# VERBAL VERSUS PICTORIAL REPRESENTATIONS IN THE QUANTITATIVE REASONING ABILITIES OF EARLY ELEMENTARY STUDENTS

by

Katrina Ashley Korb

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy degree
in Psychological and Quantitative Foundations (Educational Psychology)
in the Graduate College of
The University of Iowa

July 2007

Thesis Supervisor: Professor David F. Lohman

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

---

PH.D. THESIS

---

This is to certify that the Ph. D. thesis of

Katrina Ashley Korb

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Psychological and Quantitative Foundations (Educational Psychology) at the July 2007 graduation.

Thesis Committee: ______________________________
David F. Lohman, Thesis Supervisor

______________________________
Robert Ankenmann

Alice M. Atkinson
______________________________
Alice Atkinson

______________________________
Nancy Ewald Jackson

______________________________
Joyce L. Moore

To My Siblings:
Andrea Korb,
my favorite waterfall-repelling partner,
*and*
Andrea Gathings, Emeral Green, Ronetta Jenkins,
Tony Perry, and Danielle Washington,
my siblings in Christ.

And whatever you do,
whether in word or deed, do it
all in the name of the Lord
Jesus Christ.

Jesus

ACKNOWLEDGEMENTS

This dissertation is the product of an entire community of educators. In particular, my advisor David Lohman rescued me from the swamp of potential dissertation topics and helped me to find a project that I loved. He was always available for advice and for editing previous drafts, which I am very grateful for. Robert Ankenmann, Alice Atkinson, Nancy Jackson, and Joyce Moore also served on my committee and provided value suggestions for improvement.

I would like to thank the faculty, staff, and students of Tonganoxie Elementary School. In particular, I am grateful for the kindergarten teachers: Nancy Gibbens, Ginger Lobb, Denise Smith, and Debby Wedel; first grade teachers: Jill Conrad, Spencer Dale, Brenda Frost, Kathie Riddle, Linda Vernon, and Ruth Wickey; second grade teachers: Sarah Carttar, Mona Dibble, Natalie Frese, Kyla Sandhoefner, Kendra Schobert, and Deanna Sittner; and a fifth grade teacher: Cindy Korb. The teachers were incredibly cooperative and I greatly appreciated their support. I would also like to thank the principal, Tammie George, the secretaries (particularly Debbie Becker who provided administrative assistance), and all of the staff who allowed me to borrow empty rooms. It was a delight to spend time at TES.

I am also thankful for my fabulous peers who have offered feedback and support throughout graduate school. My two officemates, Michelle Croft and Joni Lakin, provided great companionship and insights. I also appreciated the friendship

and support of Melissa Chapman and Kyndra Middleton. Suzanne Swiderski gave me all of the advice necessary to actually *complete* a dissertation.

Last but certainly not least, I am thankful for my family, who has unwaveringly supported me in all of my endeavors and adventures. My parents, John and Cindy Korb, have provided me with innumerable educational opportunities throughout my life (including all of those Civil War battlefields). They taught me everything, including the value of work hard, the love of learning, and the most effective way of protecting a car from severe weather. My sister, Andrea Korb, has stuck by me through thick and thin and I will never outgrow her companionship.

# ABSTRACT

Quantitative reasoning primarily involves reasoning about quantitative sets that can be represented by number words, Arabic numerals, or an image-based mental model. However, most current measures of quantitative reasoning abilities rely heavily on number words and Arabic numerals. If test takers represent quantity using an image-based mental model, then most measures of quantitative reasoning demonstrate construct underrepresentation. This seriously threatens a valid interpretation of the assessment in most educational situations. The purpose of this study was to examine how kindergarten, first, and second grade students represent and reason with quantity.

Two quantitative reasoning tasks, Equivalence and Number Series, were administered to 140 kindergarten through second graders and 9 fifth graders. Both tasks were administered with a Numeral condition (Arabic numerals) and a Pictorial condition (pictures). Both tasks also had a third condition: students could choose to use pictures or Arabic numerals for the Equivalence task (Choice) and items were administered with a combination of pictures and Arabic numerals for the Number Series task (Mixed).

On the Number Series task, all students performed better in the Numeral condition than the Pictorial and Mixed conditions. However, kindergarteners performed better on the Equivalence task when using pictures whereas both first and second graders demonstrated similar performance in the Pictorial and Numeral

conditions. Kindergarteners preferred using pictures on the Equivalence task whereas second graders chose all formats equally.

Results of this study suggest that kindergarteners perform differently on quantitative reasoning tasks that afford a verbal counting structure and nonnumerical part-whole structure. Therefore, test developers should critically examine the requirements of quantitative reasoning tasks. If the task requires students to apply a part-whole schema, then concrete referents should be made available to aid early elementary students as they solve the task. On the other hand, if the task requires students to apply a verbal counting structure, then Arabic numerals or counting words sufficiently match the structure that early elementary students use to solve the task.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# CHAPTER 1

# INTRODUCTION

Quantitative reasoning abilities are one of the most important aptitudes for and outcomes of formal schooling (National Association for the Education of Young Children [NAEYC] & National Council of Teachers of Mathematics [NCTM], 2002; NCTM, 2000). Reasoning refers to the process of drawing a conclusion using evidence and strategies (Leighton & Sternberg, 2003; Wason & Johnson-Laird, 1972). Therefore, quantitative reasoning consists of reasoning in which individuals use mathematical relationships and properties to draw conclusions (Carroll, 1993; McGrew, 2005). Quantitative reasoning can be distinguished from quantitative knowledge (McGrew, 2005). Quantitative knowledge includes mathematical concepts and skills, such as knowledge of mathematical symbols, operations, and properties, that typically are acquired through formal schooling (Carroll, 1993). On the other hand, quantitative reasoning consists of making inferences or deductions with well-understood quantitative concepts. Therefore, the distinction between quantitative reasoning and quantitative knowledge reflects the degree of novelty of the procedures and strategies necessary to solve the quantitative task.

