the extent to which various arithmetic outcomes can be represented by three latent factors which are separate but hypothesized to communicate and mutually inform arithmetic cognition. A *visual Arabic* factor is hypothesized to be indicated by six measures that are formatted with Arabic numerals (Basic Facts Addition, Basic Facts Subtraction, WRAT Written Arithmetic, Computational Fluency, Double Digit Addition, and Double Digit Subtraction). An *auditory verbal* factor is hypothesized to be indicated by three measures that have language-based formats (WJ Applied Problems, Single Digit Story Problems, and Vanderbilt Story Problems). An *analog magnitude* factor is hypothesized to be indicated by two measures that involve estimation (Double Digit Estimation

Addition and Double Digit Estimation Subtraction). These three factors were hypothesized to correlate, and thus, correlations between them were freely estimated.

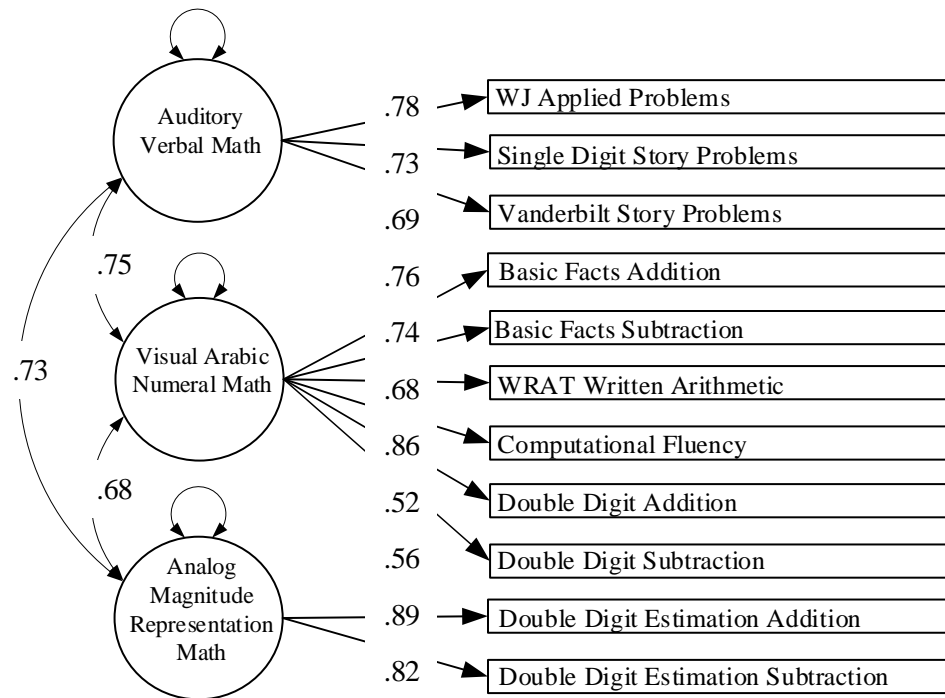
Global fit statistics indicated that this factor model was an approximate good fit for the data, ( $\chi^2 (33) = 302.59, p < .001, RMSEA = .07, CFI = .95$ ). Completely standardized factor loadings ranged from .52 to .89; indicator residual variances ranged from .20 to .73; and model  $R^2$  ranged from .27 to .80. Factor correlations ranged from  $r = .68$  to  $r = .75$ . These results support Triple Code Theory's specification that three, separate but mutually informed, format-specific modules predict arithmetic cognition outcomes. Table 10 presents standardized and unstandardized results for the Triple Code Theory Arithmetic measurement model. Table 11 presents the latent factor correlations for this model, and Figure 18 displays a model schematic.

**Table 10. Triple Code Theory Arithmetic Measurement Model CFA Results**

Latent Factor Indicators	Intercept	Factor Loadings			Residual Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Auditory Verbal						
App. Prb.	6.78 (.13)	29.06 (.11)	.78 (.02)	3.34 (.11)	.39	.61
Story Prb.	2.85 (.05)	9.90 (.08)	.73 (.02)	2.55 (.08)	.46	.54
VU Story Prb.	1.37 (.06)	8.23 (.22)	.69 (.03)	4.15 (.25)	.52	.48
Visual Arabic						
Basic Add.	2.41 (.05)	11.79 (.11)	.76 (.01)	3.71 (.10)	.42	.58
Basic Sub.	1.43 (.03)	6.85 (.11)	.74 (.01)	3.57 (.10)	.45	.55
WRAT Arth.	9.29 (.17)	23.74 (.06)	.68 (.02)	1.74 (.06)	.54	.47
Comp Fluency	2.08 (.04)	11.97 (.13)	.86 (.01)	4.96 (.11)	.26	.75
DD Add.	4.24 (.15)	16.98 (.17)	.52 (.04)	2.07 (.18)	.73	.27
DD Sub.	2.00 (.08)	11.04 (.23)	.56 (.04)	3.07 (.25)	.69	.31
Analog Magnitude						
DD Add. Est.	1.22 (.06)	8.53 (.28)	.89 (.02)	6.25 (.28)	.20	.80
DD Sub. Est.	1.11 (.06)	6.52 (.24)	.82 (.02)	4.81 (.24)	.33	.67

**Table 11. Triple Code Theory Arithmetic Latent Factor Correlations**

	1	2	3
1. Auditory Verbal Factor	1.00		
2. Visual Arabic Factor	.75	1.00	
3. Analog Magnitude Factor	.73	.68	1.00



$$\chi^2(33) = 302.59, p < .001, RMSEA = .065, CFI = .952$$

**Figure 18. Triple Code Theory: Arithmetic Measurement Model.**

#### 4.2.4 *Exact versus approximate model.*

Under Exact Versus Approximate Theory, arithmetic behavioral outcomes for different types of problem demands may call on an *analogical magnitude* module, responsible for representing semantic information about quantity, and an *auditory verbal* module, contributing verbally stored information about number facts. Exact calculation problems mainly require contribution from the *auditory verbal* module but may call on the *analog magnitude* module when they are not stored

as facts in the *auditory verbal* module (e.g., if the facts have not yet been learned or involve numbers and operations which are not commonly executed). Problems which require estimation or approximate calculations, on the other hand, mainly require contributions from the *analogical magnitude* module. Exact Versus Approximate Calculations Theory does not specify predictions for problem formatting. As an extension of Triple Code Theory, it is mainly concerned with specifying domains responsible for exact and approximate problem demands across various problem formats.

The core premise of Exact versus Approximate Calculations Theory can be represented with a factor model in which two latent factors (representing the *analog magnitude* module and the *auditory verbal* module) predict arithmetic behavioral outcomes for exact and approximate problem demands. At a minimum, these latent factors can be allowed to correlate and communicate with one another.

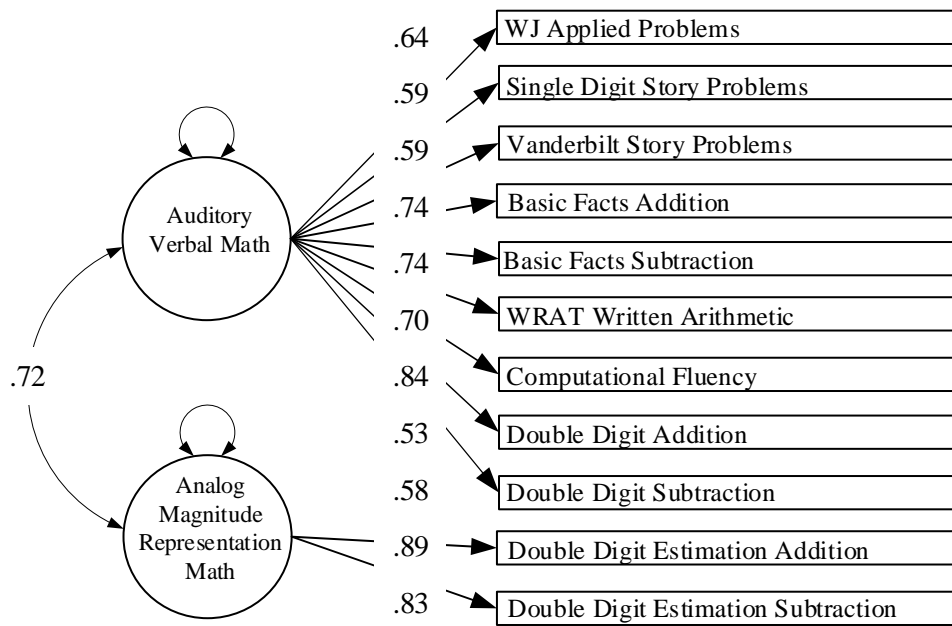
An *analog magnitude* modules is hypothesized to be indicated by two measures that involve estimation or approximate calculations (Double Digit Estimation Addition and Double Digit Estimation subtraction), and an *auditory verbal* module is hypothesized to be indicated by nine measures that involve exact calculations (WJ Applied Problems, Single Digit Story Problems, Vanderbilt Story Problems, Basic Facts Addition, Basic Facts Subtraction, WRAT Written Arithmetic, Computational Fluency, Double Digit Addition, and Double Digit Subtraction). These two factors were hypothesized to correlate, and thus, correlation between them was freely estimated.

Global fit statistics indicated that this factor model was not an approximate good fit for the data, ( $\chi^2(35) = 547.10, p < .001, RMSEA = .09, CFI = .91$ ). Completely standardized factor loadings ranged from .53 to .89; indicator residual variances ranged from .22 to .72; and model

$R^2$  ranged from .28 to .78. The factor correlation between the *analog magnitude* module and the *auditory verbal* module was large,  $r=.72$ . These results suggest that although separating estimation problem demands from exact problem demands provides an improvement in model fit (as compared to a unidimensional model of arithmetic cognition suggested for Abstract Code Theory and Encoding Complex Theory), important dimensions of the cognitive architecture are not being modeled here. Table 12 presents standardized and unstandardized results for the Exact Versus Approximate Arithmetic measurement model. Figure 19 displays a model schematic.

**Table 12. Exact Versus Approximate Calculations Arithmetic Measurement Model CFA Results**

Latent Factor					Residual	
Indicators	Intercept	Factor Loadings			Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Auditory Verbal						
App. Prb.	6.84 (.13)	29.08 (.11)	.64 (.02)	2.70 (.11)	.60	.40
Story Prb.	2.85 (.05)	9.90 (.08)	.59 (.02)	2.04 (.08)	.66	.34
VU Story Prb.	1.40 (.06)	8.41 (.23)	.59 (.03)	3.53 (.25)	.66	.34
Basic Add.	2.41 (.05)	11.79 (.11)	.74 (.01)	3.64 (.10)	.45	.55
Basic Sub.	1.43 (.03)	6.85 (.11)	.74 (.01)	3.53 (.10)	.46	.54
WRAT Arth.	9.29 (.17)	23.73 (.06)	.70 (.02)	1.79 (.06)	.51	.49
Comp Fluency	2.08 (.04)	11.97 (.13)	.84 (.01)	4.84 (.11)	.29	.71
DD Add.	4.24 (.15)	17.00 (.17)	.53 (.04)	2.13 (.18)	.72	.28
DD Sub.	2.00 (.08)	11.06 (.23)	.58 (.03)	3.22 (.24)	.66	.34
Analog Magnitude						
DD Add. Est.	1.20 (.06)	8.38 (.28)	.89 (.02)	6.17 (.29)	.22	.78
DD Sub. Est.	1.09 (.06)	6.39 (.24)	.83 (.03)	4.83 (.25)	.32	.68



$$\chi^2(35) = 547.10, p < .001, RMSEA = .086, CFI = .908$$

**Figure 19. Exact versus Approximate Theory: Arithmetic Measurement Model.**

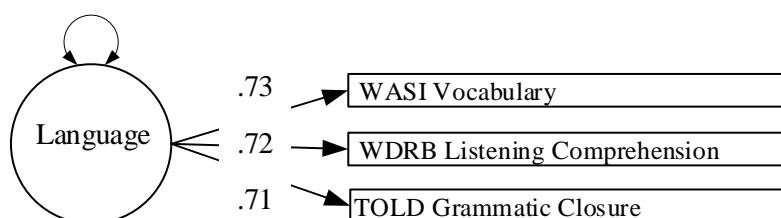
#### 4.2.5 *Language model.*

Language is commonly defined an integration of form, use, and content, a combination of skills in the areas of phonology, syntax, morphology, lexical knowledge, semantics, pragmatics, and prosody (Bloom & Lahey, 1978). Among these possible indicators of language ability, it appears that capturing listening comprehension, vocabulary knowledge, and grammatical comprehension may be essential for accurately measuring language ability (Carroll, 1993), and thus, for the purpose of the current study, these key components of language ability are the focus of measurement.

The language measurement model tests the extent to which three indicators (vocabulary, listening comprehension, and grammatical closures) measure a unitary, latent language ability. With three observed indicators, this latent *language ability* factor model is just-identified (i.e., has zero degrees of freedom), meaning that tests of global fit such as the Chi-square test of model fit, the root mean squared error of approximation (RMSEA), or the comparative fit index (CFI) are trivial. Model solutions for just-identified models are a perfect reproduction of the data input variance-covariance matrix, thus global tests of model fit will reflect a perfect fit between the model and the data (e.g.,  $\chi^2(0) = 0.00$ ,  $RMSEA = 0.00$ ,  $CFI = 1.00$ ; Brown, 2006). However, the extent to which this model reflects a unitary, latent language ability can still be evaluated using local fit indices (i.e., the quality of indicators via factor loadings, indicator residual variances, or alternately, model-based reliability via  $R^2$  statistics). Confirmatory factor analysis (CFA) of the language measurement model, although just-identified, demonstrated good local fit indicative of a single, latent *language* dimension. Completely standardized factor loadings ranged from .71 to .72, indicator residual variances ranged from .47 to .50, and model  $R^2$  ranged from .51 to .53. Table 13 presents standardized and unstandardized results for the Language measurement model. Figure 20 displays a model schematic.

