A THEORETICAL MODEL FOR THE MEASURE UP PROGRAM:
RELATIONSHIPS AMONG LOGICAL REASONING AND ALGEBRA
PREPAREDNESS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAIʻI AT MĀ NOA IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
EDUCATIONAL PSYCHOLOGY

DECEMBER 2011

By
Linda C. H. Venenciano

Dissertation Committee:
Paul Brandon, Chairperson
Barbara Dougherty
Seongah Im
Lois Yamauchi
Hannah Slovin

Keywords: Measure Up, algebra preparedness, logical reasoning, early algebra, structural equation modeling
ACKNOWLEDGEMENTS

I would like to thank the members of my dissertation committee for mentoring me and for their encouragement and patience as I traveled this journey. I appreciate your wisdom and advice on the research matters, and more importantly, the life matters. I would also like to acknowledge Ronald Heck for generously sharing his expertise as well as his time and energy to assist me in developing the statistical models for my study.

I am grateful to have been influenced by the wonderful folks at the CRDG and ULS. They have inspired me to focus my research on the teaching and learning of mathematics and have continuously supported my efforts in this significant undertaking.

A special acknowledgement to a dear friend and colleague, Fay Zenigami, for assisting me with the assessments, editing my drafts, being a sounding board, and mostly for helping me get through the ever so essential details in work and in life.

Finally I would like to acknowledge my daughters, my husband, and my parents for their patience, support, and the many special reminders about life’s priorities. Thank you for keeping me balanced and grounded in the real world.

I humbly share this achievement with you all.
ABSTRACT

The need to better prepare U.S. children for studying algebra and other high school mathematics is evidenced by studies of international student achievement. An ongoing issue in the field of mathematics education is to develop students’ high-level thinking skills. It is believed that the mile-wide-and-inch-deep approach to curriculum and instruction is responsible for deficiencies in students’ capabilities.

The Measure Up elementary mathematics program was developed from the research and curriculum development conducted in Russia. The mathematics education research from El’konin-Davydov (Davydov, 1966) focused on the development of children’s understanding of general and abstract concepts before teaching specific concepts. This builds from Vygotsky’s (1978) work on the interplay of spontaneous and scientific concepts. The Russian work showed that students were capable of developing higher level thinking and were prepared for studying more advanced mathematics. Prior studies showed that students exiting the Measure Up program were better prepared for studying formal algebra than students who had other elementary mathematics programs.

The goal of this study is to develop a statistical model, using theoretical implications from prior findings, to describe the relationships among Measure Up experience, age, prior achievement, logical reasoning capabilities, and algebra preparedness.

Structural equation modeling revealed a direct and statistically significant impact from the Measure Up mathematics experience on algebra preparedness. Other findings include a statistically significant relationship between logical reasoning capabilities and
algebra preparedness, and a nonsignificant direct relationship between prior achievement and algebra preparedness.

The findings suggest that a mathematics program that develops children’s logical reasoning capabilities also prepares them for a middle school Algebra I course. The findings from this study have implications for educators and curriculum developers who struggle with the trade-off between covering content for arithmetic manipulation skills on one hand and building conceptual understanding on the other.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS................................................................................................ ii

ABSTRACT....................................................................................................................... iii

LIST OF TABLES........................................................................................................... viii

LIST OF FIGURES ........................................................................................................... ix

CHAPTER 1: INTRODUCTION........................................................................................1

  Background of the Problem ......................................................................................... 1
  Statement of the Problem.............................................................................................. 2
  Purpose of the Study ................................................................................................... 5
  Theoretical Framework.................................................................................................. 7
    The Measure Up project........................................................................................... 8
  Research Question ........................................................................................................ 10
  Scope of the Study ........................................................................................................ 10
  Definition of Terms...................................................................................................... 10

CHAPTER 2: REVIEW OF LITERATURE.....................................................................12

  Assessing Mathematics Achievement..................................................................... 12
  School Algebra............................................................................................................. 15
    Preparing for Algebra I ......................................................................................... 18
    Early algebra .......................................................................................................... 19
     Generalized arithmetic ....................................................................................... 20
  The Measure Up Elementary Mathematics Program.............................................. 22
  The Research and Development of Measure Up ....................................................... 29
    Design research....................................................................................................... 30
LIST OF TABLES

Table 3.1 Descriptive Statistics for Exogenous Variables, \( N = 129 \) ..................................56
Table 4.1 Estimated Tetrachoric Correlations for the Logical Reasoning Variables ........76
Table 4.2 Estimated Tetrachoric Correlations for the Algebra Preparedness Variables ...77
Table 4.3 Goodness of Model Fit Statistics, \( N = 129 \) ........................................................79
LIST OF FIGURES

Figure 2.1 Measure Up task for first grade .................................................................23
Figure 3.1 Sample item used to measure logical reasoning .......................................61
Figure 3.2 Sample item used to measure algebra preparedness ..............................64
Figure 3.3 Proposed path model .................................................................................69
Figure 4.1 Path diagram of the final model for this study (Model 8) .................81
CHAPTER 1  
INTRODUCTION  

Background of the Problem  

In 1983 the National Commission on Excellence in Education published “A Nation at Risk: The Imperative for Education Reform.” This triggered reform efforts across the education fields of science, technology, mathematics, and engineering. The National Council of Teachers of Mathematics (NCTM) led the nation in substantial efforts to address the issues in mathematics education. In 1989 NCTM released *Curriculum and Evaluation Standards* to highlight the need for curricula to provide opportunities to develop mathematics applicable in various disciplines to meet society’s changing demand. NCTM’s *Professional Standards for Teaching Mathematics* (1991) followed as a companion to the earlier publication, and the most comprehensive publication in this series, the *Principles and Standards for School Mathematics* was published in 2000. Yet despite various efforts to strengthen U.S. mathematics education and produce mathematically literate high school graduates, dismal reports of academic achievement continued. In 2008 the National Advisory Panel stated in its final report, “International and domestic comparisons show that American students have not been succeeding in the mathematical part of their education at anything like a level expected of an international leader” (p. xii).  

Comparisons of mathematics achievement among industrialized nations have fueled the concerns about U.S. education. The Program for International Student Assessment (PISA) was instituted by the Organization for Economic Cooperation and Development (OECD) in 2000. PISA assesses the performance of 15-year-old students in
reading, mathematics, and science in three-year cycles. Analyzing the data collected from the 2009 administration, Fleischman, Hopstock, Pelczar, and Shelley (2010) found that among the 33 other OECD countries, 17 countries had higher average scores than the U.S. A scale of 1–6 was used to indicate mathematics literacy proficiency with Level 6 being the most advanced. The researchers found 27% of the U.S. students performing at or above Level 4. At Level 4 the students are characterized as being able to “complete higher order tasks” such as solve problems that involve spatial reasoning in unfamiliar contexts. The researchers also found that 23% of U.S. students scored below Level 2. These students are characterized as not being able to consistently “employ basic algorithms,” or make “literal interpretations of the results” of mathematical operations in real world contextual settings. Although the 2009 U.S. average score in mathematics literacy was higher than their 2006 average, it was not measurably different from their 2003 average (Fleischman et al., 2010). It seems that student performance matches mid-level standards and it is not clear that the U.S. reform efforts are creating the necessary impact to address the development of high-level thinking.

In spite of the leadership by NCTM to articulate standards for various components of mathematics education, the findings from studies such as PISA indicate little change in U.S. mathematics achievement. A closer look at mathematics education provides some insights on the pertinent issues stemming from earlier schooling experiences.

**Statement of the Problem**

Algebra I has been commonly referred to as the gatekeeper to the STEM track of higher-level mathematics and sciences. Moses and Cobb (2001) argued that mathematics illiteracy was a barrier to gaining control over one’s own economic and political life and
full participation in society. From this perspective, success in an Algebra I course is an issue of social justice as well as mathematical development. Yet for many students this mathematics course introduces mathematical concepts that are difficult for them to grasp and denies them opportunities to pursue careers in the STEM fields.

Mathematics curriculum in the U.S. elementary grades has historically focused on the learning of arithmetic skills without building the foundation for the study of more advanced topics. This approach has unfortunately resulted in a widespread lack of preparedness for the continued study of mathematics. Kaput (2008) said that the arithmetic-then-algebra curriculum structure has led to “late, abrupt, isolated, and superficial high school algebra courses” (p. 6). Several other researchers have attempted to advance the idea of incorporating algebraic conceptions in the elementary grades and began research initiatives which became known as the early algebra movement. The issues surrounding teaching and learning algebra are profound. The early algebra researchers continue to explore ways to address the need for developmentally appropriate strategies and contexts but the mainstream has not fully embraced this work.

The early grades teachers tend to be reluctant to introduce algebraic concepts. This may be due to teachers’ beliefs that students are not ready for the abstract nature of the discipline, or that algebraic concepts are not considered to be foundational to mathematics. Some teachers attempt to “help” their students by narrowing instructional goals, as it seems to extend from arithmetic, which unfortunately results in an overemphasis on operations. Pedemonte (2009) commented that school algebra is typically seen as a body of rules and procedures for manipulating symbols and that “algebra is taught and learned as a language and emphasis is put on its syntactical
aspects” (p. 249). This approach has led students to expect algebraic procedures to result in an object (e.g., a sum) as they do with arithmetic procedures (Linchevski & Herscovics, 1996). Several studies (e.g., Kieran, 1992; Knuth, Stephens, McNeil, & Alibali, 2006; Sàenz-Ludlow & Walgamuth, 1998; Falkner, Levi, & Carpenter, 1999) highlighted the narrow conception students developed in interpreting the equal symbol (=) to mean find the answer to the problem. Students tended not to consider the alternate meaning of an expression of a relationship. It seems that young students need to begin to develop a broader understanding of symbolic representation in the early grades. This issue was also pursued from an instructional perspective.

A fundamental problem in mathematics education discussed by Smith and Thompson (2008) was the absence of a linkage between numbers and symbols and the ideas that they embodied. They stated that preparing students for the study of algebra should “involve changing elementary and middle school curricula and teaching so that students come to use symbolic notation to represent, communicate, and generalize their reasoning” (p. 96).

Devlin (2009) described the current U.S. approach to K–12 mathematics as an approach that begins with a focus on counting. Counting in the early years focuses on arithmetic, specifically work with natural numbers, integers, and rational numbers. This eventually leads into algebra where mathematics is expanded to all real numbers. He contrasted this approach with an alternate approach, one that focused on measurement. Such an approach introduces students to mathematical concepts by comparing unspecified, continuous quantities. The counting or number approach emphasizes work with numeric operations, but the measurement approach is more robust because it
emphasizes work with abstractions in a meaningful way to build a conceptual system that is relevant beyond arithmetic.

The measurement approach to elementary mathematics is currently being pursued by the Measure Up curriculum research and development project at the University of Hawai‘i’s Curriculum Research & Development Group. The foundations of the Measure Up mathematics curriculum stem from the work of developmental psychologist L. S. Vygotsky (1978) and were carried forward in Russia by a team of psychologists, educators, and mathematics led by V. V. Davydov and D. B. El’konin (1975a, 1975b). The Measure Up project began in 2000 and consists of a Grades 1–5 elementary mathematics program. One of the overarching goals of the program is to develop students’ conceptual understandings to better prepare them for the rigors of higher-level mathematics. The first cohorts of students have exited the Measure Up program and have continued their education in middle school mathematics courses, including Algebra I in Grade 8.

Purpose of the Study

Prior studies (e.g., Tatsuoka, Corter, & Tatsuoka, 2004) indicated that although U.S. students appeared to develop abilities to do algebra tasks, they lacked the high-level thinking skills needed to build deeper understandings of mathematical concepts. It seemed as though students’ understandings were limited to merely extend arithmetic. The Measure Up program addressed students’ needs to develop conceptual understanding of algebra concepts in the elementary grades. The purpose of this study is to construct a theoretical model based on prior findings and hypothesized relationships among Measure
Up experience, preparation for a middle school Algebra I course, and logical reasoning capabilities.

Measure Up experience had been shown to lead to the development of students’ use of algebraic symbols and generalized diagrams (Dougherty & Slovin, 2004) and students’ preparedness for algebra (Slovin & Venenciano, 2008). This study will build on the findings from prior Measure Up research by testing for the same relationship between Measure Up and algebra preparedness and other relationships among the variables.

In developing mathematics curricula, Davydov (2008) studied the development of students’ psychological development and capabilities to think abstractly. Results from his research showed that it was possible for instruction to lead students’ development through consistent experiences using learning activity. Zak (as cited in Davydov, 2008) reported that the students using Davydov’s mathematics curriculum outperformed students in the control group on measures of students’ higher level thinking skills (referred to as theoretical consciousness in the Russian literature). This study will test for a similar relationship between students’ experience with the Measure Up elementary mathematics program and students’ cognitive capabilities.

Furthermore, findings from this study will help to disseminate knowledge gained through the creation of curriculum development. The Measure Up project used a design research method to develop the program. Gravemeijer (1994) pointed out that although this method was a slow, methodical approach to developing curriculum, it generated invaluable insights about teaching, learning, and the program. This study will identify relevant statistical relationships among these components to corroborate with or contrast with insights gained from the research and development.
Theoretical Framework

The theoretical framework for this study is based on implications from the early algebra research to introduce algebraic concepts in the early grades. My perspective is that when children learn mathematics from an approach that emphasizes conceptual understanding and work with abstractions meaningfully, they will be better prepared for algebra. Support for this perspective also comes from the research of Davydov and El’konin (1975a, 1975b).

Researchers Davydov and El’konin (1966) believed that the selection of concepts to begin the study of school mathematics was critical because it laid the foundation for the entire academic subject. Davydov (1966) stated,

Many of the students’ difficulties in mathematics in elementary and high school come about, we believe, first, because what they learn does not correspond to the concepts that actually constitute mathematical structures, and second, because general mathematical concepts are introduced into school courses in the wrong sequence. (p. 56)

Although this statement was made over forty years ago about a different society, the issues being addressed continue to persist in modern U.S. society. The Russian work challenged the popular Piagetian conception of human developmental stages. Davydov (1966) wrote,

But if we assume that the child’s actual mathematical thought develops within the very process which Piaget designates as the formation of operative structures, then these curricula can be introduced much earlier (at seven or eight, for instance), as the child begins to perform concrete operations with a high level of
reversibility. In “natural” conditions, when traditional curricula are being used, it is quite possible that formal operations develop only between the ages of thirteen and fifteen. But cannot their development be “accelerated” through earlier introduction of material which can be learned only by direct analysis of mathematical structures? (p.91)

It is suggested here that mathematics instruction and curricula could be catalysts to advancing the development of students’ cognition. This is a contrast to the more typical approach of postponing the introduction of advanced material until students indicate readiness to learn. Fostering the development of high-level, abstract thinking skills might not only be relevant by Davydov’s conception, but also necessary based on the findings from international mathematics assessment data analyzed by Tatsuoka, Corter, and Tatsuoka (2004). These findings are presented in the following chapter.

The Measure Up project. The Measure Up elementary curriculum research and development project was built upon the research conducted by Russian psychologist El’konin and mathematician Davydov (1966, 1975a, 1975b). These Russian scholars applied Vygotskian theories on learning and development and established El’konin-Davydov schools to pursue research and curriculum development. The theoretical underpinning of their work in mathematics was that instruction should focus on abstract concepts, (i.e., generalizations of mathematical properties) prior to concrete concepts (i.e., computation work with numbers). Measure Up researchers believed this approach would provide the necessary foundation for the study of a middle grades Algebra I course.
The theoretical underpinnings for the Measure Up mathematics program stemmed from Vygotsky’s (1978) work on the development of concepts and work in mathematics education from Davydov and El’konin (Davydov, 1975a, 1975b). The mathematics and curricular paradigm of Measure Up evolved from the Russian foundational theories and the research and development conducted at the Curriculum Research & Development Group (CRDG) at the University of Hawai‘i. The project began in 2000, translating curriculum developed by a team in Russia led by El’konin and Davydov (Davydov, 1975a, 1975b). Translations from Russian to English were accomplished in multiple ways and with the support of collaborative partners, including the Institute for Developmental Pedagogy and Psychology in Russia and the Best Practices in Education foundation in New York. The translations were interpreted by CRDG project staff and a Russian language consultant who, coincidentally, was educated in a Davydov school for part of her early schooling. These interpretations were then sent to the CRDG team to design lesson plans through a process called design research. This process is described in the following chapter.