Quantitative reasoning abilities figure prominently in the goals that the NCTM (2000) set for educators. One goal for mathematics education emphasizes that students should develop *number sense* by understanding various ways of representing numbers and relationships among numbers. Another goal states that all students in pre-kindergarten through twelfth grade should be able to “apply and adapt a variety of appropriate strategies to solve [mathematical] problems" (NCTM, 2000, p. 53). These skills are essential to quantitative reasoning because students must have a rich understanding of number in order to be able to reason with quantities.

NAEYC and NCTM (2002) emphasized the importance of reasoning with quantities for young children as they endeavor to make sense of their environment. Moreover, the NAEYC and NCTM acknowledged that mathematical proficiency provides a solid foundation for future success in school because understanding and reasoning with quantities are necessary not only for success in mathematics classes, but also for learning science, social studies, and for acquiring technological literacy.

Quantitative reasoning abilities also have strong relationships with important educational and occupational outcomes. Quantitative reasoning ability in middle school is associated with achievement in high school and college (Benbow, 1992), as well as graduate school (Kuncel, Hezlett, & Ones, 2001). Educational achievement, income, and adult creative attainment can also be predicted by prior quantitative reasoning abilities (Wai, Lubinski, & Benbow, 2005). In addition, quantitative reasoning predicts subsequent academic achievement considerably better for non-native English speakers than verbal measures of cognitive abilities (Kuncel et al., 2001). The relationship of quantitative reasoning with measures of academic success is most likely due to its strong relationship with the general intelligence factor (e.g., Keith & Witta, 1997).

*Validation of Measures of Quantitative Reasoning*

Validation of test scores is essential in the development and evaluation of cognitive assessments (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education

[NCME], 1999). Validation consists of evaluating the plausibility of the proposed interpretations of test scores and other outcomes of an assessment (Kane, 2006). A main threat to validity is construct underrepresentation (Messick, 1994). Construct underrepresentation reflects the extent to which an assessment does not measure important aspects of the construct that it was designed to measure (AERA, APA, & NCME, 1999). Consequently, the quality of a cognitive assessment depends upon how well it requires the fundamental cognitive processes that it was designed to measure. For this reason, psychological theory plays a vital role in the validation of cognitive assessments. Assessment tasks should require the cognitive functions that are postulated by relevant cognitive theory (National Research Council, 2001).

Measurement experts recognize the central importance of cognitive theory in test development (e.g., Embretson & Gorin, 2001; Floyd, 2005; Mislevy, Steinberg, & Almond, 2003). Embretson (1983) extended the conception of construct validity beyond the relationship of an assessment with other measures to include construct representation. According to Embretson, construct representation is achieved by aligning the demands of an assessment to the psychological processes, strategies, and knowledge that influence performance. Likewise, Messick (1994) argued that cognitive theory should guide the development of tasks on cognitive assessments. Understanding how test takers complete test tasks is therefore a crucial aspect of developing and evaluating cognitive assessments.

Many different procedures may be used to gather information on how examinees solve test tasks. Understanding how examinees comprehend the problem provides one avenue. Cognitive theory has suggested that the manner in which a student mentally represents a problem may be more important for forecasting whether it will be solved than the complexity of the procedure necessary to complete the problem (Griffin, Case, & Sandieson, 1992). The label *mental model* refers to these representations of a situation maintained in an active state in working memory when solving a problem (Halford, 1993). Mental models assist an individual in understanding a problem, choosing a strategy for solving it, and managing the implementation of that strategy (Johnson-Laird, 1983; Norman, 1983). Since mental models drive performance on a task, test developers should understand the mental models that test takers typically use when attempting problems on the assessment. If an assessment task does not evoke the mental model that test takers commonly construct when solving problems in that particular domain, then the assessment may suffer from construct underrepresentation.

The degree to which test scores are influenced by processes that are extraneous to the intended construct, termed construct-irrelevant variance, also threatens a valid interpretation of test scores (AERA, APA, & NCME, 1999). If an assessment becomes more difficult for a particular group of examinees because of the influence of a variable that is unrelated to the construct of interest, then the assessment suffers from construct-irrelevant difficulty (Messick, 1995). The influence of verbal abilities on an assessment of an unrelated cognitive process is a pervasive source of construct-irrelevant variance (Haladyna & Downing, 2004). For example, a test that measures quantitative reasoning abilities should not be unduly affected by verbal demands such as knowledge of verbal labels for quantities or reading and comprehending a quantitative scenario.