**Table 13. Language Measurement Model CFA Results**

Indicator	Intercept		Factor Loadings		Residual Variance	$R^2$
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Voc	4.24 (.09)	27.35 (.18)	.73 (.02)	4.69 (.19)	.47	.53
List	4.89 (.10)	21.08 (.12)	.72 (.02)	3.12 (.13)	.48	.52
Gram	2.83 (.06)	18.73 (.18)	.71 (.02)	4.71 (.19)	.50	.51



*Model is just identified:  $\chi^2(0) = 0$ , RMSEA = 0, CFI = 1*

**Figure 20. Language Measurement Model.**

#### **4.2.6**      *Executive functioning model.*

Defining and measuring the construct of executive functioning is largely dependent upon the theory of executive functioning or executive control to which one subscribes. For example, Baddeley's (2000) model of executive functioning is primarily focused on the specification of working memory which is defined as a network of specialized cognitive components which function in real time to monitor, process, and maintain information (Baddeley & Hitch, 1974; Baddeley & Logie, 1999; Baddeley, 1992, 2000, 2001). Conversely, Barkley's (1997) model of executive functioning is a theoretical model of self-regulation, attention, and behavioral inhibition which was formulated to add explanatory power for the constellation of poor sustained attentional capacity, impulsivity, and hyperactivity that characterize ADHD. In Zelazo's (2003) model of problem-solving, executive functioning is a temporally organized composition of sub-functions that work in different stages to (1) represent a problem, (2) plan a solution with ordered strategies for implementation, (3) maintain chosen solutions in working memory, along with the rules for their corresponding strategies, and (4) evaluate the results of problem-solving attempts, detecting and correcting errors until the problem is successfully solved (Blair, Zelazo, &

Greenberg, 2005; Zelazo & Frye, 1998; Zelazo et al., 2003). The model of executive functioning used in the current study emphasizes key pieces of executive functioning across various theories of the construct, working memory, attention/inhibition, and non-verbal problem-solving or inductive reasoning.

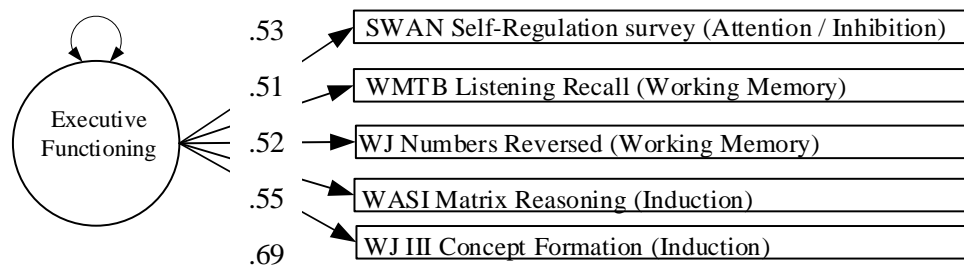
The *executive functioning* measurement model tests the extent to which five indicators measure a unitary, underlying, *executive functioning* ability which was hypothesized to be indicated by a measure of attention and inhibition (the SWAN teacher survey), two measures of verbal working memory (the WJ-III Numbers Reversed and the WMTB-C Listening Recall subtests), and two measures of inductive reasoning and problem-solving (the WASI Matrix Reasoning and the WJ-III Concept Formation subtests). This confirmatory factor model was fit in Mplus 7 via maximum likelihood estimation (ML) for items scored correct or incorrect (Muthen & Muthen, 2012). The fit statistics indicated that this one factor model of *executive functioning* was an approximate good fit for the data, ( $\chi^2(5) = 31.57, p < .001, RMSEA = .06, CFI = .97$ ). Completely standardized factor loadings ranged from .51 to .69, indicator residual variances ranged from .53 to .74, and model  $R^2$  ranged from .26 to .47.

These results are consistent with a unidimensional *executive functioning* dimension, which is being measured with adequate precision; however, the relatively high indicator residual variances would suggest that some important complexity of this dimension is not being modeled here. For the purposes of the current study, capturing several key facets of the construct of executive functioning (self-regulation, attention, inhibition, working memory, and problem-solving/reasoning skills) was the primary aim. This aim has been accomplished with a limited measurement model of executive functioning. Still, it is worth noting that other important theories and other key elements of this complex construct are not being modeled here. The

limited scope of this model of executive functioning is perhaps most apparent in the medium sized indicator factor loadings, the medium to high indicator residual variances, and accordingly, the medium to low model  $R^2$  range. Table 14 presents standardized and unstandardized results for the Executive Function measurement model. Figure 21 displays a model schematic.

**Table 14. Executive Function Measurement Model CFA Results**

Indicator	Intercept	Factor Loadings				Residual Variance	$R^2$
		STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Att.	3.20 (.07)		75.27 (.66)	.53 (.03)	12.43 (.75)	.72	.28
Recall	2.78 (.06)		9.95 (.10)	.51 (.03)	1.82 (.11)	.74	.26
Num. Rev.	3.29 (.07)		9.35 (.08)	.52 (.03)	1.47 (.09)	.73	.27
Matrix Reason.	2.40 (.05)		15.49 (.18)	.55 (.03)	3.57 (.20)	.69	.31
Con. Form.	2.20 (.05)		15.58 (.20)	.69 (.03)	4.85 (.22)	.53	.47



$$\chi^2(5) = 31.57, p < .001, RMSEA = .063, CFI = .971$$

**Figure 21. Executive Function Measurement Model.**

### 4.3 Phase 2: Full Measurement Models for Each Theory

#### 4.3.1 *Abstract code model.*

The full measurement model of Abstract Code Theory was represented with a one factor model of *abstract semantic representation*, which at a minimum, was allowed to correlate with other cognitive domains (e.g., *language*, *executive functioning*). Here, *language* and *executive functioning* are not allowed to predict arithmetic behavioral outcomes, and so, their predictions were fixed at zero (and not drawn) across formats and problem demands. Thus, the individual measurement models for Abstract Code *abstract semantic representation*, *language*, and *executive functioning* discussed in the previous section were combined in a larger measurement model in which they were allowed to correlate.

Global fit statistics indicated that this factor model was not an approximate good fit for the data, ( $\chi^2(141) = 1386.75, p < .001, RMSEA = .07, CFI = .87$ ). Completely standardized factor loadings ranged from .48 to .79; indicator residual variances ranged from .37 to .77; and model  $R^2$  ranged from .23 to .63. Although the correlation between *language* and *abstract semantic representations* was moderate,  $r = .53$ , the correlations between *executive functioning* and *abstract semantic representations* and *executive functioning* and *language* were quite high ( $r = .78$  and  $r = .82$  respectively). Although both the *abstract semantic representations* and *executive functioning* measurement model results suggest that both of these factors are contributing to the model misfit for the Abstract Code Theory full measurement model, the patterns of factor correlation would suggest that the relationships between *executive functioning* and other constructs in the model may also be important sources of model misspecification. Table 15 presents standardized and unstandardized results for the Abstract Code Theory full

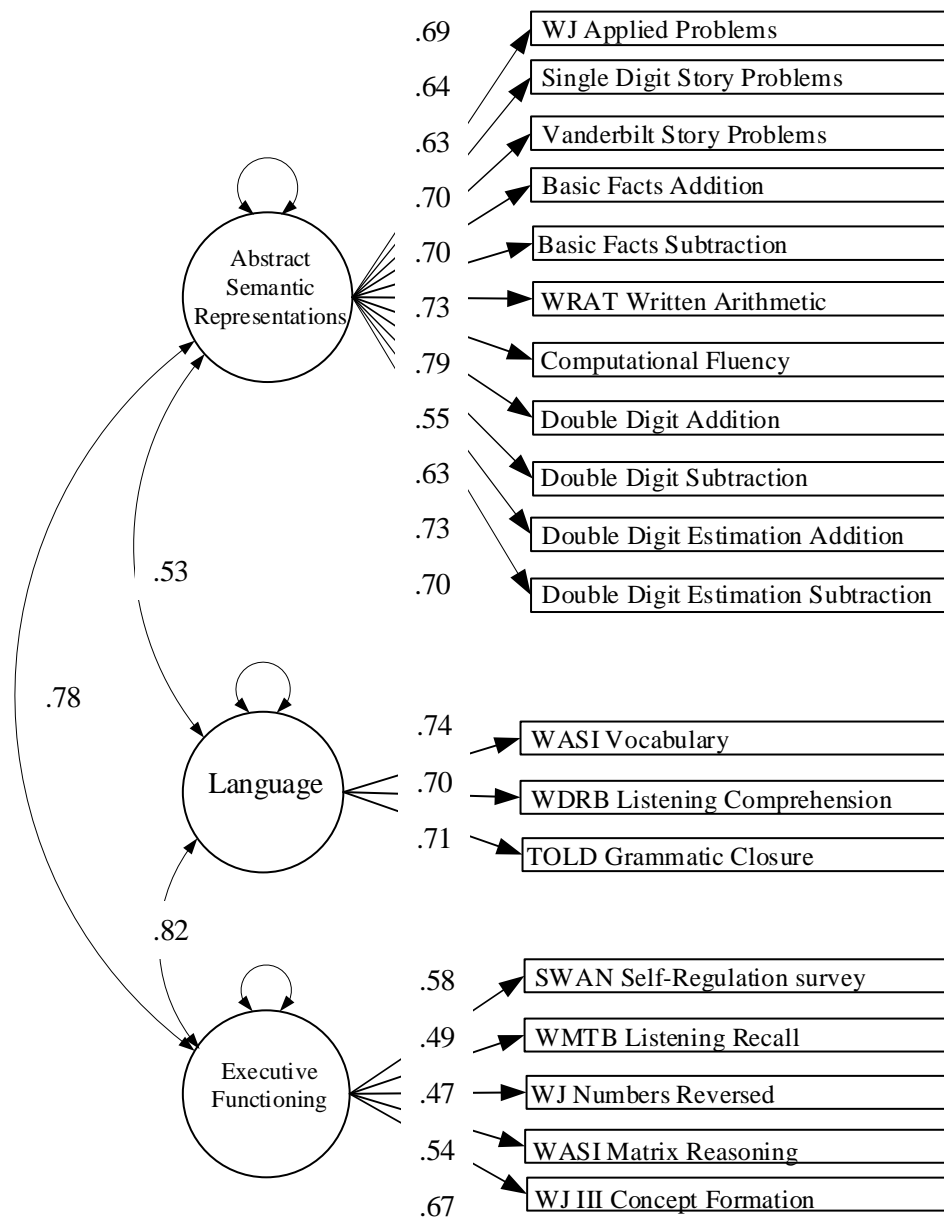
measurement model. Table 16 presents latent factor correlations, and Figure 22 displays a model schematic.

**Table 15. Abstract Code Theory Full Measurement Model CFA Results**

Latent Factor Indicator	Intercept	Factor Loadings			Residual Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Ab. Sem Rep.						
App. Prb.	6.83 (.13)	29.07 (.11)	.69 (.02)	2.95 (.10)	.52	.48
Story Prb.	2.85 (.05)	9.90 (.08)	.64 (.02)	2.23 (.08)	.59	.41
VU Story Prb.	1.40 (.06)	8.40 (.23)	.63 (.03)	3.78 (.24)	.61	.40
Basic Add.	2.41 (.05)	11.79 (.11)	.70 (.01)	3.42 (.10)	.51	.49
Basic Sub.	1.43 (.03)	6.85 (.11)	.70 (.01)	3.37 (.10)	.51	.49
WRAT Arth.	9.30 (.17)	23.71 (.06)	.73 (.02)	1.85 (.06)	.47	.53
Comp Fluency	2.08 (.04)	11.97 (.13)	.79 (.01)	4.55 (.12)	.37	.63
DD Add.	4.24 (.15)	17.04 (.17)	.55 (.04)	2.19 (.18)	.70	.30
DD Sub.	2.00 (.08)	11.10 (.23)	.63 (.03)	3.49 (.23)	.61	.39
DD Add. Est.	1.20 (.06)	8.39 (.27)	.73 (.03)	5.12 (.28)	.47	.54
DD Sub. Est.	1.09 (.05)	6.38 (.23)	.70 (.03)	4.09 (.24)	.52	.48
Language						
Voc	4.25 (.09)	27.29 (.17)	.74 (.02)	4.77 (.17)	.45	.55
List	4.90 (.10)	21.04 (.12)	.70 (.02)	2.99 (.12)	.52	.49
Gram	2.83 (.06)	18.66 (.18)	.71 (.02)	4.69 (.18)	.49	.51
Executive Func.						
Att.	3.20 (.07)	74.96 (.63)	.58 (.02)	13.61 (.64)	.66	.34
Recall	2.79 (.06)	9.92 (.10)	.49 (.02)	1.76 (.10)	.76	.24
Num. Rev.	3.29 (.07)	9.33 (.08)	.48 (.02)	1.35 (.08)	.77	.23
Matrix Reason.	2.41 (.05)	15.45 (.17)	.54 (.02)	3.45 (.18)	.71	.29
Con. Form.	2.21 (.05)	15.50 (.19)	.67 (.02)	4.68 (.19)	.56	.44

**Table 16. Abstract Code Theory Full Model Latent Factor Correlations**

	1	2	3
1. Ab. Sem. Rep.	1.00		
2. Language	.53	1.00	
3. Executive Func.	.78	.82	1.00



$$\chi^2(141) = 1386.75, p < .001, RMSEA = .067, CFI = .865$$

**Figure 22. Abstract Code Theory: Full Measurement Model.**

### 4.3.2 *Encoding complex model.*

Encoding Complex Theory specifies that the competing and sometimes interfering responses to the stimuli must be sorted for relevance, and any interference must be overcome in order for successful performance to occur. Format can interfere with both mental representation of numbers and subsequent numeric processing, and this is especially true for language-formatted problems.