The Measure Up program was developed from the El’konin-Davydov (Davydov, Gorbov, Mukulina, Savelyeva, & Tabachnikova, 1999) mathematics and emphasized the development of behaviors for learning mathematics. Measure Up students use various representations and terminology to explain their thinking both verbally and in writing. Students are encouraged to seek appropriate multiple solutions and multiple methods. The instruction is focused on developing productive reasoning skills and students’ abilities to make sense of the mathematics. Measure Up students are expected to justify their responses by engaging in problem-solving tasks and sharing their thinking with
others. From a Vygotskian perspective this approach, developing concepts by collaborating with others or intersubjectivity, is a fundamental aspect of the cultural-historical theory. Intersubjectivity provides the opportunity for cognitive growth and development through appropriation. Rogoff (1993) described appropriation as how individuals changed through their involvement in an activity and became prepared for subsequent involvement in other, related activities. From this perspective it is believed that students who engage in learning mathematics through intersubjectivity will appropriate and be prepared for learning advanced topics.

**Research Question**

My research question is, What statistical model best describes the relationships among Measure Up experience, age, prior achievement in mathematics, and logical reasoning capabilities in predicting algebra preparedness?

**Scope of the Study**

This is a study about the effects of an elementary grades program on cognitive development in mathematics and preparedness for Algebra I. Findings from this study will contribute to the early algebra literature. This study will also contribute to the literature that furthers the works of Vygotsky and Davydov.

**Definition of Terms**

For this study, use of Algebra I refers to a specific set of topics, typically included in a first year algebra course, that lead to topics in advanced courses of high school mathematics. The Algebra I topics include variables, equations, the linear function, coordinate graphing, polynomials, and solving simultaneous equations. My references to middle grades Algebra I refer to a mathematics course that has the same content as the
high school counterpart but uses an approach that is developmentally appropriate for students in Grades 6–8.

*Algebra preparedness* refers to the capabilities, as identified by Küchemann (1981), that students use to interpret the meaning of a variable in flexible ways, selecting the appropriate interpretation based on the context. This is a critical understanding for students to be able to work with abstractions, symbolic representations, and generalizations of mathematical relationships between quantities.

Defining *logical reasoning* is a more complicated issue. Although logical reasoning is widely recognized as a cognitive process, researchers have used the conception of logical reasoning very broadly. This presents some ambiguity on how to precisely define and measure it. For this study, I am using the characterization established in the Tatsuoka et al. (2004) study. Logical reasoning capabilities are characterized by the ability to solve a mathematical problem in the following ways: reasoning from cause to effect, analyzing the problem by cases (e.g., when \( x > 0 \), \( x < 0 \), or \( x = 0 \)), using deductive thinking, using if-then reasoning, and generalizing.

My reference to *curriculum* refers to the course of study and accompanying materials used by teachers and students. This includes problem-solving tasks or exercises that students are instructed to do and the directions written for teachers to use as guides in conducting the lessons. Curriculum does not include a specific pedagogical style or instructional delivery.
CHAPTER 2

REVIEW OF LITERATURE

The Measure Up elementary mathematics program was developed from the traditions of Vygotsky’s theory on learning and development and Davydov’s work in developmental and deductive education. Prior studies from the Measure Up project showed that experience in Measure Up led to advancing the development of children’s higher-level thinking and stronger preparation for learning algebra. This study builds on prior findings and contributes to the literature in elementary mathematics education and educational psychology as an application of Vygotsky’s work.

The literature used to develop this study covers topics from mathematics education and educational psychology. Pertinent topics from mathematics education include mathematics achievement in the U.S., research from the early algebra movement, and the mathematics and pedagogy from the El’konin-Davydov work and the Measure Up project. Pertinent topics from educational psychology include theoretical underpinnings from Vygotsky’s work and methods for defining and measuring achievement and cognitive capabilities. The review of these topics provides insights to test for relationships among experience in the Measure Up program, demographic background variables, logical reasoning capabilities, and preparedness for a middle school Algebra I course.

Assessing Mathematics Achievement

Tatsuoka et al. (2004) conducted a rigorous study that presented convincing evidence about weaknesses in U.S. mathematics education. The study used data from the 1999 administration of the Trends in International Mathematics and Science Study.
TIMSS is regarded as the largest and most ambitious international study of student achievement. Since the first assessments in 1995 data collection has continued every four years, with the most recent administration in 2011. In 1999 there were 38 countries participating in the study. TIMSS provides researchers with valuable data on fourth and eighth grade mathematics and science achievement. The data have been used in multiple ways to identify trends within one country and across several, and in one year’s test administration as well as over multiple administrations.

Tatsuoka et al. (2004) examined the eighth grade TIMSS (1999) data from a sample of 20 countries, \( N = 51,435 \) students. The countries in the study included the six top-achieving countries and several others to form a sample diverse in culture and achievement. The researchers examined students’ responses to identify states of mathematical knowledge and associated achievement. They analyzed 163 assessment items and characterized them as attributes of mathematical knowledge, skill, or process. Next they identified response patterns to determine the mastery or nonmastery of an attribute. These findings then served as the basis for more advanced analyses using the Rule-Space Method. This method was developed to identify underlying knowledge and cognitive processing skills used in answering assessment items. The researchers examined single attributes and found that the top five countries, Singapore, Korea, Hong Kong, Japan, and Belgium-Flemish, had also performed exceptionally well on certain tasks. Students’ performance had indicated mastery of cognitive processing attributes (e.g., deductive thinking, analytical thinking, and inductive thinking).

The researchers then conducted a principal component analysis to identify correlations between the single attributes. This analysis resulted in defining four
dimensions, one of which included three process skills (logical reasoning skills, judgmental applications of knowledge and concepts, and executive control, manage data and procedures), a content skill (geometry), and a skill attribute (proportional reasoning). Consistent with observations from their preliminary analyses, the researchers noted that U.S. students were relatively weak on all the attributes in this dimension.

Tatsuoka et al. (2004) formulated three composite variables (process, spatial, reading) using a coding scheme derived from an analysis of written student protocols, responses from domain experts, and feedback from teachers. They found that the summary scores used for the process variable carried much of the variance of the original total mean item scores of the test. That is, ordering countries by the process variable resulted in the same ordering attained by the mean standardized item proportion correct score. These findings showed that advanced cognitive processes correlated with higher overall achievement.

The principal component analysis was used to divide the once broadly defined process variable into three sub-process variables. From this analysis researchers found that the greatest weakness of the U.S. performance was with the variable that included logical and abstract reasoning skills. An examination of U.S. students’ performance on individual attributes showed that their three strongest attributes were quantitative and logical reading, approximation and estimation, and algebra. Their weakest attributes were geometry and logical reasoning. Tatsuoka et al. (2004) suggested that the U.S. emphasis of algebra in the school curricula did not result in high-level mathematics thinking skills, and that the opportunities for students to develop challenging mathematical thinking skills in American schools seemed to be insufficient.
The Tatsuoka et al. (2004) study is pertinent to my research for two reasons. First, because it was a rigorous, large-scale, international study that included diverse educational systems, it is likely that patterns identified in the data are generalizable. Second, the researchers approached data analyses by examining features of both mathematical content and processes. These findings provide interesting insights for further research in mathematics education.

Implications from this study called for the need to address the U.S. teaching and learning of algebra in ways that also contributed to the development of students’ high-level thinking skills. As evidenced by the achievement of the top countries, these objectives were not at odds with each other.

**School Algebra**

Challenges in the U.S. teaching and learning of algebra were articulated in the 2008 report of the National Mathematics Advisory Panel. This report, commissioned by the U.S. Department of Education, concluded that students in U.S. middle or high school algebra classes (a) showed a severe lack of adequate preparation for algebra, (b) lacked a firm understanding of many basic principles of arithmetic, (c) had difficulty grasping the syntax of algebraic expressions, and (d) did not understand the procedures for transforming equations. Furthermore, the report stressed that algebra teachers should not assume that all students understood basic concepts. Based on the emphasis of these remarks, the type of algebra addressed by this report is presumably a formal middle or high school Algebra I course.

Asquith, Stephens, Knuth, and Alibali (2007) examined middle school teachers’ knowledge of students’ understanding of algebraic concepts. The researchers found that
algebraic concepts (i.e., equality and variable representation) were commonly
misunderstood by students, and furthermore, teachers often neglected to recognize the
misconceptions. Interview data for the study had been collected from 20 middle school
teachers. Teachers’ responses were examined in light of student assessment data collected
from sixth- through eighth-grade students from the same district as the teachers.

Asquith et al.’s (2007) findings showed that teachers overestimated the number of
students who understood the concept of the equal sign as a relational symbol. Most of the
students were only able to interpret the equal sign as a symbol to give the answer and
could not provide the alternate meaning of a symbol that represents a relationship of
equality. This study also found that teachers tended not to consider how a narrow view of
the equal sign as an operator would hinder student performance. Such missed
opportunities for teachers to broaden students’ understanding of one of the most basic
and fundamental symbols could have potentially harmful effects on students’ progress in
a middle or high school algebra course.

One of the persistent challenges of learning algebra is making sense of letters
(Malisani & Spagnolo, 2009). Letters are used as variables and have a variety of
conceptions. Variables are versatile and can be used to represent a generalized number,
an unknown, or can be used as an arbitrary symbol. Unfortunately, early impressions
about variables may not assist students in developing a sufficiently general concept about
variable (Wagner & Parker, 1993). As Wagner and Parker (1993) have observed, students
could work with variables without fully understanding the multifaceted aspects of the
literal symbols. This was partly attributed to the way variables could be operated on, as
numbers in arithmetic. Furthermore, instruction often unwittingly created problems with
denotation as the same symbols were used for indicating different conceptions of variable (Malisani & Spagnolo, 2009). It follows that preparedness for Algebra I coursework can be attributed to a student’s ability to make sense of letters used in mathematics. Further discussion of this idea follows below in a study about the effects of the Measure Up program.

Observations from Davis’ (1985) study on algebra students’ behavior indicated a focus on mastering rituals for symbols written on paper based on following rules and memorizing. Davis (1985) attributed this to an overemphasis on teaching manipulation techniques, a perspective that resonates with the Advisory Panel’s report (2008) as well as findings from the Tatsuoka et al. study (2004). Rituals for following rules and memorizing hindered the development of high-level thinking, as students were not required to have a strong understanding of the mathematics.

Rule-following behaviors are similar to what is commonly observed of arithmetic students. Arithmetic students are often taught to apply computation rules without understanding the underlying concepts. Thus, if mathematical behaviors in the middle and high school years are indeed extensions of the behaviors from the elementary years, it seems prudent to focus on developing early behaviors that are consistent with high-level algebraic thinking. Further discussion of this idea follows below in the section on early algebra.

Findings in the 2008 Advisory Panel’s report called for stronger preparation for Algebra I coursework. In the elementary grades this could be accomplished by developing understanding of arithmetic principles and algebraic syntax (e.g., use of variable) with a particular focus on the concept of equality and multiple interpretations of
the equal sign. Elementary mathematics could also be developed from an understanding of measurement principles. The work of the early algebra movement has targeted these concerns by designing interventions and programs to build the foundational skills needed for success in a middle or high school Algebra I course.

**Preparing for Algebra I.** Dekker and Dolk (2011) advocated for developing algebraic thinking in the elementary school curriculum to prepare students for algebra in high school. The researchers analyzed the Dutch mathematics textbooks and identified a long-term learning trajectory from arithmetic in the elementary school to arithmetic in high school. They did not find a parallel development for algebra in the Dutch curricula.

The researchers (Dekker & Dolk, 2011) conducted a study with 10- and 11-year olds in a Dutch elementary school. They posed a multi-part problem that required students to use symbols, generalize a pattern, organize data, propose a formula, and analyze the property of odd numbers (i.e., odd + odd = even). The problem was selected from the algebra unit Patterns and Symbols from the *Mathematics in Context* text because it presented a clear, continuous trajectory from informal to pre-formal solution strategies and it was unlike problems the Dutch students had done in the past. The trajectory was used to compare and track students’ development for algebra.

Results from the study showed that some students were still reasoning at informal levels because they did not refer back to the mathematical property (of odd numbers). Other students were at the pre-formal level, where they were beginning to write formulas. Dekker and Dolk (2011) concluded that the elementary students were developmentally capable of advancing, and had the Dutch mathematics program consistently developed students’ mathematical capabilities along this trajectory, more 15-year-old students
would have been successful on a similar patterning task assessed on the 2004 PISA. Since 25% of the Dutch students skipped the problem or gave the wrong answer on the PISA algebraic patterning task, the researchers concluded that a discontinuity existed in what was expected from students throughout the grades.

Dekker and Dolk (2011) suggested elementary educators pursue opportunities to develop their students’ abilities to reason and generalize, as these are considered to be important aspects of algebraic thinking. The researchers confirmed the actualization of a longitudinal learning line for arithmetic and hypothesized that a continuous line of algebraic thinking from elementary to high school would support student development measured at the high school level. They also discussed the issues of differing expectations and missed opportunities for teachers to support student development of algebraic thinking during the elementary school years. Parallel discussions in the U.S. have fueled the early algebra movement in mathematics education.

**Early algebra.** Early algebra researchers believed that elementary mathematics topics (namely arithmetic) and algebra were not discrete topics. They pursued research under the premise that understanding basic algebraic principles was necessary for a deep understanding of arithmetic. Contrary to what was commonly practiced in mathematics education, the researchers contended that the mastery of arithmetic should not be a prerequisite for the introduction of algebraic concepts.

Kaput (2008) described the research associated with early algebra within three major strands. The strands differed in perspective but shared two core aspects of algebra. One aspect was the incorporation of a method to systematically symbolize generalizations of regularities and constraints. The second core aspect of algebra was the
use of syntactically guided reasoning and actions on generalizations expressed in conventional symbol systems.

From the generalized arithmetic perspective, algebra emerged from the operations, properties, and structures introduced in arithmetic. The researchers within this strand pursued topics such as building generalizations about patterns from number examples. The second strand took a function approach to algebra. From this perspective, the activity of expressing a generalization was likened to describing systematic variations of problems. For example, the practice of beginning with patterning activities in the elementary years leads to generalizations used in later mathematics. The third major strand presented modeling as an algebraic activity. In this last strand were three different approaches, including (a) a variable used to take the place of an unknown, (b) modeling a situation to express a pattern or regularity of a situation, and (c) generalizing from solution to single-answer modeling situations (Kaput, 2008).

**Generalized arithmetic.** One area of research pursued by generalized arithmetic researchers was how mathematics instruction integrated algebraic concepts. Carraher, Schliemann, and Schwartz (2008) clarified the motivation of the work in early algebra as not making Algebra I coursework accessible to younger students, but rather building upon children’s contextual experiences with arithmetic. Another area researchers focused on was the nature of the mathematics tasks. Generalized arithmetic was not merely using letters to represent numbers or blanks in the fill-in-the-blank tasks. Some researchers (e.g., Carraher, Martinez, & Schliemann, 2008) pursued this perspective as a generalized arithmetic of numbers and quantities. Mason (2008) proposed the following definition for generalized arithmetic:
[Generalized arithmetic was] the result of learners generalizing their experience with numbers, and expressing generality about properties of numbers arising from a variety of situations in the material, mental, and symbolic worlds, culminating in expressing generality about the rules of arithmetic and then taking these to be the rules of algebraic manipulation. (p. 77)

The generalized arithmetic approach required a more integrated perspective of mathematics. Researchers from this perspective developed and tested the instructional approaches and associated curricula materials. The NSF-funded project, *Impact of Early Algebra on Later Algebra Learning*, explored the long-term impact of third and fifth graders who attended algebra camp. Researchers Brizuela and Schliemann (2009) designed an algebra summer camp to better prepare students for working with algebra notation and solving equations. The students’ mathematical progress was monitored as they entered the later grades. Preliminary results showed that at Grade 6, students from the experimental group significantly outperformed students in the control group. Furthermore, this finding was more pronounced for items assessing algebraic notation and equations.