In conclusion, measures of quantitative reasoning should demonstrate both adequate construct representation and a lack of construct-irrelevant variance. Acceptable construct representation would require an assessment of quantitative reasoning abilities to evoke mental representations and processes that would be classified as quantitative reasoning. Furthermore, these assessments must not involve irrelevant process, such as an advanced knowledge of mathematical symbols and operations or comprehension of a verbally stated quantitative problem.

*Measures of Quantitative Reasoning*

Current measures of quantitative reasoning tend to assess quantitative reasoning abilities using tasks that require knowledge of Arabic numerals or comprehension of a verbally presented problem. Individually administered measures of cognitive abilities with quantitative reasoning subtests include the *Stanford-Binet Fifth Edition* (SB-5), *Wechsler Intelligence Scale for Children-IV* (WISC-IV), and *Woodcock-Johnson III Tests of Achievement* (WJ-III ACH). Group administered measures of cognitive abilities with measures of quantitative reasoning include the *Cognitive Abilities Test, Form 6* (CogAT 6), *Inview,* and *Otis-Lennon School Ability Test, Eighth Edition* (OLSAT-8).

The SB-5 (Roid, 2003) uses both verbal and nonverbal content to measure quantitative reasoning, one of five broad cognitive ability factors. The Quantitative Reasoning subtest in the nonverbal domain consists of items that depict quantity using figures such as stars or blocks and require students to manipulate the quantities or find patterns. At the lower levels, the Quantitative Reasoning subtest in the verbal domain requires test takers to count objects, recognize Arabic numerals, and perform simple calculations. At the higher levels, the verbal subtest contains story problems that are simultaneously presented in written form and read aloud to the test taker.

The WISC-IV (Wechsler, 2003) measures general intelligence with four index scores that assess children ages 6 years through 16 years and 11 months. Only one subtest on the WISC-IV, Arithmetic, contains quantitative content. The Arithmetic subtest is a supplemental measure for the Working Memory index score. This subtest requires test takers to solve arithmetic problems within a given time limit that are read aloud by the test administrator. Items on the Arithmetic subtest involve counting pictures at the lowest levels and solving story problems at the higher levels.

The WJ-III ACH (Woodcock, McGrew, & Mather, 2001) has two tests that measure quantitative reasoning abilities. The Standard Battery contains Applied Problems that requires test takers to analyze and solve quantitative tasks. At the lowest levels, the items require counting objects. Slightly more difficult items contain pictures to represent a quantitative problem that is read aloud to test takers. The more advanced items contain written story problems that are also read aloud and to the test taker. The Extended Battery contains a Quantitative Concepts measure with two subtests. The Number Series subtest contains series of Arabic numerals even for early elementary students. The Concepts subtest measures knowledge of mathematical concepts, symbols, and vocabulary. The easiest items on this subtest require counting and recognizing Arabic numerals.

The CogAT 6 Primary Battery (for kindergarten through second grade; Lohman & Hagen, 2001) includes two subtests that measure quantitative reasoning abilities, Relational Concepts and Quantitative Concepts. Relational Concepts assesses the ability to discover relationships through questions that are read aloud to the test taker, such as determining which object is longer. The Quantitative Concepts subtest requires students to solve simple verbal story problems. In the Multilevel Battery (for third through twelfth grade), the quantitative reasoning subtests are Number Series (determining a pattern in a series of numbers and continuing the pattern with the next number), Equation Building (ordering sets of numbers and mathematical operations to make a meaningful number sentence), and Quantitative Relations (comparing two quantities that occasionally requires reading text).

The *Inview* (CTB/McGraw-Hill, 2001) assesses cognitive abilities in students from second through twelfth grade. One subtest on the *Inview*, Quantitative Reasoning, measures the ability to think with numbers and to solve quantitative problems through identifying patterns, inferring relationships among quantities, and drawing conclusions from quantitative data. The Quantitative Reasoning subtest includes different item formats at every level of the assessment. In Level 1 for second and third grade, the Grid Comparison item format requires test takers to determine which grid has the most black shading, the Number Operations Puzzle requires test takers to combine two of three numbers to make an accurate number sentence, and the Algebraic Substitution-Equations format requires test takers to substitute the correct number into an algebraic equation. Other levels of the test include item formats that consist of number analogies, determining equality between sets of numbers, and manipulating a quantity according to a flow chart. While the *Inview* does have some item formats that do not contain number words or Arabic numerals, there is no assessment for kindergarten or first grade students.

The OLSAT-8 (Otis & Lennon, 2003) has three measures of quantitative reasoning included in the Nonverbal Cluster of subtests: Number Series (a format identical to the CogAT 6), Numerical Inference (analyzing how two numbers are related and applying the same rule to another set of numbers), and Number Matrix (determining the missing number in a number matrix). However, these subtests only begin at Level E for fourth grade students. No subtest assesses quantitative reasoning for kindergarten through third grade students in the Nonverbal cluster. In the Verbal Cluster for all levels of the test, Arithmetic Reasoning was designed for test takers to solve verbal problems that rely on numerical reasoning. In the Arithmetic Reasoning test for kindergarten through second grade, problems are read aloud for test takers to solve. For these items, pictures depict the problem, but test takers must comprehend the verbal statement of the problem in order to understand the meaning of the pictures. Arithmetic Reasoning items for third grade and higher contain written story problems.