Across various formats and problem demands, the task of the arithmetic under Encoding Complex Theory is largely to inhibit competing and irrelevant signals activated in the encoding complex and to enhance signals that are relevant to the problem. Failure to successfully perform arithmetic constitutes a failure of the system to inhibit. Thus, domains such as *executive systems of control* (working memory, inhibition, attention) and the *language* domain may help to resolve the conflicting signals activated in the encoding complex.

From a factor analytic framework, the roles of *language* and *executive functioning* could be modeled as separate latent variables which are allowed to directly predict arithmetic outcomes along with the *seemingly modular encoding complex for arithmetic*. *Executive functioning* is allowed to predict arithmetic behavioral outcomes across various formats and problem demands; however, *language* is allowed to predict arithmetic behavioral outcomes for language-formatted problems. The extent to which the *language* domain helps to explain performance on non-language-formatted items is not specified by Encoding Complex Theory, although we would expect that the *language* domain may make little or no contribution to non-language-formatted items.

Global fit statistics indicated that this factor model was an approximate good fit for the data, ( $\chi^2$  (128) = 478.80,  $p < .001$ ,  $RMSEA = .04$ ,  $CFI = .96$ ). Completely standardized factor

loadings ranged from .05 (non-significant) to .74; indicator residual variances ranged from .25 to .77; and model  $R^2$  ranged from .23 to .75. As mentioned in the executive function measurement model results, the residuals for this factor were undesirably high and among the highest in the model. However, *executive function* was a significant and salient predictor of all arithmetic outcomes, and *language* was a significant predictor of WJ Applied Problems and Single Digit Story Problems, though these loadings were quite low. Allowing for direct prediction of arithmetic outcomes by *executive function* and *language* left little unique predictive power for the *seemingly modular encoding complex*; however, each arithmetic outcome was still significantly predicted by something other than *executive function* and *language* (represented here by the *seemingly modular encoding complex*). Three outcomes in particular (Basic Facts Addition, Basic Facts Subtraction, and Computational Fluency, all of which were formatted with Arabic numerals and involved relatively small problem sizes) had high *encoding complex* factor loadings despite the addition of *executive function* as a predictor.

Because *executive function* was a direct predictor of arithmetic outcomes in this model, the correlation between *executive function* and the *seemingly modular encoding complex* was restricted to zero for the purpose of model specification. The correlation between *executive function* and *language* was large and positive,  $r=.77$ ; however, the correlation between *language* and the *encoding complex* was small and negative,  $r=-.11$ . These results would seem to indicate that although *language* is a small but significant predictor of outcomes in language-formatted problems, it is not a predictor of outcomes in Arabic numeral formatted problems or estimation problems. Tables 17 and 18 present unstandardized and standardized results (respectively) for the Encoding Complex Theory full measurement model. Table 19 presents latent factor correlations, and Figure 23 displays a model schematic.

**Table 17. Encoding Complex Theory Full Measurement Model Unstandardized CFA Results**

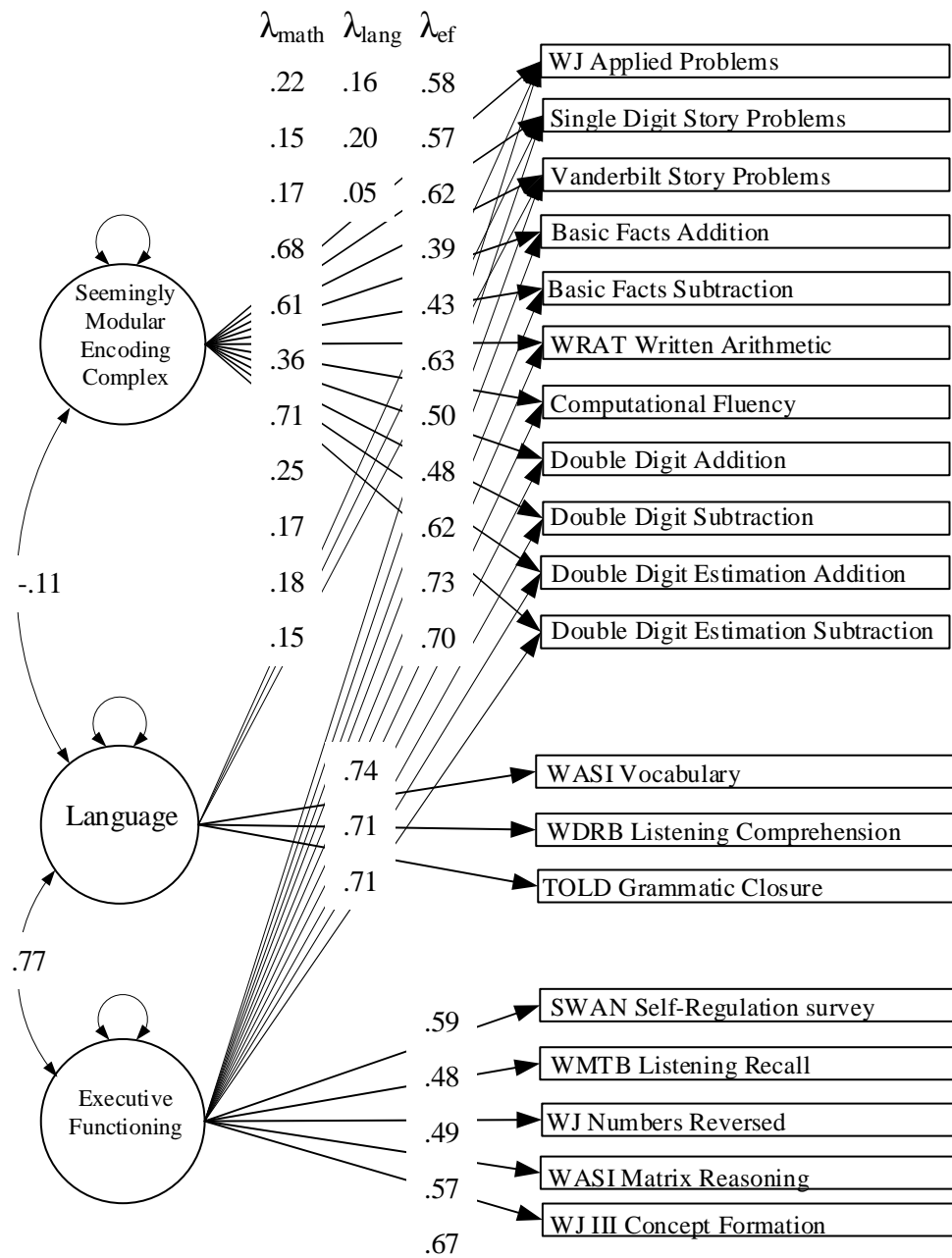
Indicator	Intercepts (SE)	Factor Loadings (SE) by Factor			Residual Variance	R <sup>2</sup>
		Seemingly Modular Encoding Complex	Language	Executive Function		
Arithmetic Measures						
App. Prb.	29.06 (.11)	.93 (.12)	.69 (.22)	2.49 (.22)	8.42	.54
Story Prb.	9.90 (.08)	.51 (.09)	.69 (.18)	1.97 (.17)	5.49	.55
VU Story Prb.	8.37 (.22)	1.03 (.30)	.29 (.67) <sup>NS</sup>	3.75 (.63)	19.44	.46
Basic Add.	11.79 (.11)	3.32 (.11)		1.90 (.13)	9.25	.61
Basic Sub.	6.85 (.11)	2.92 (.11)		2.07 (.13)	10.18	.56
WRAT Arth.	23.71 (.06)	.92 (.07)		1.60 (.07)	3.10	.52
Comp Fluency	11.96 (.13)	4.09 (.13)		2.87 (.15)	8.13	.75
DD Add.	17.07 (.17)	.99 (.20)		1.91 (.18)	11.43	.29
DD Sub.	11.23 (.23)	.95 (.26)		3.42 (.24)	18.05	.41
DD Add. Est.	8.61 (.27)	1.24 (.32)		5.11 (.28)	20.94	.57
DD Sub. Est.	6.58 (.23)	.89 (.28)		4.10 (.25)	16.77	.51
Language Measures						
Voc	27.30 (.17)		4.79 (.17)		18.81	.55
List	21.05 (.12)		3.08 (.12)		9.16	.51
Gram	18.68 (.18)		4.71 (.18)		21.86	.50
Exec. Func. Measures						
Att.	75.03 (.63)			13.88 (.64)	361.25	.35
Recall	9.93 (.10)			1.70 (.10)	9.89	.23
Num. Rev.	9.34 (.08)			1.40 (.08)	6.15	.24
Matrix Reason.	15.46 (.17)			3.68 (.17)	27.94	.33
Con. Form.	15.53 (.19)			4.71 (.19)	27.85	.44

**Table 18. Encoding Complex Theory Full Measurement Model Completely Standardized CFA Results**

Indicator	Intercepts (SE)	Factor Loadings (SE) by Factor			Residual Variance	R <sup>2</sup>
		Seemingly Modular Encoding Complex	Language	Executive Function		
Arithmetic Measures						
App. Prb.	6.77 (.13)	.22 (.03)	.16 (.05)	.58 (.05)	.46	.54
Story Prb.	2.85 (.05)	.15 (.03)	.20 (.05)	.57 (.05)	.45	.55
VU Story Prb.	1.39 (.05)	.17 (.05)	.05 (.11) <sup>NS</sup>	.62 (.10)	.54	.46
Basic Add.	2.41 (.05)	.68 (.02)		.39 (.03)	.39	.61
Basic Sub.	1.43 (.03)	.61 (.02)		.43 (.02)	.44	.56
WRAT Arth.	9.29 (.17)	.36 (.03)		.63 (.02)	.48	.52
Comp Fluency	2.08 (.04)	.71 (.02)		.50 (.02)	.25	.75
DD Add.	4.26 (.15)	.25 (.05)		.48 (.04)	.71	.29
DD Sub.	2.03 (.08)	.17 (.05)		.62 (.03)	.59	.41
DD Add. Est.	1.24 (.06)	.18 (.05)		.73 (.03)	.43	.57
DD Sub. Est.	1.12 (.05)	.15 (.05)		.70 (.03)	.49	.51
Language Measures						
Voc	4.23 (.09)		.74 (.02)		.45	.55
List	4.87 (.10)		.71 (.02)		.49	.51
Gram	2.81 (.06)		.71 (.02)		.50	.50
Exec. Func. Measures						
Att.	3.19 (.07)			.59 (.02)	.65	.35
Recall	2.78 (.06)			.48 (.02)	.77	.23
Num. Rev.	3.28 (.07)			.49 (.02)	.76	.24
Matrix Reason.	2.40 (.05)			.57 (.02)	.67	.33
Con. Form.	2.20 (.05)			.67 (.02)	.56	.44

**Table 19. Encoding Complex Theory Full Model Latent Factor Correlations**

	1	2	3
1. Encode Comp.	1.00		
2. Language	-.11	1.00	
3. Executive Func.	@0	.77	1.00



$$\chi^2(128) = 478.80, p < .001, RMSEA = .037, CFI = .962$$

**Figure 23. Encoding Complex Theory: Full Measurement Model.**

### 4.3.3 *Triple code model.*

Triple Code Theory allows for the *language* domain to inform the *quantitative* domain by providing linguistically stored math facts. Although the *auditory verbal* module is responsible for mentally representing written (graphemes) or spoken (phonemes) numbers syntactically, by place value, the extent to which the *language* domain may or may not overlap with the *auditory verbal* module of Triple Code Theory is unclear. Similarly, an *attentional control* domain is allowed to coordinate visuo-spatial attention to numbers on the internal number line, but the extent to which this *attentional control* domain helps to coordinate the working memory, inhibition, and planning required to complete numeric processing is not specified.

From a factor analytic framework, a latent *language* factor and an *executive control* factor may communicate with the three modules of Triple Code Theory's *quantitative* domain, and at a minimum, these additional domains may correlate with the numeric processing domains of Triple Code Theory. Thus, *executive function* and *language* were allowed to correlate freely with Triple Code Theory's *auditory verbal*, *visual*, and *analog magnitude* modules in the full measurement model for Triple Code Theory.

Global fit statistics indicated that this factor model was an approximate good fit for the data, ( $\chi^2 (134) = 592.06, p < .001, RMSEA = .04, CFI = .95$ ). Completely standardized factor loadings ranged from .48 to .89; indicator residual variances ranged from .21 to .77; and model  $R^2$  ranged from .23 to .79. As in the Triple Code Theory arithmetic measurement model, the arithmetic portion of this full model was very strong. Completely standardized factor loadings ranged from .52 to .89, and factor correlations for this portion of the model ranged from  $r = .67$  to  $r = .74$ , indicating that each of Triple Code Theory's arithmetic cognition modules were separable but highly related.