Similar but nonsignificant findings were found for the students in Grades 7 and 8 (Brizuela & Schliemann, 2009). This study showed that different instruction and curriculum, in the context of summer camps, yielded different results on students’ algebraic skills. These results support the conception of the sort of long-term learning trajectory in algebra as discussed by Dekker and Dolk (2011). That is, early learning of mathematics that is focused on working with algebra notation and solving equations should lead to stronger capabilities later in a child’s algebra learning trajectory.
Another approach, as characterized by Kaput (2008), within the generalized arithmetic strand was pursued by the Measure Up project. Kieran (2006) described this approach within the framework of the early algebra research. The learning activities of the Measure Up program emphasized generalizing and expressing the generalizations explicitly and systematically. The activities used symbolizing to represent relationships among mathematical quantities. The foundation for this approach stemmed from Davydov’s work (1975a, 1975b).

Students in the Measure Up program learn to express generalizations explicitly and systematically. The marked difference between the Measure Up approach and others in the generalized arithmetic strand is that the introduction of mathematics is constructed by reasoning from a measurement perspective rather than in counting and arithmetic.

**The Measure Up Elementary Mathematics Program**

The Measure Up Grades 1–5 program used a measurement context to develop critical mathematical topics. In the context of this program, measurement included aspects such as (a) comparing something with an object of a known size; (b) estimating or assessing the extent, quality, value, or effect of something; and (c) judging something by comparing it with a standard (Dougherty & Venenciano, 2007). Students explored mathematical structures and developed understanding about quantitative relationships through this perspective of measurement.

The Measure Up program presented an unusual and, by some accounts, controversial approach to counting. In Grade 2 students used different base systems to represent numbers in a measurement context (Mulligan & Vergnaud, 2006). The concept of place value was developed when students found that the base they were working in
necessitated regrouping. This approach enabled students to engage in rich mathematical discussions about topics such as why “10” is a number represented with two digits.

Unlike traditional mathematics programs that begin with the introduction of number and counting, the Measure Up program begins with having students make direct comparisons between continuous quantities (length, area, mass, and volume). Students in Grade 1 develop an understanding of the relationship between units and quantities. Through experiences with physical quantities, students see that smaller units need to be used more times than larger units to measure a given quantity. Students represent this inverse relationship in abstract and algebraic-looking forms (Dougherty, 2008).

On a Measure Up task designed for first graders, Dougherty (2008) described how students were able to make an inference when presented with task in Figure 2.1.

If a quantity $B$ is measured by unit $E$, it equals 5. And when the same quantity $B$ is measured by unit $Y$, it is equal to 8. That is, $\frac{B}{E} = 5$ and $\frac{B}{Y} = 8$. How do the units compare?

*Figure 2.1. Measure Up task for first graders in presented in “Measure Up: A quantitative view of early algebra,” by B. Dougherty (2008), in J. J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp.389–412).*

The students concluded that $Y < E$ and explained that $Y$ must be a smaller unit because it was used more times than unit $E$ to measure an equal quantity.
On a related task that built on this understanding, Measure Up students were presented with a situation where two quantities were measured with the same unit. “If $\frac{W}{E} = 10$ and $\frac{P}{E} = 7$, ” students were able to conclude, “then $W > P$ because it took more unit Es to measure quantity $W$” (Dougherty, 2008, p. 397). Students’ abilities to analyze the equations and justify their responses were attributed to their understanding of unit as it had given meaning to the values of numbers.

An inherent quality of the Measure Up program is for students to use symbols, letters, and diagrammatical representations on a regular basis. The measurement context promoted this understanding of symbolic representations. Furthermore, the Measure Up program provided consistent opportunities for students to use generalizations and make sense about mathematics and structures of the representations.

Critical themes for the other grades included conceptualizing place value and operations with multi-digit numbers in Grade 2; multiplication, division, and mathematical properties in Grade 3; and transformational geometry, geometric measurement, and rational numbers in Grade 4 (Chen, 2006). In Grade 5 the Measure Up students continued to develop concepts introduced in earlier years at more analytical and sophisticated levels. For example students continued to develop the concept of fractions through explorations of theoretical probability with an area model (Dougherty & Venenciano, 2007). Although geometry and fractions were not part of the original Davydov research, they were developed for the Measure Up program in order in order to accommodate U.S. standards of mathematics education.
One of the underlying premises was that this approach would support conceptual understanding and sense making to better prepare all students for an Algebra I course. As with the Russian work, Measure Up adopted the Vygotskian learning-leads-development theory. Measure Up researchers found that children could be brought up through mathematics content more rapidly and with deeper understanding of the content than their non-Measure Up peers (Slovin & Venenciano, 2008). This finding relied on students’ communication abilities to justify how they reasoned about a solution as well as their performance on an assessment of algebra preparedness. Researchers also found that Measure Up students were capable of using algebraic symbols and generalized diagrams to solve problems (Dougherty & Slovin, 2004). Furthermore, Measure Up researchers suggested that when children developed their own understanding of and applications for the diagrams using the Measure Up approach, their cognitive development was positively impacted.

In their 2004 study, Dougherty and Slovin found that students in their third year of the Measure Up program (8–9 years) were capable of using algebraic symbols and generalized diagrams to solve problems. The researchers conducted student interviews and linked responses to Sfard’s (1991, 1995) classification of response types to problem structure. This classification progressed over three stages. It began with student abilities to attend to problem-specific characteristics (interiorization), and then progressed to a transition stage where students used a more general approach (condensation) before evolving to a final stage where multiple forms of representation were used as means for generalizing the structure of the problem (reification).
The students in the Dougherty and Slovin (2004) study were selected from diverse backgrounds in terms of social-economic factors and prior achievement. The researchers intended to recruit a sample that would generate a wide range of responses from the student interviews. Dougherty and Slovin (2004) concluded that regardless of students’ ability to attend to a problem structure in their solution, students’ mathematical development seemed to be positively impacted by Measure Up experiences. This was shown by the children’s ability to effectively problem solve with concurrent models; (a) a physical model (e.g., two amounts of liquid), (b) an intermediate model (e.g., paper strips or a line segment diagram to represent the comparison of the quantities), and (c) symbolization (e.g., algebraic-looking statements of equality or inequality).

A comparative study about the effects of the Measure Up program indicated that more experience in the Measure Up program led to stronger preparation for algebra (Slovin & Venenciano, 2008). The researchers used the Chelsea Diagnostic Mathematics Test: Algebra (CDMTA) (Hart, Brown, Kerslake, Küchemann, & Ruddock, 1985) to measure student readiness for an Algebra I course. The CDMTA designated Levels 1–4 to characterize a student’s understanding of variable, with the higher numbers indicating more sophisticated understandings.

Other variables used in the Slovin and Venenciano (2008) study were the number of years in the Measure Up program, prior verbal and mathematics achievements, as measured by the Stanford Achievement Test v. 9, grade level, socio-economic status as coded by the school’s admission procedures, and gender. The sample consisted of all fifth and sixth graders from the same school. There were 60 students in the sample; 19 students had Measure Up experience and 41 had no Measure Up experience. The intent
of the study was to determine if the children who were in the Measure Up program were better prepared for success in a middle school Algebra I course.

Descriptive statistics suggested that although the students with no Measure Up experience had a higher mean score on the prior achievement measure, the Measure Up students had a higher mean score on the algebra preparedness measure. A multiple regression model-building technique was used to analyze the data. The maximum $R^2$ improvement technique selected the variables with the most explanatory power, progressively creating models with more variables. Slovin and Venenciano (2008) found that Measure Up experience was the variable with the most explanatory power of student performance on the algebra preparedness assessment. That is, Measure Up was the largest statistically significant contributor to explaining student achievement on the CDMTA assessment.

These results were followed with a more careful analysis of the responses from the Measure Up students. The CDMTA levels (Hart, 1981) indicated degree of problem difficulty and students’ capabilities for interpreting the letter as presented in the task. At Levels 1 and 2 students were considered to be capable of using letters with assigned values or as objects and evaluating the letter. At the more advanced levels, Levels 3 and 4, letters represented a set of unspecified values. To solve a problem at these levels required a student to accept a more abstract context, where the interpretation of the letters required the understanding of variable representations.

Measure Up students performed disproportionately better than non-Measure Up students on some of the tasks at Level 2 or higher, regardless of grade level. The researchers first identified these items and then reviewed the items to identify tasks that
would also serve as interview prompts to probe student insights. The process produced a subset of tasks for follow-up interviews to uncover student thinking beyond the written solutions on the assessment tasks. A stratified random sample was used to select Measure Up students for interviews. Three students from Grade 5 and three from Grade 6 were selected to represent low, mid, and high achievement.

During the interview the students were given copies of their previously written work and asked to say if they agreed with their original response to the problem and to explain their thinking. The researchers then analyzed these results and identified common threads in the responses. Students justified their responses with reference to the letters as representations of units, the concept of unit, or for a quantity. The researchers concluded that the evidence from this study corroborated the project’s qualitative data, which suggested that Measure Up students developed conceptual understanding of algebraic concepts (i.e., symbolic representations) and the abilities to reason and justify their answers.

Studies about the effects from the Measure Up program indicated that students were able to recognize, and use with understanding, mathematical structures in problem solving. These mathematical practices are more sophisticated than the learned behaviors of mastering rituals and following steps as indicated by Davis (1985). Furthermore, a student’s ability to look for and make use of structure was described as one of eight standards for mathematical practice in the Common Core State Standards (2010). It appears that the Measure Up program has the potential to prepare students for the rigors of a middle grades Algebra I course while addressing the broader standards of expected mathematical behavior.
The Research and Development of Measure Up

There are two distinct but interdependent Measure Up project objectives relevant to my study. One objective is to design a curriculum that develops children’s algebraic thinking, conceptualization, and skills in Grades 1–5 (Chen, 2006). The second objective is to research the effects of the curriculum on students’ mathematical development. The curriculum provides the context for the research and the insights gained from the research are applied back into the curriculum development.

Measure Up project members established a pedagogical approach that was consistent with their prior work on other mathematics curricula projects at the CRDG (e.g., Process Algebra, Geometry Learning Project, the Reshaping Mathematics Project). Russian researchers, such has Krutetskii (1976), influenced the CRDG researchers, as the designs for the mathematics tasks were similarly structured to develop skills for problem solving. Although the Measure Up mathematics program was generated in the El’konin-Davydov (Davydov, 1975a, 1975b) work, some adjustments were necessary to accommodate sociocultural and historical contextual differences. Sophian (2007) indicated that it was unclear if an implementation of the Russian mathematics program in the U.S. was desirable or possible. She advocated for further studies with the El’konin-Davydov curriculum, particularly the aspects of introducing numbers in a measurement context, making comparisons between quantitative relations, and engaging students in dialogue to construct mathematical generalizations, explanations, and conjectures, and with other students. Sophian (2007) also suggested that the Measure Up project pursue a systematic evaluation of the impact on student learning as there were promising
indications of the program’s effects. My study will address this suggestion by examining the effects of the Measure Up program on students’ mathematical development.

**Design research.** To guide the adaptations of the El’konin-Davydov mathematics curriculum (1999), the Measure Up project used a research approach that supported the data collection and analysis in the complex classroom environment. Design research (Shavelson, Phillips, Towne, & Feuer, 2003; Gravemeijer & Cobb, 2006) was selected for this project because it helped support the research and development process and was successfully used to produce other curricula materials at CRDG, the site for the Measure Up project. Design research was an appropriate choice for the development of the program because it allowed for the study of the mathematics curriculum and instruction in the context of an authentic setting. Furthermore, this approach generated the necessary insights for researchers to develop and test theories about the learning process and longitudinal mathematical development. Gravemeijer (1994) described design research as a holistic framework for curriculum development. The Measure Up project integrated design and research in a process that was guided by theory but also produced theory.

Initial implementation of the Measure Up curriculum was conducted at two sites, both public charter schools in Hawai‘i. Dougherty and Slovin (2004) explained that the criteria for selecting sites was to form a diverse group of students to reflect the state’s public school population in terms of prior achievement, socio-economic background, and ethnicity. The schools also had relatively stable populations and provided researchers with the opportunity for longitudinal studies.

The project team consisted of mathematics educators and researchers who worked closely with the teachers and students. During early stages of the Measure Up project,
project researchers led the teaching activities in order to test materials and develop instructional practices. Students in the classes were accustomed to being observed, videotaped, and interviewed by the researchers. Students developed the habits of explaining their mathematical thinking and justifying their answers (Dougherty & Slovin, 2004). The goals for experimentation in design research are to explore, probe, and investigate (Gravemeijer & van Eerde, 2009). The Measure Up project collected several formats of data (e.g., video, scripted observations, notes on recording instruments) and used them to link students’ thinking with mathematical tasks and instruction. Several iterations of the research and development process were needed to refine the curriculum.

The Measure Up mathematics program resulted in a variation of the Davydov elementary mathematics curriculum. The design research approach led curriculum researchers and developers through a process in which decision-making was based on the findings from a holistic environment. Although much of the theoretical underpinnings had been maintained, some variations in the curriculum and instructional delivery were natural consequences of the design research process.

Davydov Curriculum in the U.S.

The Davydov curriculum has a history in the U.S. Schmittau (1992) is believed to have led the first application of the Davydov mathematics curriculum in a U.S. school setting. This was a direct implementation of the Grades 1–3 Davydov mathematics curriculum translated from Russian, without variations such as those produced through design research. The Davydov curriculum consisted of carefully developed sequences of problems for students to solve and was unlike the typical U.S. mathematics curriculum on two substantial levels. On an implementation level, the curriculum lacked a teacher’s
manual, an expected part of a U.S. mathematics program. The curriculum was written to
develop children’s cognition by presenting challenging tasks that required students to
construct new methods in order to solve them. The Davydov instructional notes described
mathematical concepts and experiment-like demonstrations for teachers to present to
students and guide the formulation of mathematical observations. This was purposeful
and consistent with the theoretical perspective from Vygotsky’s work on the zone of
proximal development. Effective instruction of the Davydov program required that a
teacher had a profound understanding of the mathematics. Teachers were expected to be
able to anticipate the best approach to developing a concept. Unlike U.S. teachers’
manuals, the teachers of the Davydov curriculum did not have detailed and prescriptive
guides to follow. Davydov curriculum teachers had opportunities to reformulate their
own understanding of the mathematics concepts.

Teachers used the Davydov instructional activities to facilitate students’
mathematical development. Problems were presented to the students, and teachers
deliberately withheld judgment about students’ solutions. The impact of instructional
pedagogy was a critical aspect of the learning environment (Schmittau & Morris, 2004).
Students were assisted in arriving at the mathematically appropriate conclusions by
justifying their thinking rather than relying on a higher authority as is typical of U.S.
mathematics instruction. This process developed children’s abilities to analyze problem
situations at a theoretical rather than an empirical level. Furthermore, children were
expected to use multiple methods for problem solving and to attend to the conditions of
the problems in order to use the most advantageous method.
On a content level, the Davydov curriculum was different from typical U.S. mathematics programs in the sequencing of topics. The overarching goal was to develop the understanding of the mathematics concepts. Where the traditional U.S. approach focused on counting discrete quantities, the Davydov curriculum focused on comparing measurable and continuous mathematical quantities. From Davydov’s perspective, the genetic development of the concept of number began with working with objects (continuous quantities). Schmittau and Morris (2004) contrasted the Davydov curriculum with the type of work pursued in the generalized arithmetic strand of early algebra. Davydov’s curriculum developed algebraic structures from relationships between quantities rather than developing algebra as a generalization of number. “Thus, while children in the U.S. have pre-algebraic experiences that are numerical, Russian children using Davydov’s curriculum have pre-numerical experiences that are algebraic” (Schmittau & Morris, 2004, p.61).

Furthermore, Schmittau (2004) described the children using Davydov’s curriculum as having high levels of procedural competence and mathematical understanding. The children were able to analyze and solve problems using conceptual understanding rather than to appeal to a rule. Basic arithmetic procedures taught as concepts were captured with representational schematics and used psychological models to build understanding. Throughout the curriculum, building concepts was consistently connected to students’ prior knowledge. New problems were presented in ways to enable the development of mathematical concepts through scaffolding, and without violating
prior understanding. The introduction of a new mathematics topic was accomplished by building on prior learning and connecting to a larger conceptual system (Schmittau, 2004).