Most of these measures of quantitative reasoning assume that test takers can use number words and Arabic numerals to represent quantitative tasks. This assumption seems questionable since early elementary students have had limited exposure to the symbol system of mathematics. Quantitative reasoning largely consists of reasoning about quantities (Griffin, 2003). Indeed, Griffin defined mathematics as “a set of conceptual relations between quantities and numerical symbols” (2003, p. 8). Accordingly, quantity can be represented orally with number words, in writing with Arabic numerals, or mentally by replicating the objects in a set. In time, the concrete spatial representations can become increasingly abstract (e.g., a number line). Griffin hypothesized that competence in quantitative reasoning involves creating sophisticated relationships between these three systems. This hypothesis is similar to Halford’s (1993) suggestion that reasoning ability may consist of the ability to map one representational system onto another. The reliance of most current measures of quantitative reasoning on number words and Arabic numerals therefore neglects the mental representation of quantity through images of concrete objects. If test takers tend to use a verbal representation of quantity with number words and written Arabic numerals, then these measures of quantitative reasoning demonstrate adequate construct representation. However, if early elementary students also use a mental-image based (or nonnumerical) representation of quantity, then these measures may exhibit construct underrepresentation. Furthermore, many measures of quantitative reasoning require test takers to understand a verbally presented quantitative problem that potentially introduces construct-irrelevant variance.

If students mainly rely on number words and Arabic numerals when solving quantitative tasks, then they should use verbal working memory resources since mental models guide the processes that are used to solve tasks (Johnson-Laird, 1983; Norman, 1983). However, previous research has established that students in kindergarten through second grade tend to place greater demands on spatial resources in working memory when solving arithmetic items, whereas older elementary students tend to use both spatial and verbal resources (Holmes & Adams, 2006; McKenzie, Bull, & Gray, 2003). Therefore, early elementary students do not appear to rely on verbal processes when engaged in quantitative tasks, whereas older elementary students who have more experience with the verbal representation of number rely on both verbal and spatial processes.

Young students also tend to use overt strategies for solving difficult calculation items, such as counting their fingers to interpret the quantities necessary to solve a calculation task (Siegler & Shrager, 1984). Counting fingers to solve a mathematical task is similar to mentally replicating the objects in a set because, in both cases, the student relies on a concrete representation of the quantities to solve the problem.

Using verbal representations of quantity on quantitative tasks (number words and Arabic numerals) requires two fundamental concepts. First, students must understand that a set of objects can be described with a unique label, such as *three*, that represents the quantity of that set. This is called the cardinality principle (Gelman & Gallistel, 1978). Second, students must then use the label while reasoning about the set of objects. However, most early elementary students have not developed the strategy of verbally labeling an object (e.g., a ball) when attempting to remember it (Palmer, 2000). Until about the age of seven, children appear to simply visualize an object when instructed to remember it. If these students have not developed the strategy of verbally labeling an object to facilitate recall, then they may not have developed the strategy of verbally labeling a set of objects with the numerical name in order to facilitate reasoning about the set. As a result, young children would have to rely on a mental replication of the objects in the set to solve quantitative reasoning tasks.

*Problem Statement*

To summarize, children's quantitative reasoning primarily involves reasoning about quantities that can be represented by number words, Arabic numerals, or image-based mental models. However, most current measures of quantitative reasoning abilities rely heavily or exclusively on number words and Arabic numerals. If test takers represent quantity using image-based mental models, then most measures of quantitative reasoning demonstrate construct underrepresentation. This seriously threatens a valid interpretation of test scores in most educational situations. Likewise, the verbal demands of these assessments may introduce construct-irrelevant variance. Therefore, the purpose of this study was to examine how students represent and reason with quantity. This study specifically focused on quantitative reasoning in kindergarten, first, and second grade students because they are least likely to have developed a robust understanding of number words and Arabic numerals. In order to compare early elementary students' quantitative representations to more advanced representations, a smaller comparison sample of fifth grade students was also examined.

# CHAPTER 2

# REVIEW OF THE LITERATURE

## *Developmental Theory of Central Conceptual Structures*

Case, in his neo-Piagetian theory of conceptual development, essentially theorized that children progress through qualitatively distinct stages of reasoning in which thinking becomes more systematic (Case, 1978). A distinct central conceptual structure characterizes each stage. Central conceptual structure refers to an internal mental network of concepts used to represent and assign meaning to situations and problems (Case & Griffin, 1990). As their conceptual structures mature, children become more effective problem solvers because their conceptual structures enable them to integrate more aspects of the problem. Two key propositions of Case’s theory are that central conceptual structures (a) affect a broad range of problems within a particular domain and (b) provide the foundation on which more complex concepts are built (Griffin, 2004).

Therefore, children at different levels of development mentally represent the conditions of the same problem in fundamentally different ways. The mental representation of a problem then dictates the strategies that the child will use to find a solution. As a result, the central conceptual structure is fundamental to Case’s theory because developed central conceptual structures limit children’s cognition. Although Case’s theory of cognitive development has been empirically supported in studies on social thought, narrative, and quantitative reasoning (Case & Okamoto, 1996) the focus here was on quantitative reasoning.

According to Case's theory, children from preschool through elementary school are generally in one of four levels of the *dimensional* stage. In the *predimensional* level beginning around four years of age, children have two functionally independent conceptions of quantity: the ability to count and the ability to make nonnumerical judgments of quantity (Case & Okamoto, 1996).