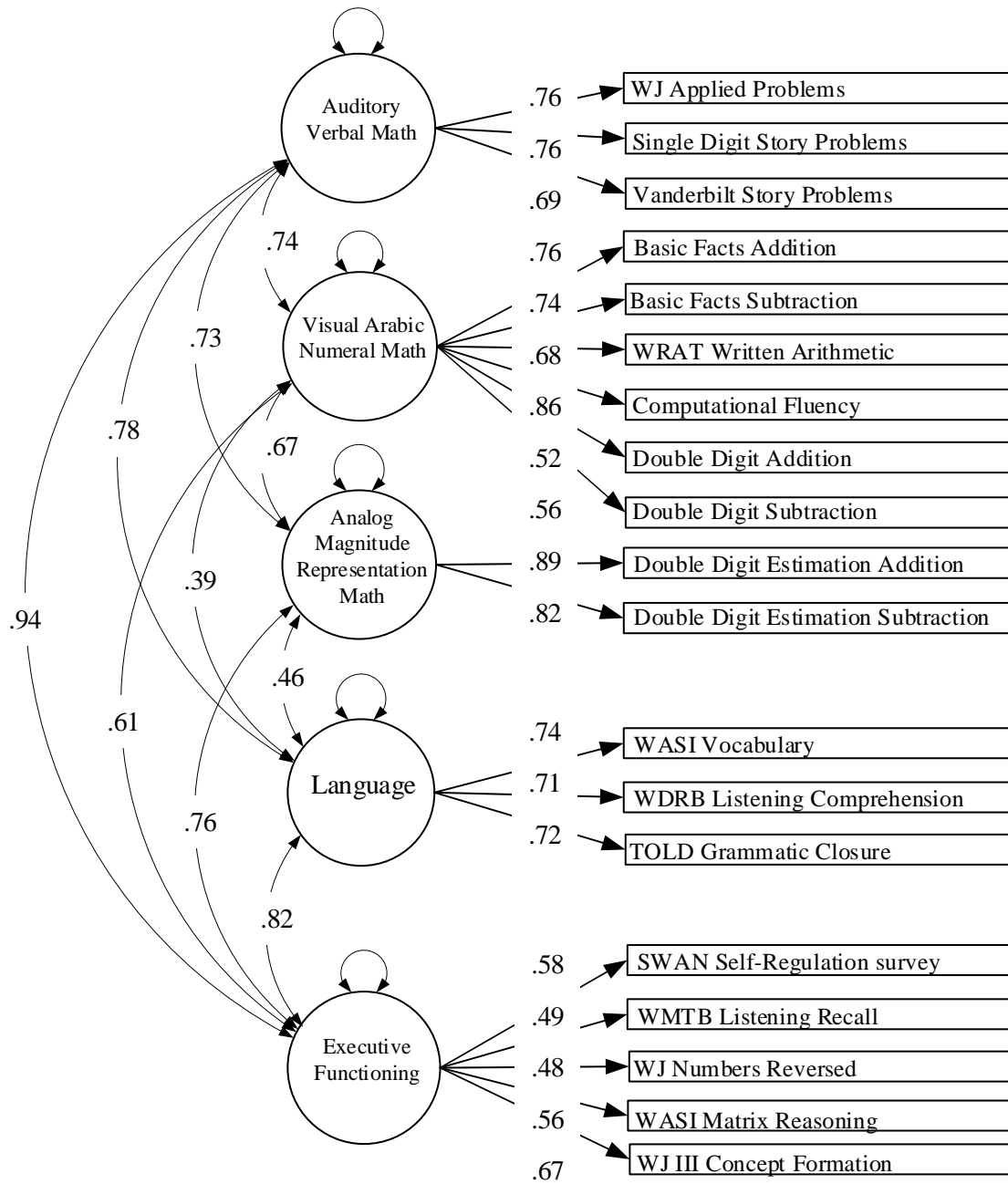
Again the *executive function* measurement model results demonstrated undesirably high residuals. However, executive functioning factor loadings indicated that the selected outcomes were all significant and salient indicators of this factor. The executive functioning factor correlated highly with all other factors in the Triple Code Theory model (see Table 20). The addition of *executive functioning*, in particular, raised some structural questions for the arithmetic portion of the Triple Code Theory model. Specifically, the relationship between *executive functioning* and the *auditory verbal module* was nearly at singularity,  $r=.94$ , and the relationship between language and the auditory verbal module was also quite high,  $r=.78$ . Taken together, these results indicate that (1) problem formatting should be explicitly accounted for in modeling arithmetic outcomes, (2) *executive functioning* and *language* may both play important roles in facilitating arithmetic cognition across various problem formats, but (3) language-formatted items in particular may be predicted by domains other than a specialized quantitative domain. Table 20 presents standardized and unstandardized results for the Triple Code Theory full measurement model. Table 21 presents latent factor correlations, and Figure 24 displays a model schematic.

**Table 20. Triple Code Theory Full Measurement Model CFA Result**

Latent Factor Indicator	Intercept	Factor Loadings		Residual Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)	
Auditory Verbal					
App. Prb.	6.76 (.13)	29.05 (.11)	.76 (.02)	3.27 (.11)	.42 .58
Story Prb.	2.85 (.05)	9.90 (.08)	.76 (.01)	2.63 (.08)	.43 .57
VU Story Prb.	1.37 (.06)	8.27 (.22)	.69 (.03)	4.18 (.24)	.52 .48
Visual					
Basic Add.	2.41 (.05)	11.79 (.11)	.76 (.01)	3.71 (.10)	.42 .58
Basic Sub.	1.43 (.03)	6.85 (.11)	.74 (.01)	3.57 (.10)	.45 .55
WRAT Arth.	9.29 (.17)	23.74 (.06)	.68 (.02)	1.75 (.06)	.53 .47
Comp Fluency	2.08 (.04)	11.97 (.13)	.86 (.01)	4.95 (.11)	.26 .74
DD Add.	4.24 (.15)	16.98 (.17)	.52 (.04)	2.08 (.18)	.73 .27
DD Sub.	2.00 (.08)	11.04 (.23)	.56 (.04)	3.08 (.25)	.69 .31
Analog Mag.					
DD Add. Est.	1.23 (.06)	8.56 (.27)	.89 (.02)	6.18 (.27)	.21 .79
DD Sub. Est.	1.12 (.06)	6.54 (.24)	.82 (.02)	4.80 (.24)	.33 .67
Language					
Voc	4.23 (.09)	27.29 (.17)	.74 (.02)	4.77 (.17)	.45 .55
List	4.88 (.10)	21.04 (.12)	.71 (.02)	3.05 (.12)	.50 .50
Gram	2.84 (.06)	18.66 (.18)	.72 (.02)	4.75 (.18)	.49 .51
Executive Func.					
Att.	3.19 (.07)	75.00 (.63)	.58 (.02)	13.55 (.64)	.67 .33
Recall	2.78 (.06)	9.93 (.10)	.49 (.02)	1.75 (.10)	.76 .24
Num. Rev.	3.28 (.07)	9.33 (.08)	.48 (.02)	1.37 (.08)	.77 .23
Matrix Reason.	2.41 (.05)	15.45 (.17)	.56 (.02)	3.59 (.17)	.69 .31
Con. Form.	2.20 (.05)	15.51 (.19)	.67 (.02)	4.74 (.19)	.55 .45

**Table 21. Triple Code Theory Full Model Latent Factor Correlations**

	1	2	3	4	5
1. Aud. Verb.	1.00				
2. Visual	.74	1.00			
3. Analog Mag.	.73	.67	1.00		
4. Language	.78	.39	.46	1.00	
5. Executive Func.	.94	.61	.76	.82	1.00



$$\chi^2(134) = 592.06, p < .001, RMSEA = .042, CFI = .950$$

**Figure 24. Triple Code Theory: Full Measurement Model.**

#### 4.3.4 *Exact versus approximate model.*

The contributions of domains other than the *analog magnitude* module and the *auditory verbal* module are unclear in the Exact Versus Approximate specification of Triple Code Theory. Language networks which also contribute to purely linguistic tasks are implicated in exact calculations, but the extent to which the *language* domain represents a unique contribution beyond the *auditory verbal* module is unclear. They may overlap so much that they do not appear to be separate domains, which would indicate that the *auditory verbal* module is in fact the *language* domain and does not have unique predictive power in numeric processing. From a factor analytic framework, this can be represented with a latent *language* factor, which is allowed to correlate with the *auditory verbal module*. An extremely high correlation would indicate that they are not distinct factors. Without a compelling reason to restrict associations with other aspects of the model, this *language* domain can also be allowed to correlate with the analog magnitude module.

Similarly, the role of executive control in coordinating processing and facilitating spatial attention is unclear. The *visual Arabic number form* module is largely absent from this specification of Triple Code Theory; however, spatial attention networks, possibly representing some of the predictive power of the *visual Arabic number form* module and possibly representing some form of executive control for attention, may contribute to coordinating both types of task. Without more information on the contributions of "spatial attention networks", this domain can be tentatively represented with a latent factor for *executive control* which is allowed to associate with the *analog magnitude* and *auditory verbal* modules as well as with the *language* domain.

The Exact versus Approximate Calculations hypothesis of Triple Code Theory may be best represented as a four factor model in which both the *analogical magnitude representation*

domain and an *executive* domain coordinating attention contribute to all numeric processing tasks, but the *language* domain and possibly a unique *auditory verbal* module contribute only to tasks requiring exact calculations. For exact calculations problems that involve number facts that are either unknown or not commonly used (e.g., double digit addition and subtraction problems), the *analog magnitude* module may also contribute to arithmetic cognition; however, one would not necessarily expect that the *analog magnitude* module contributes to most exact calculation problems.

Global fit statistics indicated that this factor model was not an approximate good fit for the data, ( $\chi^2 (138) = 1230.11, p < .001, RMSEA = .06, CFI = .88$ ). Completely standardized factor loadings ranged from .46 to .88; indicator residual variances ranged from .22 to .77; and model  $R^2$  ranged from .23 to .78. Although both the *exact versus approximate calculations* and *executive functioning* measurement model results suggest that all of these factors are contributing to the model misfit for the Exact Versus Approximate Calculations full measurement model, the patterns of factor correlation would suggest that the relationships between *executive functioning* and other constructs in the model may also be important sources of model misspecification.

*Executive function* correlated significantly and strongly with all other factors in the model (see Table 22). In so far as this *executive function* factor overlaps with Exact Versus Approximate Calculation Theory's spatial networks of control, it would seem to indicate that executive systems of control may indeed play a role in facilitating both exact and approximate calculations. *Language*, however, correlated only moderately with the *auditory verbal* and *analog magnitude* modules, but it correlated highly with *executive function*. Taken together, this pattern of correlations would seem to suggest that language is separable from traits predicting arithmetic outcomes across exact and approximate problem demands, which are in turn both

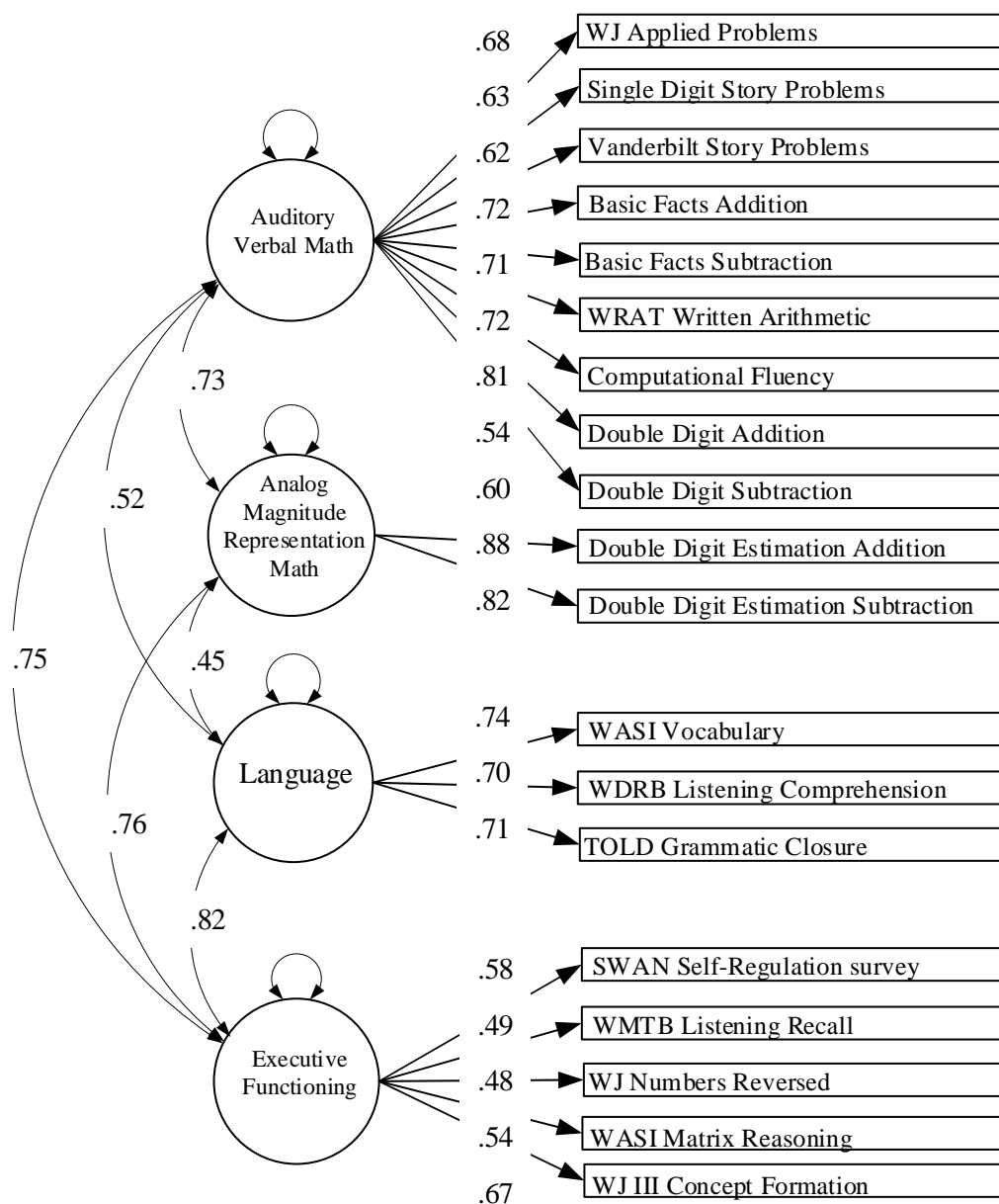
highly related and separate from each other (*auditory verbal* and *analog magnitude* modules correlated at  $r=.73$ ). Table 22 presents standardized and unstandardized results for the Triple Code Theory full measurement model. Table 23 presents latent factor correlations, and Figure 25 displays a model schematic.