Schmittau and Morris (2004) highlighted subtle but critical differences in the approach Davydov took in developing curriculum to that of a constructivist approach. The Davydov curriculum first drew students’ attention to the most general form of a concept. By contrast, the constructivist approach arrived at generality through a series of concrete examples. The constructivist approach relied on the use of inductive reasoning. Davydov’s approach intertwined the inductive and deductive to allow instructional activities to reveal the nature of a concept. Davydov believed that through the development of theoretical thinking, children would develop abilities to understand mathematical concepts at the most abstract and generalized level (Schmittau & Morris, 2004).

**Theoretical Underpinnings of Measure Up**

On a theoretical level, Vygotsky’s distinction between types of concepts formed the basis for classroom interactions. This was first forwarded by Davydov and later by Schmittau, where the primary objective of the curriculum was to develop students’ capabilities to think theoretically. It was believed that that led to the capability to understand mathematical concepts at the most abstract and generalized level. Schmittau and Morris (2004) argued that if students are to understand mathematics at the structural level they must learn to think theoretically. This is an issue of particular concern at the elementary level. The Davydov curriculum did not postpone the introduction of abstract
thinking until adolescence, instead the curriculum relied on the structure of mathematics and instructional pedagogy to elicit theoretical development.

**Vygotsky’s influence.** Vygotsky’s (1978) work led to foundational theories about the concept of human development. Vygotsky presented complex, multifaceted ideas in what later became known as the cultural-historical development of concepts. This theory led to educational applications throughout the world. Vygotsky’s work provided a perspective to view the conceptual development of children’s thinking in a social setting. Koshmanova (2007) commented that despite the existence of various applications from Vygotsky’s work, there was little research explaining how these theories became practice. Furthermore, publications from different countries have not portrayed Vygotsky’s work uniformly (Koshmanova, 2007). Vygotskian theory itself appeared to undergo a process of concept development shaped by the culture, values, and communication of the thoughts of researchers.

One of the major contributions to the Measure Up program was Vygotsky’s perspective of how concepts were developed. According to Vygotsky, the genesis of concepts was the genesis of intellectual operations. These operations were understood as abilities to generalize, to identify features of objects, to compare and differentiate, and to synthesize thoughts (Sierpinska, 1992). Vygotsky (1978) discussed two types of concepts, spontaneous or empirical concepts and scientific or theoretical concepts. Spontaneous concepts were developed when children abstracted properties from concrete experiences or instances. This included experiences in their everyday life. These concepts were related to the world in a direct but relatively ad hoc manner (Wells, 1994). Children
spontaneously appropriated everyday concepts through social interaction in joint activities (e.g. norms of fair play).

Scientific concepts were developed from formal experiences with scientific properties. Schmittau (1993) discussed scientific properties characterizing the very nature of mathematics. Scientific concepts are abstract and more general than spontaneous concepts and their primary relationship is to other concepts within a relevant system (Wells, 1994). Tharp and Gallimore (1995) suggested the use of “schooled” rather than “scientific” concepts as it was termed in translations of Vygotsky’s work. The term was intended to refer to the conceptual understanding that was produced from social and instructional interactions in formal education. Whereas schooled concepts are characterized as being learned “downward,” from generalization to concrete examples, everyday concepts are learned “upward,” from sensory experience to generalization (Tharp & Gallimore, 1995). Thus, the balance and interplay between the two approaches framed ways to artfully design education.

For example, in the Measure Up work students were presented with objects from their everyday world (e.g., two empty containers) and asked to make comparisons. From their experiences at home and elsewhere, the students had developed understandings of what was bigger. They had developed spontaneous concepts through their informal experiences and could make comparisons about size. Through Measure Up instruction the students learned how to articulate what attribute they were comparing (i.e., height of the containers, amount of volume the container could hold, areas of the bases of the containers, mass or weight of the containers). The students then used paper strips, diagrams, and equations to represent generalized forms of the mathematical relationship.
The use of concurrent representations prepared students for the development of scientific concepts. For example, a generalization such as volume T > volume C would then give meaning to specific instances like 10 > 8 and 5 + 9 > 5 + 8.

Vygotsky’s work (1978) was founded on the belief that a person’s psychological functions were developed through social interactions. This was in opposition to the belief that human development occurred spontaneously and was motivated by the internal structures of human biology (i.e., Piaget’s depiction of developmental stages). Vygotsky’s cultural-historical theory proposed that learning or education preceded development. That is, it was through the processes of appropriation, where a new perspective replaced an existing one, and assimilation, where a new idea merged with an existing one, that psychological development was advanced. A central component to Vygotsky’s theory of learning was the role of a teacher or more able peers in guiding a child’s learning experiences. Beyond observing and imitating, the child engaged in problem solving activities together with the teacher. The guidance served as scaffolding that enabled the child to problem solve at a level higher than she otherwise would be able to independently. According to Vygotsky (1978), “good learning” only occurred if it preceded a child’s development.

To emphasize the significance of teaching, Vygotsky described the distance between a child’s ability to problem solve independently and the child’s ability to problem solve with guidance from an adult or more able peers. He named this distance the zone of proximal development (ZPD). Learning stimulated a child’s internal development and this happened when the child interacted with more capable individuals.
(teachers) or in cooperation with peers. In this regard, “children grow into the intellectual life of those around them” (Vygotsky, 1978, p. 88).

Teaching toward the ZPD was a subtle yet significant component of the Measure Up program delivery. Problem solving activities worked on in small groups were typically followed by class discussions led by students. Students were the discussants who conjectured and tested their ideas with others. These discussions were strategically structured and artfully facilitated by Measure Up teachers to support students’ understanding of the concepts. Dougherty and Venenciano (2007) described class discussions about measurement and the use of a unit of measurement. The teacher did not tell students what they needed to learn from the lesson but rather provided opportunities for student understanding to develop. The pedagogical beliefs were consistent with Vygotsky’s theories. The structure of the Russian mathematics curriculum enabled the Measure Up researchers to apply these beliefs in their work.

**Davydov’s influence.** Soviet scholars Davydov and El’konin (Davydov, 1975a, 1975b) advanced Vygotsky’s work by creating school curriculum and founded the developmental approach to education (Koshmanova, 2007). This approach followed the notion that education contributed to students’ (cognitive) development if it was directed toward the ZPD. As applied in mathematics education, teachers provided challenging problems, sometimes a problem that contradicted what they understood, which the students could solve only after they developed awareness of certain behaviors, knowledge, and skills associated with their thinking processes (Koshmanova, 2007).

Davydov’s developmental approach to education was a deductive way of presenting information and organizing learning and thinking (Koshmanova, 2007).
Influenced by Vygotsky’s work, Davydov hypothesized that instruction, and more generally upbringing, could determine a child’s psychological development. He argued that empirical generalization was not an appropriate approach to studying abstractions within the domain of mathematics. Davydov believed that instruction should instead focus on the development of theoretical generalizations (John-Stiner, 1992). This was consistent with Vygotsky’s view where only scientific or theoretical concepts were believed to be true concepts (Schmittau & Morris, 2004). Thus it follows that in order for children to truly learn mathematics they must learn to think theoretically and use abstractions with understanding.

Koshmanova (2007) experimented with Davydov’s instructional approach with approximately 100 students in her teacher education class. These adult learners were not accustomed to the instructional style of developing theoretical knowledge from the presentation of problem solving activities. Although the Davydov approach was designed to stimulate the formation of students’ scientific concepts, Koshmanova (2007) concluded that it was not possible to address the appropriate theoretical level with a diverse group of that size. Koshmanova (2007) determined that the large class of students represented an overwhelming diversity of ZPDs. She concluded that the approach only worked with “intellectually homogeneous groups of students who have positive motivation to learn and are ready to grasp complex theoretical knowledge” (Koshmanova, 2007, pp. 84–85). Koshmanova (2007) proposed supporting students by gradually involving them in the process of learning and assisting them in making personal connections to the theoretical content.
The conclusion from the Koshmanova (2007) study was based on a sample of college students; a different population from that on which Davydov based his research. Perhaps if the students had been exposed to this learning approach in their earlier years of studying mathematics, they would have developed the tolerance for building theoretical knowledge. Another possible reason for the disappointing outcome may be that the class was too large for the instructor to facilitate instructional conversations. A smaller class would have enabled the instructor to facilitate instructional conversations with a larger proportion of students and steer the conversations to suit the needs of the group.

Davydov led a team in developing school curriculum (Davydov, Gorbov, Mukulina, Savelyeva, & Tabachnikova, 1999) and studied the effects on development and learning. Vygotsky had theorized that identification and discrimination were processes that operated on the level of material concrete objects only (Sierpinska, 1992). Davydov adopted the perspective that thinking and knowledge should originate in object-oriented work (e.g., comparing the attributes of two empty containers). He believed that work with concrete objects provided opportunities to develop theoretical knowledge (e.g., how to mathematically represent the relationship of the attributes). The theoretical knowledge was then used to solve a general class of problems (e.g., number problems that expressed a relationship, such as equality). This perspective was a contrast to the more common approach of delaying the presentation of generalizations or abstractions until students were developmentally ready (i.e., a Piagetian perspective) and then requiring the students to successfully solve a large number of problems at that level.

In conjunction with the theory on the development of concepts, Davydov (2008) described the development of various metacognitive abilities he suggested were
independent of the subject matter. These abilities included a person’s ability to examine the foundations of his or her own actions (reflection) and the ability to see commonalities from a previous method to a new one (generalization). Davydov’s approach to theoretical applications in the curriculum focused on a carefully guided inquiry he termed the ascent from the abstract to the concrete. It was believed that such a curriculum would transform the mathematics content and presentation of the mathematics to create a deep, systematic and theoretical understanding of the subject.

Davydov (2008) proposed a series of logical and psychological theses to (a) determine curricular content and (b) present mathematics in a manner for student thinking to evolve from abstraction to concretization. Davydov presented the study of mathematics as a genetic evolution of ideas, necessitated by given conditions. Learning was believed to be a process of observation and discovery through interactions with object-related sources. Davydov’s mathematical content and structure emerged from this perspective. Objects provided a means for the derivation of initial, essential, and universal relations. Relations were then reproduced in graphical or letter models to promote the study of properties in pure (abstract) form. Then in the final stages, students developed flexibility between performing actions mentally and physically (Davydov, 2008).

Davydov (2008) believed the opportunity for applying Vygotsky’s theory began at Grade 1. Together with his colleagues, Davydov designed an experimental school for mathematics curriculum to focus on bringing students “to the clearest possible understanding of the conception of real number” (Davydov, 2008, p. 147). Similar to typical U.S. elementary school mathematics curricula, developing understanding of the concept of real number is the goal of this curriculum. But unlike the typical curriculum
that begins with a focus on number, the Davydov curriculum begins with a study of the abstract concept of mathematical quantity. Both approaches lead to foundations for the study of mathematics but they take different paths to deriving the forms of real number (e.g., natural, fractions, negative). El’konin (1972) and Davydov (1988, 1990) are credited with applying the Vygotskian conception of learning activity and verifying the hypothesis that psychological development could be advanced by instruction (Zukerman, 2004).

**Implementation and Effects of the Russian Curriculum**

The Davydov mathematics program was studied in a cross-cultural and cross-curricular comparison study on algebraic reasoning in early and middle adolescent students. Morris and Sloutsky (1995) found differences in students’ algebraic deductive reasoning which tended to increase with age. The researchers described the Davydov curricular focus on developing abstract deductive, law-based reasoning. They characterized this curriculum as abstract-to-concrete and structural-to-procedural because of its approach to developing concepts of structure prior to numerical applications and prior to emphasizing algebraic transformations.

Morris and Sloutsky (1995) examined student outcomes in four settings; a Moscow school that implemented Davydov’s curriculum \( (n = 120) \), a non-experimental school in Moscow \( (n = 89) \), a school in England that used an experimental curriculum, the Harper et al. (1987) *National Mathematics Project* \( (n = 120) \) and a non-experimental school in England \( (n = 120) \). Three open-ended response problems and follow-up interviews were used to assess differences among the groups. One of the tasks was from
the Chelsea Diagnostic Mathematics Algebra subtest (Hart et al., 1985), an assessment that focuses on interpretation of variables and algebra preparedness.

Results showed that the Davydo group was more likely to interpret letters as variables, formulate correct equations, and formulate algebraic proofs. Furthermore, the Davydo group was more likely to (a) use algebraic deductive arguments, (b) believe algebraic proof establishes “universal validity”, (c) use arithmetical structure, (d) manipulate algebraic expressions correctly, and (e) acquire concepts of generalized numbers, variables, and givens. The researchers suggested that the different curricular approaches tended to lead to different conceptual organizations of children’s mathematical understanding.

**Results from the Russian research site.** The El’konin-Davydo work began in the late 1950s using design experiments to test their hypotheses about psychological development. Zukerman (2004) contended that the Russian psychological community found learning activity theory, an aspect of the El’konin-Davydo work, to be the most consistent and validated practical embodiment of Vygotsky’s work in education and psychological development. Presently, their curricula are used in about 10% of the elementary schools in Russia (Zukerman, 2004).

Zukerman (2004) conducted a longitudinal study of the students at Moscow school No. 91, one of the schools that systematically used the experimental curricula. The students were given a language lesson when they were in the first grade (five- and six-year-olds) and a second lesson when they were in the fourth grade. The same teacher taught both lessons. The lessons were videotaped and used to analyze the verbal and nonverbal communication.
A microanalysis of the two lessons showed that both lessons contained the characteristics of learning activity and tasks were, “aimed at the search for new means and tools for solving problems” (Zukerman, 2004, p. 11). The teacher was consistent in her instructional delivery by not guiding each step of the lesson but rather facilitated her students’ search for generalization. This was a contrast to a typical teacher’s role where the teacher’s assistance came in the form of telling. Students shared their conjectures and their peers expressed their agreement, disagreement, or uncertainty with nonverbal signs. The students then discussed their opinions.

Findings from the student interactions in this last phase revealed qualitative differences in students’ psychological development. Zukerman (2004) characterized the first grade students’ attitude as, “We (I) think otherwise, and thereby you are mistaken” (p. 12). This was contrasted with the fourth grade students’ attitude, “Following your arguments, we would arrive at this conclusion; however, the facts contradict your reasoning” (p. 12). The students who had three more years of instruction were more able to use each other’s conjecture to compare contradictory data rather than directly evaluate the hypothesis. The frequency of the remarks made by fourth graders who held that attitude was as high as 21%. In the first grade discussions this was only 3%, a highly significant difference by the chi-square criterion (Zukerman, 2004).

Zukerman (2004) described mathematics instruction in a first grade class as consistent with significant properties of learning activity. Occasionally students were presented with a problem-solving task that was incompletely defined, referred to in the Russian curriculum as a trap, in order to create cognitive dissonance and provoke different opinions from the students. These traps provided the teacher with diagnostic
information to identify student misconceptions and determine the strength of their understanding. The teacher’s role was to facilitate the class discussion rather than evaluate the students’ contributions or offer her solution to the task. Students were expected to follow the arguments of their peers and engage in tasks to link their own thinking with others through collaboration.

Zukerman (2004) tracked students from the study described above, into the seventh grade. She selected two problems from the Program for International Student Assessment (PISA) 2000 to assess students’ higher level cognitive processes. Students from the experimental school were compared with a comparison sample of students from a traditional school. The comparison school was acknowledged as a top achieving school in the standard PISA testing, relative to the other tested Russian schools. Zukerman (2004) was particularly interested in assessing students’ abilities to reflect on the knowledge and experience and to apply it to real life tasks. The ability to reflect was characterized as thinking, generalizing, and using insights to analyze the mathematical task and create one’s own problem (Zukerman, 2004). The PISA problems that assessed reflection abilities in a 2002 national report were identified by Kovaleva, Krasnovskii, Krasnokutskaya, and Krasnyanskaya (as cited in Zukerman, 2004).

The findings of the study showed that the two groups of students performed similarly on solving problems that required basic computations or relied on definitions often found on conventional assessments. However on the tasks that required the use of reflective abilities, the students from the experimental school outperformed the others. On the first reflection problem, 71% of the experimental school students solved the problem, whereas only 42% from the comparison school and 18% of all Russian students
in the study were successful. Similarly on the second reflective problem examined in this study, 53% of the experimental school students solved the problem compared to 32% and 8% of the comparison school and all students in the Russian sample, respectively. Zukerman (2004) concluded that the learning experiences at the El’konin-Davydov experimental school developed and sustained students’ reflective abilities. Furthermore, students who did not have the El’konin-Davydov education were deficient in their reflective abilities to solve mathematical problems that demanded higher-level thinking.