Most children have learned to count small sets of objects by the age of three (Fuson, 1988; Gelman & Gallistel, 1978; Siegler & Robinson, 1982). The ability to count requires three fundamental capabilities (Gelman & Meck, 1983). First, the one-to-one principle stipulates that each object should be tagged with only one verbal label. Second, successful counting involves recognizing that the tags used to label the objects must be stated in a constant order. This is called the stable order principle. Finally, the cardinality principle requires understanding that the final tag in the series is the formal name for the numerosity of the set of objects. Children tend to apply the one-to-one and stable order principles around two and one-half years of age (Gelman, 1978). However, the ability to apply the cardinality principle develops later and involves four stages of development. First, when asked how many objects are in a set, children simply repeat the last number in the counting sequence. Later, children understand that they will end at the same cardinal number across repeated counts. Next, children can determine the cardinality of a set without counting by matching the set with another set of known cardinality via one-to-one correspondence. Finally, children develop the ability to reason using a number word without having to count a concrete set of objects.

In addition to a verbal conception of quantity, preschoolers also have a variety of nonnumerical quantitative abilities. Most preschoolers understand that adding an object to a set causes the amount of the set to increase, and vice versa for subtraction (Starkey, 1992). Preschoolers can also compare sets of objects to determine which set has more and which set has less (Barth, La Mont, Lipton, & Spelke, 2005; Huntley-Fenner & Cannon, 2000; Siegel, 1974). When preschoolers compare sets and make mathematical transformations, they do not appear to use their counting abilities to solve the task (Huntley-Fenner & Cannon, 2000; Starkey, 1992).

Even though preschoolers successfully solve quantitative tasks that require either the exclusive use of verbal counting or nonnumerical quantitative skills, they appear incapable of solving tasks that require both of these abilities at the same time (Resnick, 1989; Siegler & Robinson, 1982). For example, preschoolers have difficulty answering the question "Which is bigger, 9 or 5?" These results demonstrate that preschoolers have developed separate quantitative central conceptual structures. One of these structures consists of representing a problem in terms of a verbal counting representation and the other consists of representing a problem in terms of a nonnumerical quantitative representation (Case & Okamoto, 1996). A study by Wang, Resnick, and Boozer (1971) supports this interpretation. Wang et al. (1971) administered a broad range of tasks to 78 kindergarteners that assessed the ability to count, use numerals, and compare set sizes. They found that the ability to count and use numerals were dependent on each other, but the ability to compare set sizes developed independently from counting and using numerals.

As preschoolers transition into the next stage of development, they merge their verbal counting structure and their nonnumerical structures into a new conceptual structure of a mental number line with four different components (see Figure 1; Case &

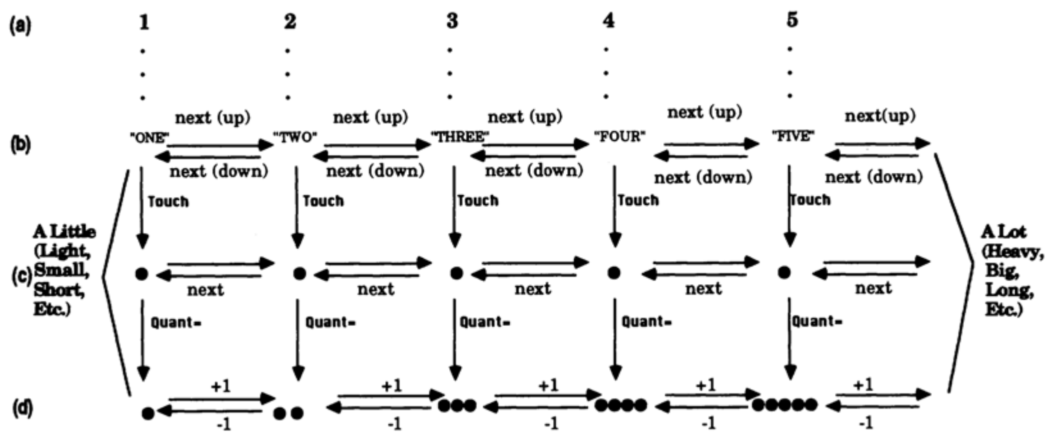



Figure 1. The central numerical structure (the *mental number line*) hypothesized to emerge around 6 years. The four rows indicate, respectively, *(a)* knowledge of written numerals, *(b)* knowledge of number words, *(c)* a pointing routine for *tagging* objects while counting, and *(d)* knowledge of cardinal set values. The vertical arrows indicate the knowledge that each row maps conceptually onto the next; the horizontal arrows indicate an understanding of the relation between adjacent items. The external brackets indicate the knowledge that the entire structure can be used as a vehicle for determining the relative amount of quantities composed of identical units (weight, height, length, etc.).

Source: Case & Okamoto (1996, p. 7).

Okamoto, 1996). The *verbal labeling* line (row b in Figure 1) consists of the ability to recognize and generate verbal labels for quantity (e.g., *one, two, three*). The *mental action* line (row c in Figure 1) indicates that children can procedurally tag objects as they state the number words using the one-to-one principle. Children progress from only tagging physical objects to tagging objects that are mentally represented (Case & Okamoto, 1996). The *conceptual interpretation* line (row d in Figure 1) signifies that each verbal tag represents a set that contains that specific quantity of objects. Finally, children recognize Arabic numerals (row a in Figure 1) that are grafted onto the three more basic number lines.