**Table 22. Exact V. Approximate Calculations Full Measurement Model CFA Results**

Latent Factor Indicator	Intercept	Factor Loadings			Residual Variance	R <sup>2</sup>
	STD (SE)	UnSTD (SE)	STD (SE)	UnSTD (SE)		
Auditory Verbal						
App. Prb.	6.84 (.13)	29.07 (.11)	.68 (.02)	2.91 (.10)	.53	.47
Story Prb.	2.85 (.05)	9.90 (.08)	.63 (.02)	2.19 (.08)	.60	.40
VU Story Prb.	1.39 (.06)	8.40 (.23)	.62 (.03)	3.75 (.25)	.61	.39
Basic Add.	2.41 (.05)	11.79 (.11)	.72 (.01)	3.50 (.10)	.49	.51
Basic Sub.	1.43 (.03)	6.85 (.11)	.71 (.01)	3.42 (.10)	.49	.51
WRAT Arth.	9.29 (.17)	23.72 (.06)	.72 (.02)	1.85 (.06)	.48	.52
Comp Fluency	2.08 (.04)	11.97 (.13)	.81 (.01)	4.65 (.12)	.34	.66
DD Add.	4.25 (.15)	17.03 (.17)	.54 (.04)	2.17 (.18)	.71	.29
DD Sub.	2.01 (.08)	11.10 (.23)	.60 (.03)	3.34 (.24)	.64	.36
Analog Mag.						
DD Add. Est.	1.23 (.06)	8.52 (.27)	.88 (.02)	6.12 (.27)	.22	.78
DD Sub. Est.	1.11 (.06)	6.50 (.24)	.82 (.02)	4.81 (.24)	.32	.68
Language						
Voc	4.26 (.09)	27.29 (.17)	.74 (.02)	4.77 (.17)	.45	.55
List	4.91 (.10)	21.04 (.12)	.70 (.02)	2.99 (.17)	.52	.49
Gram	2.83 (.06)	18.66 (.18)	.71 (.02)	4.70 (.18)	.49	.51
Executive Func.						
Att.	3.21 (.07)	75.00 (.63)	.58 (.02)	13.57 (.64)	.66	.34
Recall	2.79 (.06)	9.93 (.10)	.49 (.02)	1.76 (.10)	.76	.24
Num. Rev.	3.29 (.07)	9.33 (.08)	.48 (.02)	1.35 (.08)	.77	.23
Matrix Reason.	2.41 (.05)	15.45 (.17)	.54 (.02)	3.45 (.18)	.71	.29
Con. Form.	2.21 (.05)	15.51 (.19)	.67 (.02)	4.67 (.19)	.56	.44

**Table 23. Exact V. Approximate Calculations Full Model Latent Factor Correlations**

	1	2	3	4
1. Aud. Verb.	1.00			
2. Analog Mag.	.73	1.00		
3. Language	.52	.45	1.00	
4. Executive Func.	.75	.76	.82	1.00



$$\chi^2(138) = 1230.11, p < .001, RMSEA = .064, CFI = .881$$

**Figure 25. Exact versus Approximate Theory: Full Measurement Model.**

#### 4.3.5 *Post-hoc testing: Hybrid full measurement model.*

Results from the arithmetic only measurement models indicated that the Triple Code Theory model of arithmetic was the best fitting model; however, the Triple Code Theory full measurement model displayed some structural problems, namely a correlation between *executive function* and the *auditory verbal* module that was near singularity and very high correlations between *executive function* and the other modules of arithmetic in the model.

Conversely, results from the Encoding Complex full measurement model indicated that this model of arithmetic (and its relationships with other cognitive domains) was the best fitting model; however, the architecture for arithmetic in the Encoding Complex Theory model was unidimensional, and results from the arithmetic only measurement models indicated that a unidimensional arithmetic was not a good fit for the data.

Given the findings that (1) a three-factor model of arithmetic presented by Triple Code Theory was an excellent fit for the data, and (2) a direct prediction of *executive function* and *language* on math outcomes presented by Encoding Complex Theory was an excellent fit for the data, a final post-hoc model that combined these specifications was tested. This model represents Triple Code Theory's specification that three, format-specific modules are responsible for processing various types of arithmetic problems and that these modules are allowed to communicate via the process of transcoding. A *visual Arabic* module processes digital input and output as well as multi-digit operations. An *auditory verbal* module processes simple mathematical facts, language-based input and output, and language-based memory for numbers. An *analog magnitude* module processes semantic information for number and is responsible for performing comparison, estimation, approximate calculation, and subitizing tasks across various formats of input and output. Transcoding allows for these modules to inform one another directly

during numeric processing tasks. The post-hoc hybrid model represents each of these modules as a latent factor and transcoding as the correlation between these factors.

The post-hoc hybrid model also represents Encoding Complex Theory's specification that successful processing requires the sorting of stimulus responses for relevance to the problem-solving task and the inhibition of responses that are irrelevant to solving the problem. Format interferences are expected, particularly for language-formatted stimuli. Thus, *executive function* is a direct predictor of arithmetic outcomes across formats and problem demands, and *language* ability is expected to directly contribute to language-formatted problems. Triple Code Theory's specification of format-sensitive arithmetic cognition modules would appear to be compatible with Encoding Complex Theory's specification that format effects can interfere with mental representation of problems, subsequent numeric processing, and answer production.

Global fit statistics indicated that this factor model was an approximate excellent fit for the data, ( $\chi^2(124) = 341.71, p < .001, RMSEA = .03, CFI = .98$ ). Across outcomes, completely standardized factor loadings ranged from  $-.02$  (non-significant) to  $.74$ ; indicator residual variances ranged from  $.25$  to  $.77$ ; and model  $R^2$  ranged from  $.23$  to  $.76$ . As mentioned in the executive function measurement model results, the residuals for this factor were undesirably high and among the highest in the model.

More specifically, for the arithmetic outcomes across the three Triple Code modules, completely standardized factor loadings ranged from  $.18$  to  $.71$ . All of these loadings were significant, but only the factor loadings for the following five arithmetic outcomes were salient: Basic Facts Addition, Basic Facts Subtraction, and Computational Fluency (all Arabic numeral formatted and all involving relatively small problem sizes), as well as Double Digit Estimation Addition and Double Digit Estimation Subtraction (both involving estimation). For the language

outcomes, completely standardized factor loadings were all significant and salient, ranging from .71 to .74; however, none of the language-formatted arithmetic outcomes were significant indicators of language, meaning that the *auditory verbal* module is distinct from *language*. For the executive function outcomes, completely standardized factor loadings were all significant and salient, ranging from .49 to .67. The arithmetic outcomes were all significantly and saliently indicated by the executive function factor. Completely standardized factor loadings ranged from .39 to .69, and they were lowest for the three aforementioned Arabic numeral formatted / small problem size outcomes (Basic Facts Addition, Basic Facts Subtraction, and Computational Fluency).

Allowing for direct prediction of arithmetic outcomes by *executive function* and *language* left little unique predictive power for the three Triple Code Theory modules of arithmetic; however, all of the arithmetic outcomes were still significantly predicted by its corresponding Triple Code Theory module. This pattern of results indicates that something other than *executive function* and *language* (represented here by the *visual Arabic number form* module, *auditory verbal* module, and *analog magnitude* module) was predicting performance for each of these problem formats or analogical magnitude demands. The *auditory verbal* module factor loadings were all particularly low with *executive function* in the model, which would seem to indicate that language-formatted problems, in particular, are largely an *executive function* task.

Because *executive function* was a direct predictor of arithmetic outcomes in this model, the correlations between *executive function* and the *visual Arabic number form* module, the *auditory verbal* module, and the *analog magnitude* module were restricted to zero for the purpose of model specification. Similarly, the correlation between *language* and the *auditory verbal* module was also restricted to zero. The correlation between *executive function* and

*language* was large and positive,  $r=.80$ ; however, the correlations between *language* and both the *visual Arabic number form* and *analog magnitude* modules were small and negative,  $r=-.13$  and  $r=-.28$  respectively. Among the Triple Code Theory modules, *auditory verbal* arithmetic and *visual Arabic number form* arithmetic were moderately and positively related,  $r=.63$ , and *analog magnitude* arithmetic and *visual Arabic number form* arithmetic were slightly and positively related,  $r=.35$ . However, the *auditory verbal* and *analog magnitude* modules were not significantly related.

These results would seem to indicate that (1) language may play some role in facilitating executive function's prediction of arithmetic outcomes, (2) across formats of arithmetic problems, language ability is not related to performance when the contribution of executive function is explicitly modeled, (3) the *auditory verbal* module moderately related to Arabic numeral formatted, exact calculation problems, and (4) although Arabic numeral formatted problems may call on some of the same faculties used for estimation / analogical magnitude problems, language-formatted problems appear to have some common method variance that is distinct from analogical magnitude.

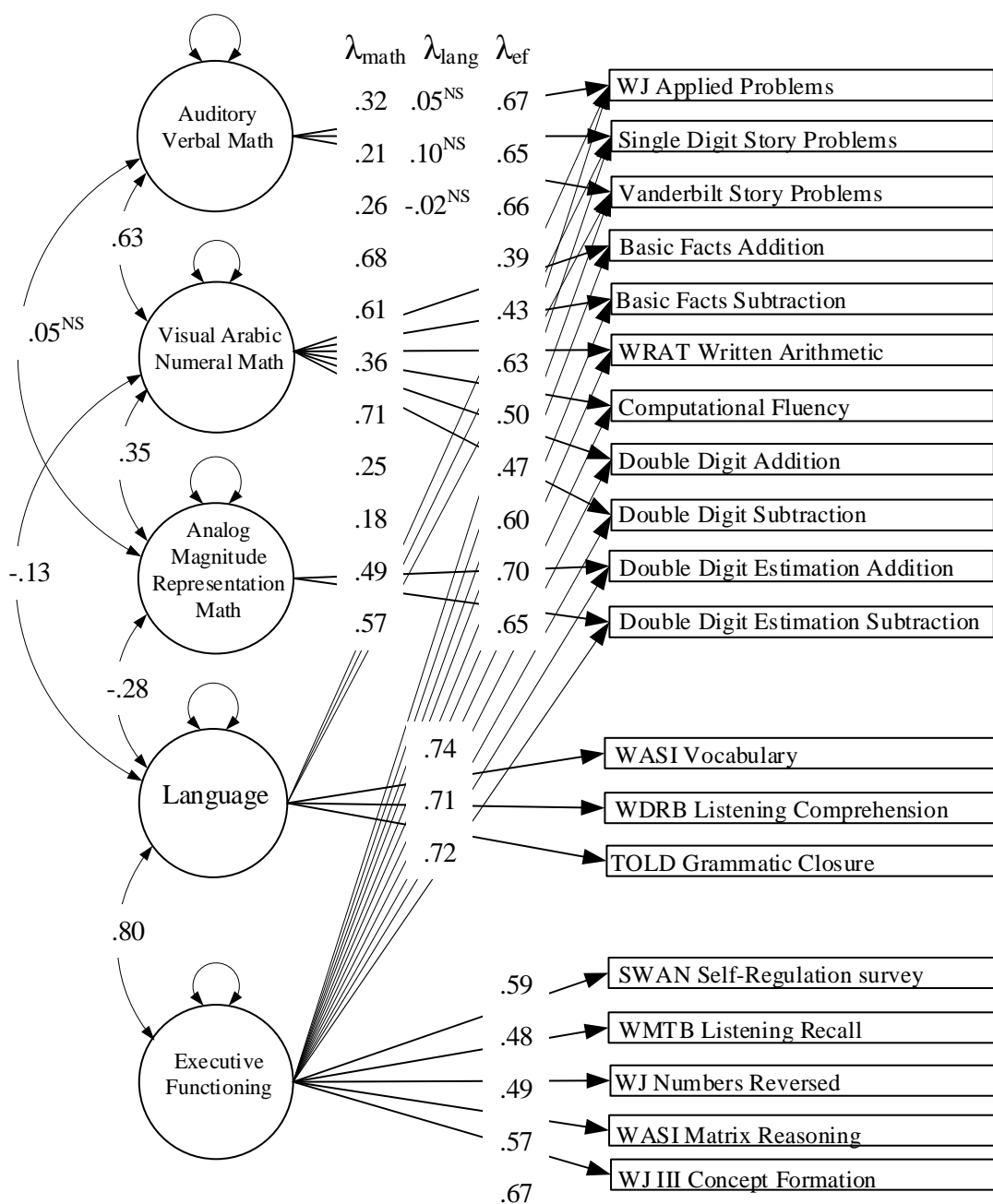
Table 24 presents completely standardized results for the Hybrid full measurement model. Table 25 presents latent factor correlations, and Figure 26 displays a model schematic.