**Theoretical consciousness.** Davydov (2008) studied the features and development of theoretical consciousness. He defined this to be the development of children’s psyche and capabilities to reflect, analyze, plan, and test hypotheses abstractly. He believed that overall results of this research showed that it was possible for instruction to lead children’s development through consistent experiences from learning activities. However, he also stated that more research was needed to create a convincing theory of developmental instruction.

Zak (as cited in Davydov, 2008) investigated the development of theoretical consciousness by conducting a comparative study between elementary school children in the Davydov experimental classes and children in typical Russian classes. A student who had developed analytical capabilities would be able to immediately find and use a general principle for problem solving several problems of a single class. External features of the problem were used as distracters but construction of the exercises remained the same. Students who solved the problems quickly were considered to have developed the target analytical skills. These students did not need the time to compare or identify an appropriate solution method for the new problem.
Reflection abilities were measured using a two-part assessment. The first part was a set of exercises, structured to fit in one of two classes, with each exercise having a different set of conditions to keep them superficially different from each other. The second part required the child to group the tasks he just solved. Correctly grouping the exercise by class was an indication of reflection capabilities.

Planning abilities were measured with practice exercises followed by a series of basic exercises. The practice exercises were designed to help a child master a relatively simple action. The child was then asked to solve a related series of basic problems, with an increasing level of difficulty. If the child solved the exercises successfully and rapidly he was considered to have planning capabilities.

Aggregated data over multiple years showed that an average of 50% of children in the experimental classes had mastered the reflection assessments after three years of instruction, whereas after the same amount of time in the control classes only a 33% of the children had done so. Similarly, after three years of instruction, analysis was used in solving problems by 73% of the students in the experimental classes and only 60% in the other classes. Results from the planning assessments showed that 64% of the students in the experimental classes mastered such planning compared to 50% of the students in the other classes. All results were statistically significant and favored the experimental classes without exception (Zak as cited in Davydov, 2008). Furthermore, the advanced abilities of students in the experimental classes were maintained after one, two, and three years of instruction. The percent of experimental class students who mastered the analysis tasks after two years of instruction was the same as the percent of other class
students after three years of instruction. This was also the case for the students who mastered the reflection tasks.

Findings from this evaluative study showed that students from the Davydov mathematics program were outperforming students in the control group on all components of theoretical consciousness. Between-group differences in achievement on the analysis and reflection assessments were similar, with the experimental students performing at levels that were a year ahead of the other students. When taken together, the assessments used to measure analysis and reflection required students to reason deductively, solve problems by satisfying necessary and sufficient conditions, and generalize. Students who have these capabilities have developed the theoretical consciousness or higher level thinking skills needed to do more advanced mathematics.

The assessments used by Zak (as cited in Davydov, 2008) were similar to assessments used by Moshman and Franks (1986) and Morris (2000). The Russian research used the assessments to measure theoretical consciousness. The works of Moshman and Franks (1986) and Morris (2000) were used to assess logical reasoning capabilities. Although different terminology was used, it seems likely that the researchers intended to measure the same construct. Differences may be due in part to English translations of the Russian work. Word-for-word translations are not seamless. Another possible explanation is that the difference is due to a lack of agreement in the field on how to concisely define and assess such components of cognitive development. This issue is addressed in my study and will be revisited in a later discussion.

Studies of the Davydov curriculum indicated that students had advanced cognitive development (theoretical consciousness) as a result of the mathematics learning activities.
The aim of my study is to build a model that best describes the relationships among the variables in my study. The model could produce parallel findings about the effects of the Measure Up program to the effects attributed to the Davydov program. One particular relationship of interest is if experience in the Measure up program led to advanced cognitive development and abilities as did the Davydov program. The term logical reasoning will be used to describe the advanced cognitive process of interest in my study.

**Logical Reasoning**

Both logical reasoning and theoretical consciousness were presumed to be constructs. Unlike behaviors that are directly observable, constructs must be inferred by patterns of performance on assessments believed to elicit their presence or lack thereof. In the Tatsuoka et al. (2004) study, the assessment tasks were coded by a team of researchers, teachers, and mathematicians. For that study a child who had developed logical reasoning was able to reason from cause to effect, and solve problems that required the problem be analyzed in cases (e.g., instances when $x > 0$, $x < 0$, and $x = 0$).

Logical reasoning also applied to types of problems that required high-level thinking, such as deductive thinking, if-then reasoning, satisfying necessary and sufficient conditions, and generalizing.

Morris (2000) studied children’s abilities to distinguish between logical and nonlogical argument structures while simultaneously negotiating their personal knowledge of extraneous, contextual factors. For example one of the assessment tasks was, “If dogs are bigger than elephants, and elephants are bigger than mice, then dogs are bigger than mice” (Moshman & Franks, 1986, p. 155) creates a situation contrary to what is true in the real world. To respond successfully to the task a child must be able to attend
to the structural validity of the argument and accept the content of the argument as inconsequential. Morris hypothesized that (a) younger children have the capacity to recognize the form of an argument but they do not attend to structural relationships and evaluate for logical consistency, and (b) these children tend not to accept contrary-to-fact argument content as a basis for reasoning. This research was conducted with 220 children, ages 8–11 years, to understand how children’s performances changed over time.

Morris (2000) used a five-tiered model to structure the assessments. First, children were given a pretest to determine if they were able to independently use “logical form” or truth of a proposition as a means for sorting a set of propositions. Second, children at each grade level (third, fourth, and fifth) were randomly assigned to one of three conditions, structural relationship, fantasy context, or combined condition. The structural and combined conditions cued the children to attend to structural relationships whereas the fantasy condition did not. The combined and fantasy conditions cued children to monitor the introduction of personal knowledge about truth of content whereas the structural relationship condition did not.

The third tier of the assessment used a sorting and ranking task from Moshman and Frank (1986). All children were assessed on this tier, regardless of their assignment in the second tier. Only those who applied validity in the sorting and ranking tasks were assessed in the fourth and fifth tiers.

Morris (2000) found that the majority of 8-to-11 year-olds did not meet the criterion for explicit understanding of logically valid arguments. That is, the children were not able to distinguish between logical and nonlogical forms of argument without cuing. However, a substantial number of children in the combined condition (cuing
children to pay attention to structural relationships and to monitor the application of their personal knowledge about the truthfulness of the relationship) were successful at making the distinction.

Findings from this study supported the idea that distinguishing between logical and nonlogical arguments involved the coordinated application of several skills. These included treating the proposition as a whole and attending to information from individual statements while ignoring or suppressing irrelevant real-world information.

The assessment Morris (2000) used to measure logical reasoning was very similar to the measure used to assess reflection as a result of the Davydov mathematics program. Thus, if the Davydov curriculum is viewed as a curriculum that advances students’ higher level thinking skills (i.e., logical reasoning), then its use in the elementary grades to address mathematical deficiencies plaguing U.S. education may produce the necessary impact on students’ overall achievement.

The assessments in the Morris (2000) study shared similar characteristics to the items that measured logical reasoning from the Tatsuoka et al. (2004) study. Subjects in the Morris (2000) study were assessed for their abilities to decipher the information, identify the relevant structure of the arguments, and respond to the assessments in ways that satisfied the necessary conditions. Success on these tasks required skills and thinking processes that assessed students’ abilities to reason from cause to effect, use deductive thinking, and use if-then reasoning. These were also aspects of the logical reasoning definition used by Tatsuoka et al. (2004). Another parallel can be made between the
sorting assessments used by Morris (2000) and mathematics problems that require analysis by cases, a component of logical reasoning defined in the Tatsuoka et al. (2004) study.

The characterization for logical reasoning, as presented in the literature, is multifaceted. An appropriate definition for logical reasoning should therefore encompass the various aspects of the construct. Based on the description of the assessments used to measure theoretical consciousness in the Russian literature, it seems prudent to accept that the similarities in the assessments indicate measurement of the same construct. Therefore the implementation of a program found to positively affect theoretical consciousness would be expected to produce similar effects on assessments of logical reasoning.

**Summary**

Kaput (2008) advocated for broadening the traditionally held view of algebra and envisioned using algebra to integrate mathematics across the grades. This is consistent with Dekker and Dolk’s (2011) call for a long-term learning trajectory of algebra across the grades. Longitudinal results from a Measure Up study indicated advanced algebra preparedness in middle school students that could be attributed to the emphasis on measurement and generalized (algebraic) notations from the elementary school experience.

Other studies have shown that Measure Up students displayed more advanced psychological development than what was typically expected of students in their grade. Measure Up students were able to analyze what seemed to be discrepant conditions and reason logically about the problem in order to solve it. Furthermore, Measure Up students
were able to explain their mathematical thinking and reflect on their problem solving strategies in class discussions. These findings mirror the results from the studies about the effects of Davydov’s curriculum where children not only developed understandings of the scientific concepts in mathematics but also developed theoretical consciousness not typically observed in their peers.

The Measure Up program was grounded in the theoretical work from Vygotsky and Davydov. Vygotsky’s characterizations of spontaneous concepts and scientific concepts were instrumental to developing the pedagogy for the Measure Up program. Students’ understandings, generated from interacting with common objects, were critical starting points for developing mathematical concepts. Through instruction the students were capable of extracting a generalization, or abstraction of the spontaneous concept, and used the abstraction to construct a meaningful mathematical structure that could be applied to solving specific cases.

Another relevant feature of Vygotsky’s theory was the depiction of the teacher’s role. The teacher was a facilitator, guiding students such that learning (teaching) led to advancing students’ cognitive development. This was especially important in establishing instruction associated with the ZPD.

Davydov’s mathematics curriculum research and development were designed to embody Vygotsky’s work. Davydov (2008) commented that more research was needed to create a convincing argument of the theory of developmental instruction. Testing the Measure Up application of these theories would address Davydov’s comment and Koshmanova’s comment on the need for more research to explain how Vygotsky’s theories become practice. Furthermore, research on the effects of Measure Up could
provide valuable insights about learning and the development of mathematics concepts, as they contribute to the work of early algebra.
CHAPTER 3

METHODS

Prior studies have shown that students’ experience with the Measure Up mathematics led to stronger preparedness for algebra (Slovin & Venenciano, 2008), and stronger abilities to attend to structure and symbolic representations (Dougherty & Slovin, 2004). Insights from Davydov’s foundational work linked the Russian mathematics curriculum and instruction with advancements in students’ cognitive abilities. It is hypothesized that the Measure Up implementation of the Davydov curricula will produce similar effects on students’ conceptual development, as it is connected with their theoretical understanding of mathematics.

The intent of my study is to build a theoretical model from these prior findings to identify the relationships between the variables described below. Structural equation modeling (SEM) is used to identify and represent the hypothesized relationships between the variables in this study. My research question is, What statistical model best describes the relationships among Measure Up experience, age, prior achievement in mathematics, and logical reasoning capabilities in predicting algebra preparedness?

Participants and Setting

Participants in this study are students from the primary research laboratory school where the Measure Up research is conducted. The sample includes fifth and sixth grade students ($M = 11.3$ years, age range = $9.9–12.4$ years) during the 2009–10 and 2010–11 school years. In accordance with the school’s mission, students are admitted by stratified random lottery to reflect the Hawai‘i public school population with respect to ethnic background, socioeconomic status, and prior achievement as measured by standardized
test for students at Grade 3 and higher. All students take the same set of courses; the school does not group the students by ability level for instructional purposes.

Additionally, the school attempts to maintain an equal gender ratio. The demographic variables, prior achievement and age, confirm a normal distribution, characterized by values within a ±1 for skewness and ±2 for kurtosis (see the descriptive statistics in Table 3.1). The Measure Up experience variable is a dichotomous measure where 1 = Measure Up experience and 0 = no Measure Up experience.

Table 3.1

Descriptive Statistics for Exogenous Variables, N =129

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Mathematics Achievement</td>
<td>5.88 (1.9)</td>
<td>2–9</td>
<td>-0.2</td>
<td>-0.9</td>
</tr>
<tr>
<td>Age</td>
<td>11.3 (0.5)</td>
<td>9.9–12.4</td>
<td>-0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Measure Up experience</td>
<td>0.31 (2)</td>
<td>0 or 1</td>
<td>-0.8</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

There are typically 10 students at each grade level K–5 and 54 students at each of the grade levels 6–12. This produces approximately 20 students in Grades 5 and 6 who have Measure Up experience in each year, with some students transferring out of the school. There were a total of 40 Measure Up students in this study. Measure Up was the only mathematics program for the elementary students. The average number of years of Measure Up experience was 4.2.

Test Administration

The logical reasoning and the algebra preparedness assessments were administered on separate days during the mathematics class. Testing occurred in November during each of the two school years. This month was an ideal time of the year because it was consistent with data collection from prior Measure Up project
administrations of the CDMTA. Furthermore, relative to the academic year, by the third month of school the students have adapted to the school and class environment. The students are well accustomed to solving problems, justifying their thinking and being facilitated through their learning.

Students were given testing instructions as a class from either their regular teacher or from me. Students who were absent on a test day took the test during a lunch break or other free time upon their return to school. The instructions for testing are in Appendix A.

**Sample Size**

McDonald and Ho (2002) conducted a meta-analysis of studies that used SEM. They found that sample sizes tended to be larger than the $N = 129$ used for my study, but some studies were as low as $N = 70$. Determining the appropriate sample size for quantitative studies had been addressed in several studies but researchers have not agreed to a standard set of rules about how large a sample needs to be in order to achieve stable parameter estimates (Raykov & Marcoulides, 2006). One suggested guideline was for a sample size to be more than 10 times the number of free parameter estimates (Hu, Bentler, & Kano, 1992). Another suggestion was for a minimum of 200 subjects and a ratio of at least five subjects for each parameter to be estimated. However, several considerations must be made for each study, such as, the psychometric properties of the variables, strength of the relationships between the variables, and the size of the model (Raykov & Marcoulides, 2006).

The number of students at the laboratory school where the Measure Up program is being researched restricts the sample size for this study. This is a limitation that will be addressed in the discussion of this study.
Variables

The predictor variables are Measure Up learning experience, prior mathematics achievement, age, and logical reasoning capability. The logical reasoning variable was simultaneously used as an outcome variable as it was hypothesized to mediate the effects of the other variables on algebra preparedness.

**Measure Up experience.** For the purpose of this study, any amount of Measure Up experience was hypothesized to result in a difference between Measure Up and non-Measure Up student performance. Although it seems likely that more Measure Up experience would result in a stronger effect, that may not necessarily be the case. Most students who begin as Kindergartners at the school remain throughout Grade 12. However, some students enter at later elementary grades due to attrition. These students were provided with several supports to “catch-up” to their peers who had prior Measure Up experience. The supports include the accessible nature of the curriculum, extra tutoring, support from collaborative interactions with their peers, and the knowledge and skills from their prior mathematics experiences. Thus it is difficult to precisely assess the impact of fewer years of Measure Up with various levels of additional support. Therefore for the purpose of my study, the Measure Up experience is treated as a dichotomous variable.

**Age.** A student’s age on the first day of test administration was calculated using an age calculator accessible via the Cornell University Medical College web site (Wells Medical College of Cornell University, 2009). This variable is a potentially useful covariate given the rapid changes typically anticipated of children in the pubescent age range.
Prior mathematics achievement. Prior achievement was a reasonable choice for an indication of what a student is prepared to learn next. This variable was measured by students’ stanine scores on the Stanford Achievement Test, v. 9. The test is a norm-referenced, standardized assessment, and is part of the assessments that make up the Hawai‘i State Assessment. The stanine scale is based on a normal curve and the scores are on a scale of 1–9, with 5 representing the mean of the normal distribution. The sample for my study is slightly above the mean of the normal distribution.

Logical reasoning. Logical reasoning capability is not directly observable but is a hypothetical construct. Previous studies (Tatsuoka et al., 2004; Zak as cited in Davydov, 2008; Moshman & Franks, 1986; Morris, 2000) established the existence of logical reasoning and developed ways to measure it. The characterization of logical reasoning as identified by the Tatsuoka et al. (2004) is applied in this study. That is, the ability to solve mathematical problems in the following ways: reason from cause to effect, analyze the problem by cases (e.g., \( x > 0 \), \( x < 0 \), \( x = 0 \)), use deductive thinking, use if-then reasoning, and generalize.