In addition to the four mental number lines, children must also understand associations between them (Case & Okamoto, 1996). First, they must recognize that each mental number line maps directly onto the other number lines in a one-to-one fashion. Secondly, children must understand that moving from one place on a number line to the next involves addition or subtraction of one unit. Finally, children need to know that movement along one mental number line is necessarily accompanied by the same movement along the other mental number lines.

At the age of six, the *unidimensional* level, children's conceptual structure of quantity consists of mental objects that are mentally manipulated (Okamoto, 1996). In this stage, children can represent a situation using only one mental number line. By age eight, the *bidimensional* level, children have mastered the use of the mental number line and so they can begin relating two number lines to each other (Case & Okamoto, 1996). Because of this ability to use multiple number lines, children are able to go beyond thinking about mental objects and can reason using numbers as symbols for quantitative sets (Okamoto, 1996). At ten years of age, the *integrated bidimensional* level, children are capable of explicitly relating two mental number lines and generalizing the relationship to an entire number system (Case & Okamoto, 1996).

To summarize, in the *predimensional* level, children possess two separate quantitative structures that allow them to interpret problems in terms of a counting representation or a nonnumerical representation. These two systems then merge and children begin to construct a system of symbolic relations that allow them to map a set of real objects to number words and Arabic numerals via the mental number line. At this level, children interpret quantitative tasks using a single array of mental objects. From ages six through ten, children attain complete mastery of the mental number line, allowing them to coordinate multiple number lines.

*Empirical evidence to support central conceptual structures.* Support for the theory of central conceptual structures has come from research in which a group of children completed a range of tasks that assessed each level of conceptual development. Analyses test whether performance on the tasks followed the hypothesized pattern of conceptual structures (Case & Okamoto, 1996). The theory of central conceptual structures was confirmed if participants succeeded on both tasks that defined a particular conceptual structure and the tasks at lower conceptual structures, but failed at tasks that defined higher conceptual structures.

Tasks that assessed the first quantitative conceptual structure did not require precise quantification skills, but instead could be answered by general polar classifications such as evaluations of more or less. Only one dimension of the task had to be accurately quantified on tasks that assessed the second conceptual structure. The third conceptual structure was assessed by two dimensions that had to be quantified precisely but did not need to be precisely related, and the fourth conceptual structure was assessed by precisely quantifying and relating two dimensions.

Following this paradigm, Okamoto (1996) conducted a study to determine whether conceptual structures adequately described children's performance on quantitative word problems. Previous research has established that quantitative word problems vary in difficulty even when the basic mathematical operations were the same (Arendasy & Sommer, 2005; Carpenter & Moser, 1984; Hudson, 1983; Riley & Greeno, 1988). Because of the differences in difficulty, the linguistic structure of a word problem has been hypothesized to afford a specific mental representation. The mental representation in turn required more or less advanced quantitative knowledge for successfully solving the problem (Kintsch, 1988). In other words, the description of the quantitative scenario in the word problem elicited a particular representation of the problem. Performance on the problem then depended on whether the child's conceptual structure could accommodate the quantitative representation that the linguistic structure afforded (Okamoto, 1996).

Since six year old children were hypothesized to have a conceptual structure that consisted of a single mental line of objects, they would only be able to solve word problems that afforded a representation using one dimension of mental objects. For example, "Joe had six marbles. Then he gave Tom two marbles. How many marbles does Joe have now?" On this item, children could mentally represent an array of six marbles, then remove two marbles and count the number of marbles left.

Eight year olds were hypothesized to have a conceptual structure that enabled them to compare two mental number lines using numbers to symbolize mental objects. At this level, children should be able to solve the problem "Joe has two marbles. Tom has six marbles. How many more marbles does Tom have than Joe?" At this stage, students could create one number line to represent Joe's two marbles and a separate number line that represented Tom's six marbles. Comparison of the two number lines would allow the student to determine that four numbers come between two and six.

Ten year olds were hypothesized to have a conceptual structure that could use two mental number lines that are well integrated. These children could therefore reverse operations that eight year olds could perform. Ten year olds could successfully solve problems such as “Joe has six marbles. Joe has two more marbles than Tom. How many marbles does Tom have?” In order to determine how many marbles Tom has, the phrase “Joe has two more marbles than Tom” must be reversed to “Tom has two fewer marbles than Joe.” A student could then calculate *2 fewer than 6*. Therefore, with the sophisticated *dimensional* conceptual structure, children could create a mental number line for Joe’s six marbles. A second mental number line represented the phrase *two more*. This must be reversed to represent “two fewer.” By comparing the first mental number line and a reverse of the second mental number line, the student could arrive at the solution.

To conduct the study, Okamoto administered sixteen word problems representing three levels of conceptual structures to children in kindergarten through fourth grade. The experimenter read aloud the items and presented a representation of the items on a card. In kindergarten through second grade, the cards contained pictures that represented the word problems while the text was printed on cards for grades three and four.