**Table 24. Post Hoc Hybrid Full Measurement Model Completely Standardized CFA Results**

Indicator	Intercepts (SE)	Factor Loadings (SE) by Factor					Residual Variance	R <sup>2</sup>
		Auditory Verbal	Visual	Analog Mag.	Language	Exec. Function		
Arithmetic Measures								
App. Prb.	6.76 (.13)	.32 (.05)			.05 (.06) <sup>NS</sup>	.67 (.06)	.40	.60
Story Prb.	2.85 (.05)	.21 (.04)			.10 (.06) <sup>NS</sup>	.65 (.05)	.43	.57
VU Story Prb.	1.38 (.05)	.26 (.07)			-.02 (.12) <sup>NS</sup>	.66 (.11)	.52	.48
Basic Add.	2.41 (.05)		.68 (.02)			.39 (.03)	.39	.61
Basic Sub.	1.43 (.03)		.61 (.02)			.43 (.02)	.44	.56
WRAT Arth.	9.28 (.17)		.36 (.03)			.63 (.02)	.48	.52
Comp Fluency	2.08 (.04)		.71 (.02)			.50 (.02)	.24	.76
DD Add.	4.27 (.15)		.25 (.05)			.47 (.04)	.72	.28
DD Sub.	2.04 (.08)		.18 (.05)			.60 (.03)	.60	.40
DD Add. Est.	1.24 (.06)			.49 (.06)		.70 (.03)	.28	.72
DD Sub. Est.	1.12 (.06)			.57 (.06)		.65 (.04)	.25	.75
Language Measures								
Voc	4.23 (.09)				.74 (.02)		.45	.55
List	4.87 (.10)				.71 (.02)		.50	.50
Gram	2.81 (.06)				.72 (.02)		.49	.51
Exec. Func. Measures								
Att.	3.19 (.07)					.59 (.02)	.65	.35
Recall	2.78 (.06)					.48 (.02)	.77	.23
Num. Rev.	3.28 (.07)					.49 (.02)	.76	.24
Matrix Reason.	2.40 (.05)					.57 (.02)	.68	.32
Con. Form.	2.19 (.05)					.67 (.02)	.55	.45

**Table 25. Post Hoc Hybrid Full Model Latent Factor Correlations**

	1	2	3		
1. Auditory Verbal	1.00				
2. Visual Ar. Num.	.63	1.00			
3. Analog Magnitude	.05 <sup>NS</sup>	.35	1.00		
4. Language	@0	-.13	-.28	1.00	
5. Executive Func.	@0	@0	@0	.80	1.00



$$\chi^2(124) = 341.71, p < .001, RMSEA = .030, CFI = .976$$

**Figure 26. Hybrid Model of Triple Code Theory Arithmetic and Encoding Complex Theory Structure as a Full Measurement Model**

#### 4.3.6 *Summary of Model Testing Results.*

Model testing began with the examination of distinct portions of what would later become full measurement models. This phase of model testing began with an examination of the arithmetic portions of measurement for each of the four theories considered in this study. The Abstract Code Theory model of arithmetic tested the extent to which arithmetic behavioral outcomes could be explained by one, latent form of mental representation (*abstract semantic codes*) across a variety of problem formats and demands. This model was not a good fit for the data. The Encoding Complex Theory model of arithmetic tested the extent to which arithmetic behavioral outcomes could be explained by a unitary, latent *encoding complex* across a variety of problem formats and demands which appears to be modular with practice. This model was structurally identical to the Abstract Code Theory model of arithmetic and was also not a good fit for the data. The Triple Code Theory model of arithmetic tested the extent to which arithmetic behavioral outcomes could be explained by three, latent modules with format and problem demand specific responsibilities in numeric processing. This model was an approximate good fit for the data. The Exact Versus Approximate Calculations model of arithmetic tested the extent to which arithmetic behavioral outcomes could be explained by two, latent modules with problem demand specific responsibilities in numeric processing. This model was not a good fit for the data.

Measurement models for *language* and *executive function* also were examined during this phase of model testing. The language measurement model was just-identified with three indicators of *language* ability. Though global fit could not be examined for this model, local fit statistics indicated that these three indicators were measures of the same underlying dimension. The *executive function* measurement model was over-identified with five indicators of *executive*

*function* ability. Though the global and local fit statistics indicated that this model was an approximate good fit for the data, all indicators in this model demonstrated relatively high residuals. The *executive function* model, though adequate for the purposes of the current study, evidenced issues of fit that could be interpreted to mean that important complexity in this construct was not being modeled with a unitary conceptualization. Ultimately, the Triple Code Theory model of arithmetic was the best fitting model for arithmetic during this phase of model testing. Both the *language* and *executive function* models were also retained for the next phase of testing.

The next phase of model testing examined each of the four theories of arithmetic cognition with the inclusion of *language* and *executive function* abilities in full measurement models. Results from the first phase of model testing were crucial for identifying sources of model misfit during this phase of testing.

The Abstract Code Theory full measurement model tested the extent to which both *language* and *executive function* contributed to but did not directly predict arithmetic outcomes across a variety of problem formats and demands, which were represented by one, format-independent, *abstract semantic code*. This model was not a good fit for the data. Phase one results indicated that both the *abstract semantic code* and *executive function* measurement portions of this full model were important sources of model misfit. However, patterns of high correlations between factors indicated that the specifications of relationships between *language*, *executive functions*, and *abstract semantic code* may also have been important sources of model misspecification. This model was ultimately rejected.

The Encoding Complex Theory full measurement model tested the extent to which *executive function* was a direct predictor of all arithmetic behavioral outcomes and *language*

ability was a direct predictor of outcomes on arithmetic problems with language formats.

Arithmetic behavioral outcomes were modeled as a unitary, *encoding complex*, which appears to be modular with practice. This model was an approximate good fit for the data, despite the contributions of the *seemingly modular encoding complex* and *executive function* to model misfit. *Executive function* was a significant predictor of all arithmetic outcomes in the model, and *language* was a significant predictor of two language-formatted arithmetic outcomes.

Furthermore, once *executive function* and *language* were directly modeled as predictors of arithmetic outcomes, *language* evidenced a negative correlation with the remaining (non-language-formatted) indicators in the model.

The Triple Code Theory full measurement model tested the extent to which *executive function* and *language* contributed to but did not directly predict arithmetic behavioral outcomes across three, format and problem demand specific modules with specific responsibilities for numeric processing. This model also was an approximate good fit for the data; however, patterns of high correlations between factors in this model raised questions about the extent to which direct prediction should be allowed. Specifically, *executive function* and the *auditory verbal* module (responsible for language-formatted problems) correlated near singularity, and the *auditory verbal* module correlated highly with *language* (possibly because *executive function* correlated highly with *language*). Results indicated that *executive function* played a role in facilitating arithmetic outcomes across problem formats and demands but that language-formatted problems were particularly affected by contributions from domains other than a specialized arithmetic module. Despite these issues, the findings from the first phase of testing for the Triple Code model of arithmetic (only) held; the three modules of arithmetic evidenced correlations that indicated they were highly related but distinct.

The Exact Versus Approximate Calculations full measure model tested the extent to which *executive function* and *language* contributed to but did not directly predict arithmetic behavioral outcomes across two, problem demand specific modules with specific responsibilities for numeric processing. This model was not a good fit for the data. Results from phase one of testing indicated that the specifications for the arithmetic (only) and *executive function* portions of the full measurement model were important sources of misfit for the model. However, as in other full measurement models, patterns of factor correlations indicated that *executive function*'s relationships with other factors in the model may have been important sources of model misspecification. This model was ultimately rejected.

Finally, because full measurement model results supported both Encoding Complex Theory and Triple Code Theory, an unplanned, post-hoc model, incorporating key measurement hypotheses of each theory, was examined. This model combined the three-module arithmetic (only) portion of Triple Code Theory with Encoding Complex Theory's specification that *executive function* could be a direct predictor of all arithmetic outcomes and that *language* could be a direct predictor of outcomes on language-formatted arithmetic problems. This model was an approximate good fit for the data, and Chi-square difference testing indicated that this model significantly improved fit as compared to all other full measurement models tested (see Table 25). This model represented a synthesis of hypotheses from two theories of arithmetic cognition that were supported by patterns of results from all model testing, and as such, this model was ultimately retained as the most parsimonious presentation of results.

**Table 26. Summary of Model Testing Results**

Initial Measurement Models	$\chi^2$	df	p	CFI	RMSEA	Note
Abstract Code Arithmetic	705.68	36	<.001	.88	.10	Structurally Identical to Abstract Code Arithmetic
Encoding Complex Arithmetic	705.68	36	<.001	.88	.10	
Triple Code Arithmetic	302.59	33	<.001	.95	.07	
Exact V. Approximate Arithmetic	547.10	35	<.001	.91	.09	Model is just-identified
Language	0.00	0	N/A	1.00	.00	
Executive Functioning	31.57	5	<.001	.97	.06	
Full Measurement Models	$\chi^2$	df	p	CFI	RMSEA	Note
Abstract Code Theory	1386.75	141	<.001	.87	.07	$\Delta\chi^2 (17) = 1045.05, p < .001$
Encoding Complex Theory	478.80	128	<.001	.96	.04	$\Delta\chi^2 (4) = 137.10, p < .001$
Triple Code Theory	592.06	134	<.001	.95	.04	$\Delta\chi^2 (10) = 250.36, p < .001$
Exact V. Approximate Theory	1230.11	138	<.001	.88	.06	$\Delta\chi^2 (14) = 888.41, p < .001$
Post Hoc Hybrid	341.71	124	<.001	.98	.03	Baseline Model for $\chi^2$ Difference Testing

## 5 CHAPTER 5: DISCUSSION

The purpose of this study was to evaluate several leading theories of arithmetic cognition with special attention to possible systematic measurement error associated with item formatting and to possible contributions of cognitive abilities other than a quantitative domain that is specialized for numeric processing. Four leading theories of arithmetic cognition were used to guide hypotheses about (1) the structure of mathematics abilities involved in arithmetic cognition, (2) the role of language versus Arabic numeral symbolic formats in predicting arithmetic outcomes, (3) the role of exact versus approximate calculation demands in predicting arithmetic outcomes, and (4) the possible contributions of language and executive function in predicting arithmetic outcomes.

### 5.1 Summary of Major Findings

#### 5.1.1 *The structure of arithmetic cognition.*

As predicted by Triple Code Theory, the structure of arithmetic cognition was best supported by several modules of quantitative ability with specialization for particular formats and problem demands. An *auditory verbal* module was largely responsible for problems that were language-formatted. A *visual Arabic number form* module was largely responsible for problems that were formatted with Arabic numerals. An *analog magnitude* module was largely responsible for problems that involved estimation across formats. This three-module architecture of arithmetic cognition was valuable for explaining arithmetic outcomes across the models tested in the current study.

#### 5.1.2 *Symbolic formatting and calculation demands.*

Abstract Code Theory's stipulation that *abstract semantic codes* predict arithmetic outcomes across various formats of problem was not supported, nor was a specification of

Encoding Complex Theory in which a unitary, *seemingly modular encoding complex* predicts arithmetic outcomes across formats. Exact Versus Approximate Calculations Theory's specification that exact and approximate problem demands would be predicted by separable cognitive architectures was somewhat supported, but ultimately, among problems with exact calculation demands, different formats were predicted by different modules. Among symbolically formatted problems, language and Arabic numeral formats were distinct but related. Among calculation demands, exact and approximate calculations were distinct but related. However, within exact problems, those problems with language formatting were separable from problems with Arabic numeral formatting.

### 5.1.3 *Contributions from executive function and language.*

Although the unitary, practiced, *seemingly modular encoding complex* model of arithmetic-only was not supported, another important tenet of Encoding Complex Theory was instrumental in predicting arithmetic outcomes. As predicted by Encoding Complex Theory, across all problem formats and calculation demands, *executive function* was a major predictor of arithmetic outcomes. The inclusion of *executive function* as a direct predictor of arithmetic outcomes overwhelmed the arithmetic-only models of cognition. Little variance remained for modules of arithmetic cognition to explain; however, each retained some unique predictive value.

#### 5.1.3.1 *The relationship between executive function and language.*

Interestingly, *executive function* left no predictive value for *language* ability on language-formatted problems. Language-formatted problems were explained mostly by *executive function* and somewhat by the *auditory verbal* module of arithmetic in the current study, and *language* ability evidenced a negative relationship with Arabic numeral formatted problems and estimation

problems. This outcome suggests that *language* ability was not directly contributing to arithmetic cognition. However, the lingering, large correlation between *language* and *executive function* suggests that *language* has some role to play in arithmetic cognition. It raises questions about the possibility that *language* may play a facilitative role in reasoning, particularly for language-formatted problems.

Explaining the relationship between *language* ability and *executive function* in a theoretical model of arithmetic cognition will be a challenge for future research. Given that (1) *language* is not positively associated with modules of arithmetic, (2) nor is *language* a direct predictor of language-formatted arithmetic, but (3) *executive function* is a direct predictor of arithmetic outcomes across modules of cognition, this research suggests that language may play an indirect role in helping executive systems of control to predict arithmetic outcomes.

Several theories of executive function have implicated language ability as a facilitator of systems of executive control. Most often, this relationship has been conceptualized in terms of the construct of internal speech, also called self-directed speech or private speech. As a construct it can be defined as language that is generated and directed internally, not directed socially toward communication partners other than the self, for the purpose of facilitating cognition and behavioral control (see for example, Berk, 1999). In Baddeley's (see for example Baddeley & Logie, 1999; Baddeley, 1992, 2000) model of working memory, internal speech may play a critical role in helping to maintain mental representations of stimuli in the phonological loop of working memory via an articulatory rehearsal system. In Barkley's (1997) model of self-regulation, internal speech helps to regulate inhibitory control by guiding rule-governed behaviors and self-evaluation during problem-solving. Similarly, in Zelazo's (see for example Zelazo & Frye, 1998) model of problem-solving, self-directed, internal speech plays a crucial

role in problem-solving, particularly during planning and inhibition. In this model, self-directed speech helps to link mental representations of problems, rules for problem-solving, and consequences of problem-solving efforts.