The instrument used to measure logical reasoning capabilities and the processes to select and score indicator items are described below.

Algebra preparedness. Algebra preparedness is also a construct. For this study it is defined as a student’s capability to interpret the meaning of a variable in flexible ways (e.g., to represent a number or a set of numbers).

The instrument used to measure algebra preparedness and the processes to select and score indicator items are described below.
Instruments

The challenge in developing appropriate measures for hypothetical constructs was addressed in the Tatsuoka et al. (2004) study. They said, “Measurement of such unobservable latent variables can be performed only indirectly from observable item scores by making inferences about what misconceptions, leading to what incorrect responses, did a tested individual most likely have” (p. 904). Manifestations of a latent construct, or lack thereof, can be observed and measured. Measurement of the manifestation (behavior) is accomplished by using an appropriate instrument like a test, questionnaire, or self-reports (Raykov & Marcoulides, 2006). Instrument development for each of the constructs in my study is described below.

Logical reasoning instrument. Mathematical problems prompting the application of logical reasoning skills were identified in the Tatsuoka et al. (2004) study. The problems were from the released TIMSS-R (1999). The subset of problems identified by Tatsuoka et al. (2004) also served as an appropriate instrument for my study because it was a neutral measure of logical reasoning capabilities and it did not bias prior Measure Up experiences. Another reason for using these problems was that item analyses were available and that provided me with additional information for comparing my findings.

The appropriateness of the mathematics content was a concern because the TIMSS-R items were designed for students who experienced two-to-three years more schooling than the students in this study. Therefore the set of items identified by Tatsuoka et al. (2004) were first reviewed to select a subset of problems that were developmentally appropriate for my study. Problems that relied on knowledge about
advanced mathematical topics not typically experienced by the end of Grade 5 were eliminated.

A sample problem from my instrument is presented in Figure 3.1. Solving the problem correctly requires students to (a) make sense of the given information presented on the number line, (b) analyze the result of unit movement along the number line, and (c) reason about which of the multiple choice options would be the outcome that satisfies the conditions of the problem.

Point $P$ (not shown) on the number line is 5 units from point $N$ and 2 units from point $M$.

Where is point $P$ located?

A. Between $O$ and $L$

B. Between $L$ and $M$

C. Between $M$ and $N$

D. To the right of $N$

*Figure 3.1.* Sample item used to measure logical reasoning. Originally written for TIMSS-R (1999) and identified by Tatsuoka et al., (2004) as an assessment item that measures logical reasoning. Problem #N12 on the TIMMS-R (1999). Problem #A9 for my study.
Following my selection of items I sent a copy of both the complete set of released TIMSS-R items and a copy of my selected problems to my advisory group. This advisory group consisted of an expert in measurement and evaluation and two mathematics education researchers, one who was also an expert in mathematics assessment. The review from the advisory group was then compared against the items coded as logical reasoning in the Tatsuoka et al. (2004) study. Finally, discrepancies between the reviews were discussed with a former researcher from the Tatsuoka et al. (2004) study and I then produced a draft instrument.

A pilot test of the draft was needed to establish the approximate amount of time needed to complete the assessment. A volunteer (son of a colleague) who was within the age range of the students in this study was recruited to assist in the pilot test. He was instructed to do as many problems as he could in 30 minutes. The volunteer was recognized as a high-achieving student at his school. His pacing and achievement on the test were used as a gauge for the upper limit of a student’s potential. Final adjustments to the draft were made to produce the final instrument. The results from each of the reviews are compared in Appendix B.

Scoring for each item was recorded as 1 = correct and 0 = incorrect. This data were then used in the analyses to select indicator variables. The indicator selection process is described below in the development of the measurement model.

**Algebra preparedness instrument.** The instrument developed in Küchemann’s (1981) research was used as the algebra preparedness instrument for this study. Participants in Küchemann’s algebra assessment study (1981) were students 13–15 years old from urban and rural areas in England. The major testing occurred toward the end of
the school years in 1976 and 1977. To establish item difficulty levels, approximately \( N = 10,000 \) students were tested on this and the other subtests that comprised the Chelsea Diagnostic Math Tests (CDMT) (Hart et al., 1985).

The algebra assessment emphasizes students’ understanding of a variable. Unlike other algebra readiness assessments, this assessment does not rely on more mature reading abilities or algebraic manipulation abilities. The problems were designed to assess children’s interpretation of letters in mathematical problems and their ability to accept abstract context.

At Levels 1 and 2 students are considered to be capable of using letters with assigned values or as objects and evaluating the letter. At the more advanced levels (3 and 4) letters represent a set of unspecified values. To solve a problem at the highest level requires a student to accept a more abstract context, where the letters are used as variables.

The algebra subtest of the CDMT continues to be highly regarded by contemporary researchers. In addition to its use for determining a student’s preparedness for an algebra course, the items from the algebra assessment were previously used and referenced in several related areas of research. These include work in early algebra (Schliemann et al., 2003), students’ understanding of algebra symbols (Kieran, 2007; Coady & Pegg, 1993), the development of algebraic skills (Falle, 2005), cognitive and linguistic aspects of learning algebra (MacGregor & Stacey, 1994), and the development of students’ algebraic deductive reasoning (Morris & Sloutsky, 1995).

The CDMT algebra subtest (CDMTA) is an appropriate instrument for this study because of the emphasis on assessing conceptual understanding of variables. The tasks
were used in prior Measure Up studies as well as in several other recent studies, validating the merits of the tasks. Furthermore, the data from prior Measure Up administrations of the CDMTA provided a basis for comparison with this new set of data.

There are 53 items on the CDMTA that require response; some items are part of a multipart problem. One of the problems used as an indicator variable in this study is presented in Figure 3.2.

\[
\text{If } e + f = 8 \\
\text{Then } e + f + g = \underline{\quad} 
\]

*Figure 3.2. Sample item used to measure algebra preparedness. Originally written for the Chelsea Diagnostic Math Tests (Hart et al., 1981). Problem # 5c on the CDMTA. Problem #X14 in this study.*

Scoring for each item was recorded as 1 = correct and 0 = incorrect. This data were then used in the analyses to select indicator variables. The indicator selection process is described below in the development of the measurement model.

**Research Design**

I am conducting a correlational study that uses statistical modeling techniques to compare the effects of variables on preparedness for algebra, with particular attention on the effects of Measure Up, for the purpose of constructing a theoretical model.

The statistical method I used was *structural equation modeling* (SEM). SEM is a method that that provides researchers with a comprehensive approach to quantify and test
theories (Raykov & Marcoulides, 2006). SEM is preferred over the traditional regression analysis when more complex multivariable relationships need to be studied. SEM also enables the testing for direct and indirect effects of variables in the model, thereby treating them as a system of interactions. I hypothesized that Measure Up experience, age, prior achievement in mathematics, and logical reasoning capabilities were variables in a system that predicted algebra preparedness. Furthermore, I hypothesized logical reasoning capabilities mediated the effects of the other variables on algebra preparedness. SEM is an appropriate method for this study to analyze the data for this study.

The structural equation model is a composite of a measurement model and a path model (McDonald & Ho, 2002). The path model is used to describe hypothesized causal relationships between latent variables. The measurement model is used to link observable or manifest variables as indicators of latent factors. For this study, logical reasoning and algebra preparedness were treated as endogenous, latent variables. That is, they are constructs that can only be inferred by the ways they each influence the observed indicator variables within the model. The indicator variables for logical reasoning and algebra preparedness were identified through the psychometric processes described below.

Prior to building the SEM models it is necessary to first develop the measurement model. The instruments for each of the constructs were selected as described above but only a subset of the items for each measure was used in the model. The subsets contained the indicator variables for each latent factor (logical reasoning and algebra preparedness). The process for selecting indicator variables is described below.
The measurement model. This model was created from first conducting an exploratory factor analysis (EFA) to identify the number and structure of the components that made up each of the measures. Then a confirmatory factor analysis (CFA) was used to verify the results. Theoretical conceptions, based on the definitions used in this study, of logical reasoning and of algebra preparedness were used to characterize each factor. The scree plots of the eigenvalues were generated by the statistical analyses in SAS (v. 9.1.3). These plots indicated the number of factors the data sufficiently defined.

The criteria for selecting indicator variables was based on strong correlations and conceptual relatedness with the latent variable they were intended to measure. Determining the composite reliability of the set for each latent construct was accomplished by using a method presented by Raykov (2007). This was an alternative to the Cronbach’s (1951) alpha coefficient. Although Cronbach’s (1951) alpha is one of the most widely used indices for determining internal consistency reliability (Hatcher, 1994), Raykov (2007) presented evidence to suggest this statistic is too limited and should be considered to be a lower bound of reliability.

Raykov (2007) presented a latent variable modeling procedure to generate point and interval estimates of instrument reliability following deletion of each component in a tentative scale. This was a more rigorous approach to addressing composite reliability. Unfortunately for my study, this statistic favors large samples. Nevertheless, discussion of composite reliability for this study is based on Raykov’s (2007) method as calculated using the Mplus software, v. 4.1.

The data for the selected indicator variables was dichotomously scored and necessitated the use of a nonlinear approach. Tetrachoric correlations were produced
within the statistical programs (e.g., SAS and Mplus) and used to generate appropriate results for factor analyses.

**The logical reasoning factor.** Unlike the CDMTA, the logical reasoning instrument was newly created for this study. Both EFA and CFA were conducted with the same data set \(N = 129\). Although it was recommended that the procedures not be used with the same data set, the sample was too small to attempt each process with half the class. The results were still useful in the SEM model building process. The selected assessment items resulting for EFA and CFA analyses were used as indicator variables for the logical reasoning factor in the SEM model.

**The algebra preparedness factor.** Data from prior Measure Up project administrations of the CDMTA assessment was used to conduct the EFA. The data was collected in November of 2006 and 2007 from students in Grades 5–8 from two public charter schools, an urban laboratory school \((n = 275)\) and a rural charter school \((n = 108)\). The EFA helped identify a subset of variables that was used to define the algebra preparedness factor.

A CFA then followed using the newly collected data \(N = 129\). The use of two different data sets was a more rigorous approach to analyzing the assessment items and created potentially more stable results (i.e., it is likely that similar findings about the items would be found with different samples).

**The path model.** The path or structural model for this study was built from the relationships among variables found in earlier studies. The Slovin and Venenciano (2008) algebra readiness study used a multiple regression model building technique, MAX-R. The maximum \(R^2\) improvement technique selected the variables with the most
explanatory power and included them first. In Step 1 the model was built with the variable with the highest $R^2$ and in Step 2 it used the highest two variables. The model building continued until all independent variables were included in a model (the final step). The researchers determined that a two-variable model that included Measure Up experience and prior academic achievement in mathematics was the most parsimonious and meaningful model. It significantly explained 40% of the variance in algebra preparedness. The relationship between these variables will be studied further in the current study.

Figure 3.3 displays the path model for this study. It represents the relationships among the three exogenous variables and the two latent variables or constructs. Measure Up experience, age on the first day of testing, and prior mathematics achievement provided background information about a subject. These variables were used as covariates in the model and were hypothesized to impact the logical reasoning factor and algebra preparedness factor.

The assumptions for the use of SEM are (a) appropriate sample size, (b) correctly specified and multivariate normal distribution of the data, and (c) linear relationships among variables in the model. The assumption of linear relationships is a result of the development of the statistical procedures stemming from linear regression analysis.

**SEM model development.** The process for developing the SEM model necessitated comparisons between several competing models, as well as goodness-of-fit indices in order to justify the selection of a final model. The competing models were generated by restricting or freeing the hypothesized paths between variables and factors, and by comparing the fit indices of the models.
In the first model, all the paths from the predictor variables were allowed to load on the algebra preparedness factor, directly, and indirectly as they were mediated by the logical reasoning factor. This model was nested within all subsequent models to compare changes in the chi-square statistic.

In contrast to the design for Model 1 where all paths are free, Model 2 was designed to establish a conceptual boundary of the effects. In Model 2, all the direct paths to algebra preparedness were restricted to zero, and the paths to logical reasoning

Figure 3.3. Proposed path model showing all paths between the variables.
remained free. This model was a test for the effects of the observed predictors on algebra preparedness by only those mediated by logical reasoning capabilities (i.e., preparedness for algebra is completely explained by one’s logical reasoning capabilities).

Since previous findings about Measure up experience indicated that the age variable had a minimal effect, the age effects were checked. Model 3 was a test of the effects from age, with both paths from age is restricted to zero. Model 4 was a test of the direct effects from age on algebra preparedness.

The design of subsequent models was dependent on the findings from these first four SEM models. Other models incorporated the paths from age that contributed to stronger models. Models 5–7 were tests of the direct, indirect, and both direct and indirect effects from Measure Up on algebra preparedness. Model 8 was a test of the direct effect of prior mathematics achievement. All model development was conducted using the Mplus, v. 4.1 software.

Data Analysis

The Mplus software was also used to perform confirmatory factor analysis for the logical reasoning and algebra preparedness assessment data. The results helped construct the measurement model. Indicator variables with strong reliabilities and conceptual relatedness were identified for each of the latent constructs. Identifying this set of variables provided robust measures of the factors of interest. Raykov and Marcoulides (2006) advocated for at least three indicator variables per latent variable to obtain a more complete and reliable measure. Too many variables create more complicated and statistically strenuous models; thus, only five of the items from each assessment were used in my analyses.
The Mplus software was also used for structural equation modeling. Researchers (Muthén & Kaplan, 1992) found that standard estimators developed for SEM were not suitable for noncontinuous variables because that violated the assumption of normality. This study included variables that were scored dichotomously (Measure Up experience and the indicator variables) and therefore an alternate estimator was needed. An appropriate choice developed in Mplus is the robust estimator, the means and variances-adjusted weighted least-squares or WLSMV. This estimator was used in prior studies (Flora & Curran, 2004; Wegmann, Thompson, & Bowen, 2011) and was believed to produce accurate estimates and standard errors under a variety of different conditions.

The path diagram illustrates the hypothesized causal relationships among the variables in a study. It represents the structural equation model in graphical form. Standardized path coefficients were generated using the Mplus software. These values indicate the nature and strength of the correlations of the independent variables with the dependent variables, as well as the correlations among the independent variables. In each of the models tested, fit indices were compared to determine if the model was acceptable, that is, how closely the model’s correlation estimates were to the observed correlations. If the model was acceptable, the paths were then checked for significance. Comparison of the fit indices from competing models was necessary to find the most parsimonious model.

Hu and Bentler (1999) discussed concerns for determining adequacy of fit indices for their study and selecting goodness-of-fit cutoff scores for evaluating model fit. McDonald and Ho (2002) conducted a meta-analysis and found that the most widely used
indices were the Bentler comparative fit index (CFI) and the root mean square error of approximation (RMSEA). Therefore, I included these measures in my study.

The CFI represents the ratio between the discrepancy of the tested model with the baseline model, which is a more restricted model in which the observed variables are assumed to be uncorrelated. Although some researchers (e.g., Hilikari, Nevgi, & Komulainen, 2008; Charalambos, Pitta-Pantazi, 2007; Hatcher, 1994) used a criterion of .90, Hu and Bentler (1999) recommended a criterion of .95 as the acceptable CFI cut-off for a good fit.

In a simulation study using four models differing in sample size, model complexity, and distribution of data (i.e., normal or non-normal), Yu (2002) found that suitable cutoff criteria for some fit indices were strongly dependent on models. For a model with a sample size $N = 100$ and dichotomous outcomes, Yu’s findings suggested that the rejection rule of chi-square with a cutoff value of $p < .05$ might not be suitable for small samples, as it tended to overreject true models. He proposed lowering the cutoff value for such studies. The CFI cutoff value of 0.95 also seemed to overreject models with small samples. Since I am working with dichotomous variables and a small sample size, of $p < .05$ and CFI > .93 seemed to be reasonable values for this study.

The root mean square error of approximation (RMSEA) is an absolute fit index; that is, it assesses how well an a priori model reproduces the sample data. RMSEA values less than .06 are considered a good fit and values less than 0.08 are considered an acceptable model fit (Browne & Cudeck, 1993). However, Yu (2002) found that although cutoff values of .06 for RMSEA tended to overreject models, increased cutoffs of .07 or
.08 tended to underreject more complex models with slightly larger sample sizes. For this study, an RMSEA < .06 is reasonable.