As expected, the proportions of correct responses for all items within a particular level of conceptual structure were very similar. Between levels of conceptual structures, however, there was a significant decrease in proportion correct. The mean proportion of correct responses for the three level 1 items was .906 (range from .867 to .950), the mean of the nine level 2 items was .600 (range from .467 to .683), and the mean of the four level 3 items was .267 (range from .233 to .300). Cluster analysis and latent structure analysis also generally confirmed that the word problems that were hypothesized to measure each conceptual structure did in fact cluster together to represent separate structures. Examination of individual children’s correct and incorrect responses revealed that all sixty children in the sample conformed to the hypothesized pattern of responses whereby level 1 knowledge was prerequisite to level 2 knowledge that was then prerequisite for level 3 knowledge.

Case and Okamoto (1996) also conducted a similar study with additional items that were hypothesized to measure four levels of quantitative conceptual structures, including the *predimensional* structure represented by two separate counting and nonnumerical structures. Level 1, *predimensional thought*, assessed the conceptual structure that precedes the mental number line, such as “Which of these two piles of chips has more?” Level 2, *unidimensional thought*, was measured by items that tested for the presence of the mental number line, such as “What number comes after seven?” At level 3, *bidimensional thought*, children were hypothesized to be capable of integrating two mental number lines. This could be reflected by understanding the relations between the ones and tens column of a base-ten number system because the ones and tens column are essentially separate number lines. A sample item at level 3 was “What number comes 5 numbers after 49?” Level 4, *integrated bidimensional thought*, could be reflected by understanding the relationship between multiple columns in the base-ten system. This level was measured by items such as “What number comes 10 numbers after 99?”

Kindergarten through fourth grade children completed a battery of similar items. Results supported the presence of the different conceptual structures because the proportions of correct responses were comparable within level, but decreased as the levels increased. Similar studies have also been conducted with items that reflect the various levels of conceptual structures in the domains of time and handling money (e.g., Griffin et al., 1992).

A different experimental paradigm to test the theory of central conceptual structures consisted of giving the same task to children at a range of ages. Experimenters then coded children's answers and justifications based on the characteristics of the problem that they took into account when solving the task (Marini, 1992). For example, students were asked to predict which side of a balance beam would go down when different numbers of weights were placed in different locations along the beam. Reponses to the task were coded based on whether they only focused on global aspects of the task *(predimensional),* one dimension *(unidimensional),* two dimensions *(bidimensional),* or two dimensions and the variation in each *(integrated bidimensional).* Additional tasks from this paradigm have consisted of a conceptually similar task where students made predictions of shadow projections, as well as three different proportional reasoning tasks such as determining which mixture of juice would be more concentrated when various amounts of juice and water were combined (Marini, 1992; Marini & Case, 1994). The results of these studies demonstrated that children within age a particular age group tended to focus on similar characteristics of the problem, while children across age groups tended to take into account more conditions of the problem. Each child also produced similar responses across tasks.

These studies by Case, Okamoto, and colleagues showed that quantitative tasks afforded different problem representations based on the characteristics of the task that children used to solve the problem. Children were successful on tasks with demands that could be represented by their conceptual structures. However, children were unsuccessful on the tasks where the complexity exceeded the capacity of their conceptual structures.

Furthermore, less advanced conceptual structures were prerequisite to more sophisticated conceptual structures.

Central conceptual structures were additionally hypothesized to affect children's interpretation of diverse problems within the domain. To test this hypothesis, children's conceptual structures were experimentally manipulated by training children in the next advanced conceptual structure. If conceptual structures did indeed influence a wide range of tasks, then instruction in a more complex conceptual structure would transfer to unrelated tasks within the domain.

To test this hypothesis, Case and Sandieson (1992) assigned junior kindergarten students (the first of two years of kindergarten; mean age of 4.9 years) in Canada to either a treatment group that received instruction in the mental number line, which was not hypothesized to develop until six years of age, or to a control group that received instruction in the letters of the alphabet. The treatment consisted of instruction on reciting the number sequence, counting objects, adding and subtracting one unit from a set, deciding the larger of two numbers, and using numbers to compare sets of objects. Both the treatment and control groups were pre- and post-tested on tasks that required near transfer (comparing amounts of money), intermediate transfer (solving a basic proportional reasoning task), and remote transfer (comparing amounts of money that were misleading such as quarters and dimes, analyzing the passage of time on a clock, and determining the amount of blocks after a large amount of blocks had been added to a set).

Results showed that at the pretest, most of the children in both groups failed all of the transfer tasks. After training, children in the control group still failed the transfer tasks. However, most of the children who had mastered the training activities in the treatment condition were successful on all of the transfer tasks. Since instructional training in the next level of conceptual structure influenced performance on tasks that were only distally related to the instruction, Case and Sandieson (1992) concluded that a quantitative central conceptual structure did indeed influence a wide variety of problems within the domain.

Additional training studies were conducted with low and middle socioeconomic status students in kindergarten (Griffin, Case, & Siegler, 1994). In these studies, the treatment and control groups were matched on number knowledge and cultural background. Griffin and colleagues found results similar to those of Case and Sandieson (1992) on measures of transfer. In addition, both the treatment and control groups were assessed on quantitative tasks one year later in first grade. The treatment group outperformed the matched control group on measures of oral arithmetic, written arithmetic, and word problems.

In conclusion, Case and his colleagues provided evidence that children interpreted tasks in a wide range of quantitative situations according to central conceptual structures that helped children organize their thought about a problem. As children developed, their central conceptual structures became fundamentally altered, producing qualitatively different conceptual structures that could represent more complex problems.