Internal speech may have properties that are qualitatively different than socially-directed speech with communication partners, and measuring it may require methodologies that utilize careful behavioral observation and self-reporting during and after the performance of problem-solving tasks (Berk, 1999). Though this was beyond the scope of the current study, future research should investigate the construct of internal speech as an indirect predictor of arithmetic problem-solving.

#### 5.1.3.2 *The relationships between arithmetic modules after accounting for executive function.*

The addition of *executive function* as a direct predictor of arithmetic outcomes also impacted the relationships between the three modules of arithmetic cognition. Although the three modules of Triple Code Theory evidenced a pattern of strong, positive relationships when modeled in isolation, this was no longer true when *executive function* was explicitly modeled. Problems involving exact calculations remained highly related across language formats (on the *auditory verbal* module) and Arabic numeral formats (on the *visual Arabic number form* module); however, the relationships of these modules with the *analog magnitude* module changed when *executive function* was included. With explicit modeling of *executive function* in arithmetic outcomes, the *visual Arabic number form* module was only slightly related to the *analog magnitude* module, and the *auditory verbal* module was no longer related to the analog magnitude module.

These correlations represented Triple Code Theory's specification of transcoding, or direct communication between modules of arithmetic cognition during numeric processing, and it is this notion of transcoding that allows Triple Code Theory's arithmetic modules to avoid necessarily communicating via abstract semantic codes. Though direct communication between Triple Code Theory's modules is assumed during numeric processing, only the *analog magnitude* module is hypothesized to contain semantic information about number. These findings suggest that when the role of *executive function* in arithmetic cognition is directly modeled, transcoding with the *analog magnitude* module may be minimal or non-existent. Perhaps numeric processing for problems involving language-formats, Arabic numeral formats, multi-digit operations, and language-based memory for numbers relies more heavily on executive function (attention, inhibition, working memory, and reasoning) than it does on semantic information about number.

## **5.2 Implications for Measuring Arithmetic**

The findings from the current study raise important questions about the inferences that can confidently be made from testing instruments. The assumption that all assessments which involve arithmetic are inherently measures of arithmetic ability is not warranted. Features of problem formatting and problem demands may influence the extent to which arithmetic is being captured by measurement instruments, and even when measures appear to reliably and validly capture arithmetic skill, they may also be measures of executive systems of control.

When attempting to measure arithmetic cognition, measurement formatting and problem demands are important, but all of the arithmetic outcomes in the current study were largely predicted by domain general capacities in executive control. Despite the overwhelming effect of executive function, several measures of arithmetic did retain unique predictive value that was salient. These measures either involved Arabic numeral formatting and small problem sizes or

estimation problem demands. Such formats and problem demands may be promising methods of assessing arithmetic competence because these types of problems remained strong predictors of arithmetic cognition despite the contributions of executive function.

Conversely, language-formatted arithmetic items may yield results with dubious inferential value for assessing arithmetic cognition. Language-formatted items retained little unique predictive value with an auditory verbal arithmetic module once executive function was added as a direct predictor of arithmetic outcomes, suggesting that language-formatted items may be mostly measures of executive function and, by extension, the role of language ability in facilitating linguistic problem-solving. Thus, language-formatted “arithmetic items,” may more accurately be labeled “linguistic problem-solving tasks that involve some arithmetic”.

### **5.3 Limitations and Future Directions**

#### **5.3.1 *Adapting theories toward specific measurement hypotheses.***

The specificity required by the factor analytic framework is a limitation of the current project. Factor models represent abilities or commonalities between various measures, but they do not represent processes unless a process is specifically being modeled (Carroll, 1993). Such a model would necessitate structural hypotheses among traits, with the specific allowance for traits to influence one another in the time-scale specified by the process (e.g., over seconds, minutes, days, years). Arithmetic cognition is a process. Executive control is arguably a process. Linguistic facilitation of executive control is also a process. Inferences in the current study are limited to traits, but the relationships among traits at a single time point can give important clues about underlying processes, and factor analysis can help to answer important questions about the properties of measurements.

It is important to note that these theories of arithmetic cognition were not specified with factor analysis methodologies in mind, and so, translation into factor analytic frameworks becomes difficult when theories of arithmetic cognition are not explicit in specifying their parameters. For example, “contributions” could be conceptualized as direct predictions of latent factors, correlations between latent factors, or perhaps residual error terms. Some specific aspects of each theory lend themselves to formulations with factor models, while other aspects were not necessarily testable with this method. For example, modeling Abstract Code Theory's highly complex mechanism of numeric processing was beyond the scope of the current study.

In general, theories vary in the extent to which they directly consider measurement of the constructs they specify and in the extent to which their recommendations to users are explicit about methods of capturing those constructs. Theories of arithmetic cognition tend to be somewhat terse in their measurement specifications and methodological recommendations. Ideally, this research may inform theories of arithmetic cognition with regard to both measurement method effects and the possibility of factor modeling as a methodological approach to evaluating theoretical postulations of latent, cognitive domains and problem formatting effects. Measurement hypotheses in the current study were carefully constructed with the aim of striking a balance between faithfully representing theoretical postulates and holding the research to the methodological rigor demanded by factor analysis. Still, the measurement hypotheses for theories of arithmetic cognition are open to other interpretations. Future research should explore alternate measurement hypotheses with these theories of arithmetic cognition.

### **5.3.2      *Adapting theories toward developmental hypotheses.***

The second limitation of the current project lies in the generalization of theory to a population at an earlier developmental stage. Although these theories of arithmetic cognition

largely pertain to the skilled performances of adults, this project aimed to understand the arithmetic cognition of school-aged children and the facets of numeric cognition that may predict their development into skilled adults. The extent to which these theories apply to school-aged children is unclear, but generally speaking, theories that specify one factor structure across all possible populations (regardless of experience and development) have little room for realistic evaluation.

Although some theories of arithmetic cognition make specifications about growth and the ways in which one might become a skilled adult, others do not. Invariance testing (the idea that one can test the hypothesis that the same cognitive architecture that is specified for adults can be assumed for children) is implicit in the current project, because the theoretical specifications of arithmetic cognition pertained to adults but were used to inform hypotheses about children. However, future research should examine the development of arithmetic cognition in children, adolescents, and adults utilizing a longitudinal design and explicit testing of longitudinal measurement invariance. Extant neuroimaging research has indicated that quantitative cognition of children and adolescents may be qualitatively different from that of skilled adults (e.g., Cantlon et al., 2006). Future behavioral research that is explicitly designed to examine the arithmetic cognition of children and adolescents should be (1) guided by the possibility that their cognition may be qualitatively different from the arithmetic cognition of skilled adults as opposed to a deficient form of adult-like cognition, and (2) followed by theoretical extensions of existing theories of arithmetic cognition designed to address the developmental continuum of quantitative cognition.

### 5.3.3 *Generalizability of symbolic formatting.*

A third, major limitation of the current project is that it is exclusively focused on numeric processing with symbolically formatted measures of arithmetic (e.g., language or Arabic numeral formats) and does not include non-symbolically formatted measures of arithmetic (e.g., dot arrays). Although the arithmetic that children will encounter in most formalized assessment settings is symbolically formatted, developmental research on the quantitative domain is focused largely on children's performance with non-symbolically formatted measures. Including non-symbolically formatted measures of arithmetic in measurement batteries will be essential for establishing common scaling and examining developmental continuity in the quantitative domain. Future research should explore arithmetic cognition, formatting effects, and domain specificity (contributions of cognitive abilities other than a quantitative domain) with the inclusion of non-symbolically formatted arithmetic items in the measurement battery.

Similarly, many other aspects of item modality (e.g., timed/untimed, problem size, number of steps required to solve a problem) as well as item content (e.g., arithmetic, algebraic reasoning, geometry) are often controlled or varied in order to approximate item difficulty across various types of mathematics tasks. Implicit in these studies is the idea that items are (a) becoming more difficult as a result of varying certain aspects of their modality (e.g., speededness or number of steps to solution), and (b) items may be becoming more difficult because varying certain aspects of their modality taps cognitive abilities other than the quantitative domain (e.g., processing speed or working memory). The difficulty of items requires empirical examination, as does the assumption that these items may begin to measure cognitive domains other than the quantitative domain. The purpose of the current study was to examine symbolically formatted arithmetic items with regard to theoretical specifications of the cognitive abilities involved in

solving them; however, future research should examine other aspects of item modality and their effect(s) on the measurement of cognitive abilities across a variety of tasks involving differing mathematical content.

#### **5.3.4 *Overlap in features of item modality.***

Although children were instructed to use estimation to solve the double digit estimation problems, and although these items were speeded in order to encourage the use of the most efficient strategy for solution, it should be noted that these problems could have been solved by using the strategy of calculating the exact answer and then rounding. In other words, depending upon the strategies employed by children during numeric processing, the double digit estimation problems may have been solved using a combination exact calculations and approximation. Unfortunately, the strategy usage employed by children during numeric processing was beyond the scope of the current study. It is indeed probable that certain formats may be better suited for eliciting certain problem-solving strategies (e.g., nonsymbolic formats may be better suited to eliciting approximate calculation strategies).

Similarly, the WJ Applied Problems subtest items are language-formatted problems designed to measure children's knowledge of and ability to solve everyday problems (e.g., telling time). These problems served different roles in different models in the current study. They were alternately loaded onto unitary factors (abstract semantic representations or a seemingly modular encoding complex), an exact calculations factor, and an auditory verbal factor. Their treatment as exact calculation items was perhaps the most questionable. Problems on the WJ Applied Problems subtest require children to produce exact answers, but they do not necessarily require children to perform exact calculations. Of the 39 problems designed for examinees who are not above average adults or who are below college-level in education, most require knowledge of

numbers and operations; however, 12 items (approximately 31%) involve the production of exact answers requiring specific, applied knowledge of telling time, recognizing American money, or reading a thermometer. Thus, unfortunately, the WJ Applied Problems subtest represented a mixture of traditional word problems and applied problems. Though this subtest was consistently significant and salient as an indicator in the models tested for the current study, generalizing of the WJ Applied Problems subtest as a test of traditional word problems requiring exact calculations is limited by the extent to which it includes applied problems.

In both the case of the double digit estimation problems and the WJ Applied Problems, issues of item-formatting overlapped with issues of item calculation demands in ways that may have led to model misfit. This caveat is particularly relevant to the exact versus approximate calculations model. This research found some support for a central tenet of exact versus approximate calculations theory; problems requiring the production of exact solutions appeared to be separable from problems requiring the production of approximate solutions. Other features of item modality, in this case symbolic formatting, were also important contributors to the dimensionality of arithmetic measures. However, examination of the possibility that item features may interact to predict examinee responses was beyond the scope of the current study. Future research should examine the relationship between item modality and the measurement of arithmetic cognition with explicit control in the design of item features (e.g., formatting, calculation demands), observation of children's strategy usage during numeric problem-solving, and allowances for the possibility that features of item modality may interact to predict children's responses.

### 5.3.5 *Measuring and modeling executive function.*

For the purposes of the current study, executive function was indicated by a combination of two measures of working memory, one measure of inhibition and attention, and two measures of inductive reasoning or problem-solving. These five measures were combined in an *a priori* specified, latent factor model with the aims of (1) synthesizing important facets of executive function, while (2) explicitly accounting for measurement error. However, it should be noted that across all of the full measurement models and in the executive function-only measurement model, the executive function factor evidenced some problems.

Although this unitary executive function factor displayed good model fit in most ways, patterns of residual variance indicated that much of the complexity of these indicators was not accounted for by a single factor. Recent research has indicated that what is popularly referred to as executive function may in fact be three, distinct, but highly related constructs, (1) updating or working memory, (2) inhibition or controlled attention and response generation, and (3) shifting or cognitive flexibility during problem-solving (Friedman, Miyake, Corley, Young, & DeFries, 2006; Miyake et al., 2000). Because executive function in the current study utilizes measures of each of these facets, the unitary executive function construct included in this study likely represents a hierarchical ‘EF,’ the correlation between each of these facets or ‘EFs’. For the purposes of the current study this ‘EF’ was interpreted as an overall relationship between executive systems of control and arithmetic performances; however, important nuances in the facets of ‘EF’ are not captured here. Future research should explore the extent to which updating, inhibition, and shifting may make unique contributions to arithmetic outcomes.

## 5.4 Summary and Conclusions

Because this study aimed to examine the construct of arithmetic cognition by examining the formatting and dimensionality of arithmetic measures, a factor analytic framework in conjunction with a multi-trait, multi-method approach was appropriate. The factor analytic framework requires explicit statements of hypotheses about model parameters, which can reveal areas of theoretical misspecification, implications of measurement techniques for construct-level inferences, as well as areas of theoretical ambiguity. Though the specificity required by a factor analytic framework can be challenging, this approach is a promising method for evaluation of theories of arithmetic cognition.

Each of the theories examined in the current study were designed to explain the arithmetic cognition of skilled adults. This study sought to understand the arithmetic cognition of developing children who have some formal education and exposure to arithmetic, but are still actively engaged in mathematics education. Describing a developmental continuum that links the arithmetic cognition of developing children to the cognition of skilled adults will be a crucial next step for researchers and theoreticians.