In addition to the CFI and RMSEA, I am using the weighted root mean square residual (WRMR) and the Tucker-Lewis Index (TLI). The WRMR had the same tendency to overreject small samples with simple models at the .95 cutoff and underreject the complex misspecified models. Furthermore, Yu (2002) suggested an acceptable cutoff WRMR value of 1.0 for dichotomous outcomes. For this study, a criteria of WRMR < 1.0 is considered reasonable.

The TLI is a relative fit index that takes into account the number of parameters in the model. This index was found to overreject true population models at samples < 250 and severe non-normality when the cutoff value was set at .95 (Yu, 2002). Although Taub (2001) suggests that values over .90 are considered as excellent, for my study I am considering values over .95 as excellent fit and less than .90 as poor fit.

My model comparisons were also informed by chi-square difference testing. The use of the chi-square statistic alone is limited, as it tends to be sensitive to sample size. Jöreskog (1978) first proposed the rationale for chi-square difference testing. This test is likened to the test of change in $R^2$ when adding predictors to a set of explanatory variables in regression analysis (Raykov & Marcoulides, 2006). The chi-square difference test is used for testing models that are nested. For example, model A’ is said to be nested within model A if one or more of its parameters is restricted to zero or restricted to have a specific relationship (a numerical value) with the other variables. Researchers impose parameter restrictions based on theory and findings from prior
studies. Values for the test are generated by the difference in the fit between the two models relative to the difference in the degrees of freedom.

Nonsignificant chi-square difference values indicate that constraining the parameters in the nested model does not significantly worsen the fit (Raykov & Marcoulides, 2006). If this is the outcome, the more restrictive model, with constrained parameters, is preferred as it is more parsimonious. This series of model development and analyses led to the selection of a final model, informed by the theoretical basis of the study and refined by insights gained from the fit indices.

Calculation of chi-square is based on the assumption of normally distributed data, a problematic statistic for this study because of the dichotomously scored items. Therefore the DIFFTEST option in Mplus was used in conjunction with the WLSMV estimator in order to conduct chi-square difference tests. The WLSMV estimator transformed the data to emulate a continuous variable by creating a matrix of tetrachoric correlations. The DIFFTEST first estimated the less restrictive model. The computed derivatives were saved for the chi-square difference testing. This was followed by the analysis of a more restrictive model, which generated comparison values for Mplus to compute the chi-square difference test.
CHAPTER 4

RESULTS

Selecting Manifest Indicators for Logical Reasoning

In the preliminary factor models that I investigated, one factor was retained. That factor initially accounted for 63% of the common variance. Through subsequent testing, as discussed below, I later reduced the number of items defining the factor.

Using tetrachoric correlations to analyze the data, nine indicator variables were retained. The model was nonsignificant with $\chi^2 (19) = 22.5, p = 0.26$. This suggests the selected variables were representative indicators of the latent construct, a better indication than what would be found had they been randomly selected. In order to generate a concise set of items that produced a good measure of the construct, the factor loadings for each of the items were checked. Items with low loadings (less than .40) were eliminated. The remaining items were then analyzed to recheck the conceptual relatedness of the assessment tasks as a group. Through this process, a few more items were eliminated to produce the final set of indicator variables, all with relatively strong factor loadings (0.44–0.81). The five-indicator model fit the data very well, with $\chi^2 (5) = 7.20, p = 0.21$. Other fit indices (CFI = .97, RMSEA = 0.06, WRMR = 0.57) also supported these indicator variables for the logical reasoning construct. See Appendix C for the assessment items.

The values used to compute the composite reliability value are found in the tetrachoric correlation matrix in Table 4.1. These values were used to compute the composite reliability scale for the logical reasoning items. Using Raykov’s (2007)
method, the value for composite reliability was .76. This value shows that the items do a reasonable job of measuring the factor.

Table 4.1

Estimated Tetrachoric Correlations for the Logical Reasoning Manifest Variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>A5</th>
<th>A6</th>
<th>A9</th>
<th>A13</th>
<th>A14</th>
</tr>
</thead>
<tbody>
<tr>
<td>A5</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>.53</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>.32</td>
<td>.30</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>.56</td>
<td>.35</td>
<td>.30</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>.30</td>
<td>.17</td>
<td>.12</td>
<td>.55</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*Note.* Items A5–A14 are the logical reasoning indicator variables.

Selecting Manifest Indicators for Algebra Preparedness

In the preliminary factor models that I investigated, one factor was retained. That factor initially accounted for 51% of the common variance. Through subsequent testing, as discussed below, I later reduced the number of items defining the factor.

Using tetrachoric correlations to analyze the data, twelve indicator variables were retained. The model produced a mixed set of goodness-of-fit measures ($\chi^2 (20)= 45.79$, $p = 0$, CFI = .98, RMSEA = .10, WRMR = 1.098). Although all the variables had high factor loadings (0.58–0.99), the fit indices did not confirm a good fitting model. The variables were checked for conceptual relatedness against two frameworks. First, the problems were reviewed to confirm that each addressed the definition of algebra preparedness for this study. Another framework for reviewing the items was from level indicators for which the problems were designed in the CDMTA. Of the 12 items, five were at Level 3, four were at Level 2, and three were not used for level determination. At Level 3 students were able to use a letter as a generalized number where the letter can have more than one value in the problem. Level 3 was the highest level indicated for the
items under consideration for this study. Ultimately five items (all at Level 3) from the original 53 assessment items were selected as indicator variables for the algebra preparedness construct. The fit indices of the five-indicator model were a significant improvement over the twelve-indicator model ($\chi^2 (5) = 5.25, p = .39$, CFI = 1.0, RMSEA = 0.01, WRMR = .45). See Appendix D for the items.

The values used to compute the composite reliability value for algebra preparedness are found in the tetrachoric correlation matrix in Table 4.2. These values were used to compute the composite reliability scale for the logical reasoning items. Using Raykov’s (2007) method, the value for composite reliability was .89. This value showed that the items do a good job of measuring the factor.

Table 4.2

*Estimated Tetrachoric Correlations for the Algebra Preparedness Manifest Variables.*

<table>
<thead>
<tr>
<th>Variable</th>
<th>X14</th>
<th>X30</th>
<th>X38</th>
<th>X40</th>
<th>X25</th>
</tr>
</thead>
<tbody>
<tr>
<td>X14</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X30</td>
<td>.50</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X38</td>
<td>.36</td>
<td>.45</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X40</td>
<td>.75</td>
<td>.66</td>
<td>.56</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>X25</td>
<td>.57</td>
<td>.59</td>
<td>.42</td>
<td>.72</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*Note.* Items X14–X40 are the algebra preparedness indicator variables.

**SEM Models**

Results from model comparisons are in Table 4.3. In Model 1 I allowed the exogenous variables to load on all paths to the latent variables. Aside from the CFI, the fit indices were acceptable and therefore warranted further analyses. This first model explained fairly large variances in the constructs, 63.6% of the variance in logical reasoning and 87.1% of the variance in algebra preparedness. The overall results from Model 1 were acceptable and were consistent with theory. This justified further model
development to test the hypotheses and establish a model with stronger fit indices. Model 1 served as the baseline to compare against all other models. Subsequent models introduced restrictions on the paths. The restrictions were generated from theoretical implications and findings from earlier studies. Comparative information to evaluate how well the model fit the data provided insights for further modifications.

Models were tested to compare the results for fit and parsimony. Model 1 was a fully saturated model with effects of the three exogenous variables mediated by logical reasoning as well as being allowed to directly affect algebra preparedness. This served as the baseline model to which each subsequent model was compared.

Model 1 was the least restrictive model. All exogenous variables were allowed to load on algebra preparedness directly and indirectly. The indirect effect is the result of allowing the variables to be mediated by logical reasoning. Nearly all of the goodness-of-fit indices were acceptable, thus supporting the hypothesis that the model fits the data. A CFI = .93 was below the recommended .95 cut-off. However, since some researchers (e.g., Hilikari, Nevgi, & Komulainen, 2008; Charalambos, Pitta-Pantazi, 2007; Hatcher, 1994) use the criterion of .90, these fit indices as a whole support further analyses in developing a final model with stronger fit indices.
Table 4.3

*Goodness of Model Fit Statistics, N = 129*

<table>
<thead>
<tr>
<th>Model</th>
<th>( \chi^2 (p) )</th>
<th>df</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>WRMR</th>
<th>( \chi^2_{\text{diff}} (p) )</th>
<th>df</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All variables partially mediated</td>
<td>40.2 (.08)</td>
<td>29</td>
<td>.93</td>
<td>.91</td>
<td>.06</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. All variables fully mediated, no direct effects</td>
<td>39.4 (.10)</td>
<td>29</td>
<td>.93</td>
<td>.92</td>
<td>.05</td>
<td>0.97</td>
<td>2.67 (.45)</td>
<td>3</td>
</tr>
<tr>
<td>3. Partial mediation, no direct or indirect age effects</td>
<td>47.3* (.02)</td>
<td>29</td>
<td>.88</td>
<td>.86</td>
<td>.07</td>
<td>1.07</td>
<td>10.27* (.01)</td>
<td>2</td>
</tr>
<tr>
<td>4. Partial mediation, no direct age effect</td>
<td>39.4 (.10)</td>
<td>29</td>
<td>.93</td>
<td>.92</td>
<td>.05</td>
<td>0.95</td>
<td>0.11 (.75)</td>
<td>1</td>
</tr>
<tr>
<td>5. Partial mediation, no direct age or MU effects</td>
<td>39.7 (.09)</td>
<td>29</td>
<td>.93</td>
<td>.92</td>
<td>.05</td>
<td>0.97</td>
<td>1.81 (.41)</td>
<td>2</td>
</tr>
<tr>
<td>6. Partial mediation, no direct age, no indirect MU effects</td>
<td>38.6 (.11)</td>
<td>29</td>
<td>.94</td>
<td>.93</td>
<td>.05</td>
<td>0.96</td>
<td>0.76 (.68)</td>
<td>2</td>
</tr>
<tr>
<td>7. Partial mediation, no direct age, no direct or indirect MU</td>
<td>43.6 (.05)</td>
<td>30</td>
<td>.91</td>
<td>.90</td>
<td>.06</td>
<td>1.02</td>
<td>6.82 (.08)</td>
<td>3</td>
</tr>
<tr>
<td>8. Partial mediation, no direct age or SAT9, no indirect MU</td>
<td>39.4 (.12)</td>
<td>30</td>
<td>.94</td>
<td>.93</td>
<td>.05</td>
<td>0.96</td>
<td>1.22 (.75)</td>
<td>3</td>
</tr>
</tbody>
</table>

Note. The chi-square and degrees of freedom for WLSMV were estimated in Mplus. The comparison of the nested models (Model 1 compared to each alternate, presented in the table as \( \chi^2_{\text{diff}} \)) relied on calculations that accounted for the ordinal variables in the model. CFI = comparative fit index; TLI = Tucker-Lewis index; WRMR = weighted root mean square residual; RMSEA = root mean squared error of approximation.

*\( *p < .05 \)
An antithesis of a partially mediated model, where the exogenous variables are allowed direct and indirect effects, is a fully mediated model, where the exogenous variables are only allowed to indirectly affect algebra preparedness. This is accomplished in Mplus by setting a path to zero. In Model 2 restrictions were placed on the exogenous variables, preventing them from directly affecting the algebra preparedness factor. Only effects that are mediated by logical reasoning were allowed to influence algebra preparedness. The chi-square difference test resulted in a nonsignificant result with acceptable model fit. The chi-square difference test, the new RMSEA value, and the improved TLI all supported Model 2 over Model 1. Moreover, the CFI was the same, and although the WRMR value was slightly larger, it was within acceptable limits.

Findings from the Slovin and Venenciano (2004) study indicated that experience in Measure Up and prior achievement were the two highest contributors to algebra preparedness. The multiple regression analysis used in that study left unclear to what extent age had on algebra preparedness. The direct and indirect effects of age on algebra preparedness could be restricted to verify and provide more insights to the findings from the previous study. However, completely restricting age from direct and indirect effects resulted in a poorly-fitting model. Because all goodness-of-fit indices in Model 3 weakened, this model could not be supported as a plausible representation of the data.

Model 4 restricted only direct age effects; that is, indirect effects mediated through logical reasoning were allowed. The goodness-of-fit indices were all strong (e.g., RMSEA = .05, WRMR = .95), with CFI at an acceptable (if not optimal) value of .93.

Models 5–7 restricted direct age effects and tested the significance of the direct and indirect Measure Up effects to the model. The fit statistics supported the restriction
of either the direct or indirect effects but not both. As suggested by Model 7, completely restricting Measure Up effects resulted in marginally-fitting model (e.g., CFI = .90, WRMR = 1.02).

Figure 4.1. Path diagram of the final model for this study (Model 8).

Standardized path coefficients are on the single-headed arrows. Correlations appear on the double-headed arrows. All estimated path coefficients were significant at $p < 0.05$ or lower.

Model 6, partial mediation with restrictions on direct age effects and indirect Measure Up effects, resulted in a better fitting model than Model 1. As suggested in
Table 4.3, there was a slight improvement on the CFI value (.94 to .93, respectively). Although all other goodness-of-fit indices were reasonable, the path coefficient from prior achievement to algebra preparedness was nonsignificant and negative. For purposes of parsimony, this path was removed, suggesting that only prior achievement indirectly affected algebra preparedness through logical reasoning. The resultant model is summarized as Model 8 in Table 4.3.

Model 8 retained all the goodness-of-fit indices found in Model 6 but with one fewer parameter. The total standardized effect on algebra preparedness from Measure Up experience was .26, from age .24, and from prior achievement .67; all coefficients were significant at \( p = .05 \). Final factor loadings and path coefficients are displayed in Figure 4.1.
CHAPTER 5
DISCUSSION

The aim of the study was to uncover the nature and strength of the relationships among Measure Up experience, age, prior achievement in mathematics, and logical reasoning capabilities in predicting algebra preparedness. The findings presented in the previous chapter led to the development of a model that was theoretically justified and best described the statistical relationships for this sample.

My second SEM model resulted in a general improvement over Model 1 ($\chi^2_{\text{diff}} = 2.67$). This indicated that the logical reasoning factor mediated some of the effects on the algebra preparedness factor. Further model development was used to test specific effects from each predictor variable and to develop a model with acceptable fit indices.

Results from the Slovin and Venenciano (2008) study of Measure up experience and algebra preparedness showed that the age variable had a minimal effect on predicting algebra preparedness. Age was dropped from the final model of that study and was therefore worthy of reconsideration in developing models for this study. In Model 3, the two paths from age were restricted to zero and all fit indices declined below levels of acceptable fit. However when the path from age to logical reasoning was freed, the fit indices improved to levels similar to those in Models 1 and 2 (see results in Table 4.2).

I restricted the direct age effect and developed Models 5–7. Model 7 was not viable. Findings from Model 7, where both paths from Measure Up were restricted to zero, clearly indicated that some Measure Up effect was needed to establish a model with reasonable fit indices. Restricting the indirect effect of Measure Up on algebra
preparedness (Model 6) produced slightly better results in the fit indices compared to restricting direct effects (Model 5).

Since Model 6 seemed viable I proceeded to examine the path coefficients. I discovered a negative value for the path that led from prior achievement to algebra preparedness. Although it was nonsignificant, this was surprising because prior achievement is frequently used as an indicator of future achievement. The (λ = −0.16) path value indicated that when prior achievement increased by one standard deviation from its mean, algebra preparedness was expected to decrease by 0.16 its own standard deviations from its own mean, ceteris paribus. Regardless of significance, in this study the direct prior achievement effect seemed to have a far smaller impact than what was found in other studies. Since this path was nonsignificant I restricted it to zero and generated Model 8. This model retained the goodness-of-fit indices that were found in Model 6 and all path coefficients were significant.

Findings from the final model showed that algebra preparedness was mediated by logical reasoning capabilities. Measure Up experience was the only exogenous variable that significantly contributed to algebra preparedness; age and prior mathematics achievement were found to be positive, statistically significant, indirect contributors to algebra preparedness. The path coefficient from the Measure Up experience to algebra preparedness was positive, indicating that the Measure Up experience led to an increase in algebra preparedness and that students who had Measure Up experience were better prepared for algebra than students who did not. This supports earlier findings reported in the Slovin & Venenciano (2008) study.
I found that the CFI statistic was the most resistant to the changes introduced through the model testing process. Even when other indices pointed to a reasonable fit, the CFI hovered just below the cut-off. This might be due to the small sample used in this study. Being one of the most commonly studied statistics in model testing, a lack of improvement in the CFI was troubling.