In general, results from this study provided support for both Triple Code Theory and Encoding Complex Theory, and to some extent, Exact Versus Approximate Calculations Theory is also supported. As predicted by Triple Code Theory, arithmetic outcomes with language formatting, Arabic numeral formatting, and estimation demands across formats were related but distinct from one another. This finding is also compatible with Encoding Complex Theory's stipulation that formatting effects exist for arithmetic cognition. The large and enduring relationship between problems that required exact calculations (across formats) also provides

support for Exact Versus Approximate Calculations Theory's stipulation that exact calculation problems may draw from the same cognitive processes.

Executive function was a direct predictor of all arithmetic outcomes. This finding is compatible with Triple Code Theory's stipulation that other cognitive domains, in particular domains responsible for coordinating visuospatial attention, may contribute to arithmetic cognition. The construct of executive function is complex, and modeling that complexity was beyond the scope of the current study; however, the facets of working memory, inhibition and attention, and induction and reasoning ability shared a unitary predictive power in explaining arithmetic.

Given the strong and enduring relationship between executive function and language ability, this synthesized executive control may have been facilitated by language ability in a collaborative relationship that was beyond the scope of the current study. Future research should investigate the extent to which internal or self-directed speech may facilitate executive function and indirectly predict performance on arithmetic problem-solving tasks. This pattern of results may be particularly pertinent for language-formatted arithmetic items. Results from the current study support the growing body of literature indicating that caution should be used in interpreting the results from language-formatted arithmetic items (e.g., Abedi & Lord, 2001; Martiniello, 2009; Rhodes, Branum-Martin, Morris, Ronski, & Sevcik, in press). These items may have little construct validity as pure measures of mathematics ability. Inferences about arithmetic mastery should be made with caution when they are based on results from language-formatted testing instruments, and this caution is particularly relevant to national achievement assessments that utilize language-formatting in their assessment of mathematical competence.

## REFERENCES

- AAIDD. (2010). *Intellectual Disability: Definition, Classification, and Systems of Supports* (11th ed.). Washington D.C.: The American Association on Intellectual and Development Disabilities.
- Abedi, J., Hofstetter, C., Baker, E., & Lord, C. (2001). *NAEP math performance and test accommodations: Interactions with student language background* (Vol. CSE Techni). Los Angeles. Retrieved from <http://www.cse.ucla.edu/products/reports/newTR536.pdf>
- Abedi, J., & Lord, C. (2001). The Language Factor in Mathematics Tests. *Applied Measurement in Education*, 14(3), 219–234. doi:10.1207/S15324818AME1403\_2
- Abedi, J., Lord, C., & Hofstetter, C. (1998). *Impact of Selected Background Variables on Students* (Vol. CSE Techni). Los Angeles.
- Abedi, J., Lord, C., & Plummer, J. (1997). *Language background as a variable in NAEP mathematics performance (CSE Technical Report No. 429)*. Los Angeles, CA.
- Ansari, D., & Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: An event-related functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 18(11), 1820–1828.
- Baddeley, A. (1992). Working memory. *Science*. doi:10.1126/science.1736359
- Baddeley, A. (2000). The episodic buffer : a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. Retrieved from <http://eprints.whiterose.ac.uk/66016/>
- Baddeley, A. (2001). Is working memory still working? *The American Psychologist*, 56(2), 851–864. doi:10.1037/0003-066X.56.11.851
- Baddeley, A., & Hitch, G. J. (1974). Working Memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory Vol. 8* (pp. 47–89). New York, NY: Academic Press.
- Baddeley, A., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 28–61). New York, NY: Cambridge University Press.
- Barkley, R. a. (1997). Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychological Bulletin*, 121(1), 65–94. doi:10.1037/0033-2909.121.1.65

- Benson, D., & Denckla, M. (1969). Verbal paraphasia as a source of calculation disturbance. *Archives of Neurology*, 21, 96–102.
- Berk, L. E. (1999). Children's private speech: An overview of theory and the status of research. In *Lev Vygotsky: Critical assessments: Thought and language* (pp. 33–70).
- Blair, C., Zelazo, P. D., & Greenberg, M. T. (2005). The measurement of executive function in early childhood. *Developmental Neuropsychology*, 28(2), 561–571. doi:10.1207/s15326942dn2802\_1
- Bloom, L., & Lahey, M. (1978). *Language development and language disorders*. New York, NY: John Wiley & Sons, Inc.
- Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Brodinsky, B. (1977). Back to basics: The movement and its meaning. *Phi Delta Kappan*, 58, 522–527.
- Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.
- Campbell, J. I. (1994). Architectures for numerical cognition. *Cognition*, 53, 1–44. doi:10.1016/0010-0277(94)90075-2
- Campbell, J. I., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General*, 117(2), 204–214. doi:10.1037/0096-3445.117.2.204
- Campbell, J. I., & Epp, L. (2005). Architectures for arithmetic. In J. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 347–360). New York, NY: Psychology Press.
- Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. a. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, 4(5), 844–854. doi:10.1371/journal.pbio.0040125
- Carroll, J. B. (1993). *Human Cognitive Abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511571312
- Clark, J., & Campbell, J. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition*, 17(2), 204–239.
- Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The Cognitive and Academic Profiles of Reading and Mathematics Learning Disabilities. *Journal of Learning Disabilities*, 45(March 2011), 79–95. doi:10.1177/0022219410393012

- Cote, J. a., & Buckley, M. R. (1987). Estimating Trait, Method, and Error Variance: Generalizing Across 70 Construct Validation Studies. *Journal of Marketing Research*, 24(August), 315–318. doi:10.2307/3151642
- Cote, J. a., & Buckley, M. R. (1988). Measurement Error and Theory Testing in Consumer Research: An Illustration of the Importance of Construct Validation. *Journal of Consumer Research*, 14(March), 579. doi:10.1086/209137
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44(1), 1–42.
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics* (Revised an.). New York, NY: Oxford University Press.
- Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396.
- Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: A case study of severe acalculia with preserved approximation. *Neuropsychologia*, 29(11), 1045–1074.
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1, 83–120.
- Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, 14, 218–224. doi:10.1016/j.conb.2004.03.008
- Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487–506. doi:10.1080/02643290244000239
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science*, 284(May), 970–974. doi:10.1126/science.284.5416.970
- Eid, M., Lischetzke, T., & Nussbeck, F. W. (2006). Structural equation models for multitrait-multimethod data. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 283–299). Washington, D.C.: American Psychological Association.
- 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. doi:10.1207/S15328007sem0803\_5
- Feigenson, L., Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: Object files versus analog magnitudes. *Science*, 13, 150–156.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., & DeFries, J. C. (2006). Not All Executive Functions Are Related to Intelligence. *Psychological Science*, 17(2), 172–179.

- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., ... Fletcher, J. M. (2006). The Cognitive Correlates of Third-Grade Skill in Arithmetic, Algorithmic Computation, and Arithmetic Word Problems. *Journal of Educational Psychology*, 98(1), 29–43. doi:10.1037/0022-0663.98.1.29
- Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology*, 100, 491–509.
- Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*, 100(1), 30–47. doi:10.1037/0022-0663.100.1.30
- Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). Test of computational fluency. Available from L. S. Fuchs, Department of Special Education, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). Grade 3 Math Battery (Unpublished paper). Available from L. S. Fuchs, Department of Special Education, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLoS ONE*, 8(1). doi:10.1371/journal.pone.0054651
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early Identification and Mathematics Difficulties. *Journal of Learning Disabilities*, 38(4), 293–304. doi:10.1177/00222194050380040801
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197–218.
- Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, 6(2), 0275–0285. doi:10.1371/journal.pbio.0060011
- Jordan, N. C., & Hanich, L. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, 33, 567–578.
- Kelly, D., Xie, H., Nord, C. W., Jenkins, F., Chan, J. Y., & Kastberg, D. (2013). *Performance of U.S. 15-Year-Old Students in Mathematics, Science, and Reading Literacy in an International Context: First Look at PISA 2012*. (NCES 2014-024) (p. 23). Washington

- D.C.: National Center for Education Statistics, U.S. Department of Education. Retrieved from <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2014024>
- Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: The Guilford Press.
- Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: Dissociable systems. *Neuropsychologia*, 41, 1942–1958. doi:10.1016/S0028-3932(03)00123-4
- Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. a., Torgesen, J. K., Wood, F. B., ... Olson, R. (2001). Rethinking learning disabilities. In C. Finn Jr., C. E., Rotherham, R., & Hokanson Jr. (Ed.), *Rethinking special education for a new century* (pp. 259–287). Washington D.C.: Thomas B. Fordham Foundation and Progressive Policy Institute. doi:10.1080/09297049.2011.557651
- Marsh, H. W., Hau, K., & Grayson, D. (2005). Goodness of fit in structural equation models. In *Contemporary psychometrics* (pp. 275–340). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341. doi:10.1207/s15328007sem1103
- Marsh, V., Beard, M., & Bailey, C. (2002). Multitrait-multimethod matrix in scientific inquiry. *Journal of Theory Construction and Testing*, 6(1), 94–97.
- Martiniello, M. (2009). Linguistic Complexity, Schematic Representations, and Differential Item Functioning for English Language Learners in Math Tests. *Educational Assessment*, 14, 160–179. doi:10.1080/10627190903422906
- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*, 4(April), 169. doi:10.3389/fpsyg.2013.00169
- McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, 44(1-2), 107–157.
- McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition*, 4(2), 171–196.
- McCloskey, M., Macaruso, P., & Whetstone, T. (1992). The functional architecture of numerical processing mechanisms: Defending the modular model. In J. Campbell (Ed.), *Advances in psychology (vol. 91): The nature and origins of mathematical skills* (pp. 493–537). Amsterdam: Elsevier.

- McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). Washington, D.C.: American Council on Education.
- Messick, S. (1996). Validity of performance assessment. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1–18). Washington, D.C.: National Center for Education Statistics (NCES 96-802).
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “Frontal Lobe” tasks: a latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. doi:10.1006/cogp.1999.0734
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User’s Guide* (7th editio.). Los Angeles, CA: Muthen & Muthen.
- National Center for Education Statistics. (2013). *The Nation’s Report Card: A First Look: 2013 Mathematics and Reading. (NCES 2014-451)* (pp. 1–12). Washington D.C.: Institute of Education Sciences, U.S. Department of Education. Retrieved from papers3://publication/uuid/45C8F3B4-8A8D-459C-A3C3-2D59E6429AA4
- Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development* (Revised ed.). Austin, TX: Pro-Ed.
- Noël, M. P., Fias, W., & Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition*, 63, 335–374. doi:10.1016/S0010-0277(97)00009-7
- Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children*. London, UK: Psychological Corporation.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, 88(5), 879–903. doi:10.1037/0021-9010.88.5.879
- Rhodes, K. T., Branum-Martin, L., Morris, R. D., Ronski, M., & Sevcik, R. A. (n.d.). Testing math or testing language?: The construct validity of the KeyMath-Revised for children with mild intellectual disability and language difficulties. *American Journal on Intellectual and Developmental Disability*.

- Rivera, S., Reiss, A., Eckert, M., Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, 15, 1779–1790.
- Seethaler, P. M., & Fuchs, L. S. (2006). The Cognitive Correlates of Computational Estimation Skill Among Third-Grade Students. *Learning Disabilities Research and Practice*, 21(4), 233–243. doi:10.1111/j.1540-5826.2006.00220.x
- Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences*, 21(5), 536–542. doi:10.1016/j.lindif.2011.05.002
- Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The Impact of Language Characteristics in Mathematics Test Items on the Performance of English Language Learners and Students With Disabilities. *Educational Assessment*, 11(2), 105–126. doi:10.1207/s15326977ea1102\_2
- Singer, H., & Low, A. (1933). Acalculia (Henschen): A clinical study. *Archives of Neurology and Psychiatry*, 29, 476–498.
- Stanescu-Cosson, R., Pinel, P., van De Moortele, P. F., Le Bihan, D., Cohen, L., & Dehaene, S. (2000). Understanding dissociations in dyscalculia: a brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain : A Journal of Neurology*, 123 ( Pt 1), 2240–2255. doi:10.1093/brain/123.11.2240
- Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, 210, 1033–1035.
- STEM Coalition. (2000). Before it's too late: A report to the Nations from the National Commission on Mathematics and Science teaching in the 21st century. Retrieved from <http://www2.ed.gov/inits/Math/glenn/report.pdf>.
- Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem-solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, 96, 471–479.
- Terry, J. M., Hendrick, R., Evangelou, E., & Smith, R. L. (2010). Variable dialect switching among African American children: Inference about working memory. *Lingua*, 120, 2319–2534.
- Tolar, T. D., Fuchs, L., Cirino, P. T., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2012). Predicting Development of Mathematical Word Problem Solving Across the Intermediate Grades. *Journal of Educational Psychology*, 104(4), 1083–1093. doi:10.1037/a0029020

- Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.
- Whalen, J., McCloskey, M., Lindemann, M., & Bouton, G. (2002). Representing arithmetic table facts in memory: Evidence from acquired impairments. *Cognitive Neuropsychology*, 19(302), 505–522. doi:10.1080/02643290244000086
- Wilkinson, G. S. (1993). *The Wide Range Achievement Tests* (3rd ed.). Wilmington, DE: Jastak Associates.
- Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside.
- Woodward, J. (2004). Mathematics education in the United States: past to present. *Journal of Learning Disabilities*, 37(1), 16–31. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15493464>
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358(6389), 749–750.
- Xu, F., & Spelke, E. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, 1–11.
- Zelazo, P. D., & Frye, D. (1998). Cognitive Complexity and Control: The Development of Executive Function in Childhood. *Psychological Science*, 7(4), 121–126.
- Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., ... Carlson, S. M. (2003). The Development of Executive Function in Early Childhood. *Monographs of the Society for Research in Child Development*, 68(3), 1–151.