I expected to find a large, significant, direct effect from prior achievement on algebra preparedness but that was not an outcome in this study. This may be due to the nature of the CDMTA items as they were focused on the interpretation of the letters in the mathematical context and did not require application of diverse, more typical mathematics concepts. However, the total effect of prior achievement on algebra preparedness was larger than the combined Measure Up experience and age effects. Results from this study were consistent with prior research, that is, prior achievement was a reliable predictor of future learning success.

I expected to find a significant, positive effect from Measure Up experience on logical reasoning, as I hypothesized a parallel connection with the Davydov curriculum and its effects on cognitive development. The findings from this study did not support that hypothesis. The logical reasoning indicator variables covered diverse mathematical topics. The items did not create a narrowly defined set. This was a consequence of the decisions I made with defining and measuring the logical reasoning construct. The number contexts of two of the indicator items (A5 and A13) are not focus areas in the Measure Up program. Yet each task requires the use of logical reasoning. My definition of logical reasoning may have been too broad to identify any specific advantages attributable to the Measure Up experience.
The logical reasoning factor was highly correlated with the algebra preparedness factor. One possible explanation is that students needed to have high-level thinking skills in order to indicate sufficient preparedness for algebra. This would be a consistent finding with what other researchers (e.g., Tatsuoka et al., 2004; Davis, 1985) have concluded. A justifiable conclusion for this study is that students who have strong logical reasoning capabilities are more likely to be prepared for algebra.

An alternative explanation is that the two factors may have inadvertently been defined as one underlying construct not identified in this study (e.g., general reasoning capabilities). However this explanation seems unlikely due to the differences in significance found along the paths from the exogenous variables. If this were true I would expect to find similar significant effects or non-effects coming from Measure up experience, age, and prior achievement on each of the two latent factors.

Logical reasoning capability was a major predictor of algebra preparedness. This was a tremendous effect, to the extent that effects from prior achievement could only be observed through logical reasoning. A possible explanation for this is that only a particular aspect of prior achievement, that which affected logical reasoning capabilities, contributed to a conceptual understanding of variables. As a result, logical reasoning capabilities mediated the effects from prior achievement.

Implications

Theoretical foundations for the Measure Up program were developed from Vygotsky’s work on concept development and Davydov’s work on the development of pedagogy and mathematics curriculum. Studies on the Davydov approach showed that students in the Davydov program developed advanced cognitive abilities (i.e., reflective
abilities) when compared to their peers in traditional Russian schools (Zukerman, 2004). The findings from the present study show that the Measure Up program successfully prepared students for algebra but that could not be attributed to the students having advanced logical reasoning capabilities. Since logical reasoning was highly correlated with algebra preparedness, a relevant question to pursue next is if the Measure Up students were deficient in some aspect of their mathematical experiences.

An alternate explanation to why the Measure Up experience did not correlate with advanced logical reasoning capabilities is that my characterization of logical reasoning was flawed. The characterization I used was based on the definition from the Tatsuoka et al. (2004) study which I found to be aligned with the tasks used to assess logical reasoning in the Morris (2000) and Zak (as cited in Davydov, 2008) studies. Establishing a more concise definition for logical reasoning might help to address this concern. Since the choice of a definition for logical reasoning determined the items in which the construct was manifested, a different definition or a different measurement instrument could lead to different results.

If I were to keep the existing definition, a test of the discriminant validity could be done to check if each instrument actually measured the construct for which it was developed. This could be accomplished by collecting data using multiple measures for each construct and conducting a multi-trait multi-method study. A different measure might assess the hypothetical construct in another way (e.g., the instruments used by Zak [as cited in Davydov, 2008] to measure theoretical consciousness), or it could validate the measure developed in this study.
**Implications for Measure Up.** The goal of this study was to develop a statistical model to describe the relationships among Measure Up experience, age, prior achievement, logical reasoning capabilities, and algebra preparedness. The results from this study supported those found in the Slovin and Venenciano (2008) Measure Up study linking experience in Measure Up and prior achievement to algebra preparedness. Other pertinent variables, such as the instructional delivery of the program, were not included in this study. Further studies could investigate the instructional impact to generate insights for better supporting Measure Up students in their mathematical development. One possible approach is to use a mixed methods study to collect data on various aspects of instruction. This might include measures on the degree to which teachers instruct for the development of logical reasoning (or theoretical consciousness), the degree to implement the program with fidelity, and the teacher’s mathematics content knowledge for teaching.

The Measure Up program’s emphasis on conceptual understanding and its approach to introducing relational thinking results in a delay of the introduction of numbers and computation. The program was developed with the premise that this understanding better prepares students for the study of more sophisticated mathematics, specifically, a middle school Algebra I course. Findings from my study suggest that the Measure Up program could be further developed to contribute more strongly to developing children’s logical reasoning capabilities. This might be in the way of a focus on providing students with more opportunities to think through problem contexts or broadening the coverage of topics in the later years of the program (e.g., introducing basic topics from number theory in Grade 5).
Implications for algebra preparation and logical reasoning. The findings from this study have implications for educators and curriculum developers who struggle with the trade-off between covering the content needed so that children will develop fluency in arithmetic manipulation skills on one hand and building conceptual understanding on the other. Some automaticity in arithmetic is necessary to enable students to focus their thinking and energies on making mathematical connections and other problem solving activities. However, if instruction in the elementary grades neglects to build the foundation for understanding symbolic representations, principles of arithmetic, and other early algebra concepts, students’ may not be sufficiently prepared for a middle grades Algebra I course. This implication is also consistent with the findings from the National Mathematics Advisory Panel’s 2008 report regarding the achievement of students in middle and high school algebra classes. Among the conclusions in the report were that students lacked a firm understanding of many basic principles of arithmetic and had difficulty grasping the syntax of algebraic expressions.

The findings of this study are consistent with what was found in the Tatsuoka et al. (2004) study regarding the correlation between advanced cognitive processes and higher overall achievement. The strong and positive correlation between logical reasoning capabilities and algebra preparedness implies that the development of logical reasoning may be necessary to prepare students for algebra. Not enough research exists about ways to influence the long-term development of logical reasoning. Additional research regarding how to develop logical reasoning capabilities is needed to continue to build insights about preparing children for the study of more advanced mathematics.
The findings about the relationship between prior achievement and algebra preparedness ($\lambda = -0.16$) are noteworthy. Further studies are needed, first, to validate these findings, and second, to understand the reasons for the unexpected result. If a trajectory were extended from the mathematics assessed on the standardized tests of earlier learning to the mathematics assessed on the algebra preparedness measure, I would expect to find a direct, positive, significant path from prior achievement to algebra preparedness. This was not the case. I suspect a mismatch exists between the mathematics learned for standardized tests and the mathematics for algebra preparedness. Research is needed to assess the possibility of a mismatch and, if it is determined that a mismatch exists, develop achievement tests that are more closely aligned with the development of mathematical skills and foundational processes that support the study of advanced mathematics.

The findings suggest that only the aspects of prior achievement that impact logical reasoning will contribute to algebra preparedness. If the educational objective is to better prepare elementary students for the rigors of a middle school Algebra I course, elementary mathematical experiences that emphasize the development of logical reasoning capabilities will contribute to that goal.

**Limitations**

**Method choice.** SEM allowed for the inclusion of several variables. Selection of variables, and conversely exclusion of others, was driven by the theories being tested. SEM provided fit indices based on the data included in this study. Consequently, there may exist an alternate model (i.e., the true model) with a different set of variables, also supported by theory. My use of SEM in this study does not allude to such models.
Selection of the indicator variables for the SEM factors was based on the data sets and the definitions I imposed for the constructs. My objective was to balance the mathematics content in the set of indicator variables with the statistical measures. I conducted several iterations of EFA and CFA in order to arrive at a concise number of items that had the desired psychometric values. Although the variables seemed to work well for this model, due to the nature of this process it is possible that I did not identify the true set of indicator variables. More research is needed to understand the complex nature of these constructs and how they can best be characterized for statistical models.

**Defining logical reasoning.** Comparing what other researchers described as logical reasoning and how they measured students’ capabilities was one of the challenges I faced in the review of the literature. I differentiated it from the field of mathematics known as *logic* and focused on the research about student learning in an educational setting. The relevant studies for my research were limited. Terms like *theoretical consciousness* and *contentful analysis* emerged from the Russian literature. I considered using Davydov’s characterization of the construct but decided to use a more modern definition as it was more closely connected with the assessments relevant to our society. I ultimately used the definition proposed by the Tatsuoka et al. (2004) study.

**No random assignment.** In this study there was no random assignment to the groups. That might have caused a selection bias and impacted the internal validity of the study. The school had admitted students using a stratified random sampling method. This study included all the students from that sampling process. Therefore Measure Up experience was not given to a completely random group. Thus it is unclear how the results of this study generalize to the population. The descriptive statistics show that the
sample was normally distributed by prior achievement giving an indication that the Measure Up experience may have similar effects with other students within the age range of those in this study.

The major limitation with testing the long-term effects of a curriculum is that instruction is situated in a school setting and there are several extraneous variables that cannot be controlled. This included class size and student attrition.

**Small sample.** Hu, Bentler, and Kano (1992) suggested using a sample size 10 times the number of free parameters in the model. The final model in this study had 15 free parameters, therefore a sample of at least 150 would be recommended. Unfortunately even though I was able to include all eligible students, my sample was lower than some of the recommended numbers.

**Limited number of variables.** Other aspects of the Measure Up learning experience were not addressed in this study. One aspect of particular interest is teacher training and support to implement the program. The pedagogical demands to teach the program are complex and the mathematics is more rigorous than what U.S. teachers expect from an elementary mathematics program. Kaminskaia (2003) commented on the profound challenges a teacher faces in developing the necessary skills and attitudes to become a master teacher of the El’konin-Davydov instructional system. The issues she discussed are very similar to the issues concerning the Measure Up research and development team. It may not be enough to measure if a student had Measure Up experience but also what Measure Up experience was received. Developing an instrument to measure teacher effect was not considered for this study but is a pertinent concern if
programs such as Measure Up are to be disseminated beyond the research and development stages.
APPENDIX A

Directions for Assessment Administration

Directions to the students: Today you will be taking a mathematics assessment. You will have 30 minutes to do as many of the problems as you can do. Don’t worry if you aren’t familiar with the mathematics, you may not have seen these types of problems before. Think about each problem to show your work and to use as scratch paper. Be sure you clearly mark your answer. This assessment is not part of your class grade. It is part of the on-going research efforts to help us understand how students learn mathematics and something you and your parents agreed to participate in when you enrolled at the Lab School.

For the logical reasoning assessment: 30 minutes maximum time allowed

For the CDMTA (Chelsea): 30 minutes maximum time allowed

Practice Problems for the Chelsea: As time permits, do the class problems as a class, allow students to think independently first and then have one or two share their answer. Do not verify correct answer and/or explanation; simply tell students where to record their answers on the blank.
## APPENDIX B

### Review of Recommended Items for Logical Reasoning

<table>
<thead>
<tr>
<th>Subset identified by my definition</th>
<th>Items from Tatsuoka et al. study</th>
<th>Experts’ Feedback</th>
<th>Used in Pilot</th>
<th>Recoded</th>
<th>Logical Reasoning Instrument</th>
<th>Used as indicator variable, based on CFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>B09</td>
<td>X</td>
<td></td>
<td>X</td>
<td>A1</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>B11</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A2</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>D11</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F07</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F08</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F11</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A3</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>H11</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J10</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>A4</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>J14</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J15</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>A5</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>J17</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>A6</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>L12</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L13</td>
<td>X</td>
<td></td>
<td>X</td>
<td>A7</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>L15</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>A8</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>L16</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N12</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A9</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>N13</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N14</td>
<td>X</td>
<td></td>
<td>X</td>
<td>A10</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>N16</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A11</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>N17</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N19</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A12</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>P08</td>
<td>X</td>
<td></td>
<td>X</td>
<td>A13</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>P10</td>
<td>X</td>
<td>?</td>
<td>X</td>
<td>A14</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>P12</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A15</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R10</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A16</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>R14</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T01</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A17</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>R15</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T02A</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A18</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>T02B</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V01</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A19</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>V04C</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>A20</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td></td>
<td>20</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. “?” = Uncertainty about including the item
All logical reasoning items are from the released items of the TIMSS-R 1999 Mathematics Assessment.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Assessment Task</th>
</tr>
</thead>
</table>
| #A5           | The height of a boy was reported as 140 cm. The height had been rounded to the nearest 10 cm. What are two possibilities for the boy’s actual height?  
  Answer: __________________cm and __________________cm |
| #A6           | Shade in $\frac{3}{8}$ of the unit squares in the grid. |
### #A9

Point $P$ (not shown) on the number line is 5 units from point $N$ and 2 units from point $M$.

Where is point $P$ located?

A. Between $O$ and $L$
B. Between $L$ and $M$
C. Between $M$ and $N$
D. To the right of $N$

### #A13

If 4 times a number is 48, what is $\frac{1}{3}$ of the number?

A. 4
B. 8
C. 12
D. 16
The figures show four sets consisting of circles.

<table>
<thead>
<tr>
<th>Figure 1</th>
<th>Figure 2</th>
<th>Figure 3</th>
<th>Figure 4</th>
</tr>
</thead>
</table>

a) Complete the table below. First, fill in how many circles make up Figure 4. Then, find the number of circles that would be needed for the 5th figure if the sequence of figures is extended.

b) The sequence of figures is extended to the 7th figure. How many circles would be needed for Figure 7?

Answer: ____________

c) The 50th figure in the sequence contains 1275 circles. Determine the number of circles in the 51st figure. Without drawing the 51st figure, explain or show how you arrived at your answer.
APPENDIX D
Algebra Preparedness Items Used as Indicator Variables

All algebra preparedness items are from the Chelsea Diagnostic Mathematics Tests: Algebra.

Variable #X14, only part (c) is used for this item.

a. If \( a + b = 43 \)  
then \( a + b + 2 = \) ______.

b. If \( n - 246 = 762 \)  
then \( n - 247 = \) ______.

c. If \( e + f = 8 \)  
then \( e + f + g = \) ______.
Variable #X25, only part (d) is used for this item.

This square has sides of length $g$. So for its perimeter, we can write $p = 4g$.

What can we write for the perimeter of each of these shapes?

a. $p = \underline{\hspace{2cm}}$

b. $p = \underline{\hspace{2cm}}$

c. $p = \underline{\hspace{2cm}}$

d. $p = \underline{\hspace{2cm}}$
Variable #X30, only part (b) for this item

\[ a + 3a \text{ can be written more simply as } 4a. \]

Write these more simply, where possible:

a. \[ 2a + 5a \] ____________

b. \[ 2a + 5b \] ____________

---------------------------------------------------------------------------------------------------------------------

Variable #X38

What can you say about \( r \) if \( r = s + t \) and \( r + s + t = 30 \)?

________________________________________________________________________

---------------------------------------------------------------------------------------------------------------------

Variable #X40, only part (b) for this item

In a shape like this you can work out the number of diagonals by \textit{taking away} 3 from the number of sides.

So, a shape with 5 sides has 2 diagonals.

a. \ A shape with 57 sides has _____ diagonals. 

b. \ A shape with \( k \) sides has _____ diagonals.
REFERENCES


Davydov, V. V. (1966). Logical and psychological problems of elementary mathematics as an academic subject (A. Bigelow, Trans.). In D. B. Elkonin & V. V. Davydov (Eds.), *Learning capacity and age level: Primary grades*, (pp. 54–103). Moscow: Prosveshchene.


P. Boero (Eds.), *Handbook of research on the psychology of mathematics education* (pp. 11–49). Rotterdam: Sense.


Mplus (Version 4.1) [Computer software]. Los Angeles, CA: Muthén & Muthén.


Arzarello (Eds.), *Proceedings of the sixth congress of the European Society for Research in Mathematics Education* (pp. 241–250). Lyon, France: Institut National De Recherche Pédagogique.