*Descriptive Theory of Mental Models*

Huttenlocher and colleagues developed a mental model theory of quantitative reasoning to describe preschoolers' performance on quantitative tasks. According to Huttenlocher and colleagues, children construct a mental representation of the features of a situation that are critical for quantitative reasoning (Mix, Huttenlocher, & Levine, 2002). In this process, a mental image of the quantitative task is created by imagining each discrete entity and then visualizing any quantitative transformation that may be completed on the set, such as adding or removing objects (Mix et al., 2002). These mental models are inherently symbolic because the abstract mental images represent physical objects and exclude information irrelevant to the quantitative task. Consequently, mental models are hypothesized to develop in early childhood during the same time that other symbolic activities develop, such as language and pretend play (Huttenlocher, Jordan, & Levine, 1994).

Three steps are necessary when using a mental model to perform basic calculation transformations (Mix et al., 2002). First, a mental-image of each unit in a set must be created. To transform the set, the child must recall the image of the original set, the amount by which the set must be transformed, and the direction of the transformation. The transformation is executed by imagining mental units either being added to or taken away from the original set (Huttenlocher et al., 1994). Lastly, working memory maintains the image of the final set while the solution is communicated.

The use of mental models of objects to solve quantitative tasks has obvious limitations. First, memory capacity limits reasoning because each image has to be held in mind at the same, so mental models are only useful for small set sizes. The large number of steps necessary to reason using a mental model also affords many opportunities for error, such as misrepresenting the initial set, completing an inaccurate transformation, and incorrectly communicating the final quantity when an answer is required.

These potential sources of error can be reduced by using number words or Arabic numerals as a substitute for the mental image of a quantitative set. Verbal representations of quantity can reduce memory burdens during quantitative transformations so the entire problem does not have to be reconstructed in memory (Mix et al., 2002). Using either number words or Arabic numerals also greatly increases the quantitative operations that can be performed, both by expanding the range of numbers that can transformed and by allowing for more complex mathematical functions such as algebra and statistical analysis.

Mix et al. (2002) proposed that mental models of objects act as the bridge that unites nonnumerical quantitative reasoning skills to the conventional symbol system of mathematics that utilizes number words, Arabic numerals, and other symbols for mathematical operations. Accordingly, mental-images of quantity may be used as conceptual referents for number words and other mathematical symbols. As children become more proficient relating verbal number words to the associated mental images, they gradually depend more on the verbal number words while reasoning than the mental images.

*Empirical evidence to support mental models.* Support for Huttenlocher and colleagues' mental model theory of quantitative reasoning has come from studies comparing the difficulty of quantitative tasks that are presented using different stimuli. Levine, Jordan, and Huttenlocher (1992) examined quantitative reasoning in children ages four to six years. The older children were kindergarteners who had some formal instruction in calculation, whereas the younger children were preschoolers and had no formal arithmetic instruction. The students completed three tasks, a nonverbal quantitative task, verbal story problems, and traditional number-fact problems.

In the nonverbal condition, the experimenter displayed a set of chips in full view of the child. These chips were subsequently hidden by a box. Next, the experimenter placed an additional set of chips one-by-one into the box. To solve the task, the child used their own set of chips to display the final set of chips hidden by the box. The two verbal conditions had equivalent mathematical requirements but without concrete objects as in the nonverbal condition. The story problems consisted of items such as “Mike had one ball. He got two more. How many balls did he have altogether?” Number-fact items simply stated “How much is one and two?” The experimenter read aloud both story problems and number-facts problems and children responded with the correct number word. None of the items included numerosities greater than six.

Since the chips afforded envisioning the quantitative transformation, the nonverbal quantitative task was hypothesized to be the easiest if children represented quantity using a mental-image representation. The number-fact problems, on the other hand, were hypothesized to be easiest if children used a verbal representation of quantity because the number words did not afford visualizing the quantity that the number words represent. Although the story problems used number words, the addition of a concrete referent (balls) was hypothesized to potentially afford picturing the number of balls.

An analysis of variance (ANOVA) showed a significant main effect of problem type. Tukey HSD follow-up tests confirmed that nonverbal items were significantly easier than story problems, and story problems were easier than number-fact problems. These results demonstrated that young children performed better on a task that afforded using a mental-image representation of quantity. However, the standard errors for all three conditions in the 6.0 to 6.5 age group appeared to overlap in the plots of the mean scores, particularly for the addition items. Since the authors did not report whether they tested the task by age interaction, it is not certain whether the three conditions were significantly different for the oldest age group. This issue was particularly important since children in the two oldest age groups had formal instruction in arithmetic that most likely consisted of instruction in number facts.

The authors did report a significant interaction between problem type and numerosities. For the nonverbal task and story problems, items with numerosities of five or six were significantly more difficult than items with smaller numbers. However, there was no significant difference between small and large numbers on number-fact problems. Therefore, numerosity did not appear to affect performance if the task used a verbal representation. On the other hand, the size of the set did affect performance if the task afforded a mental-image representation where reasoning would be limited by the amount of objects that children could hold in memory at one time.

The study additional found that the older children used their fingers as a strategy most frequently for number-fact problems, with an intermediate frequency for story problems, and least frequently on the nonverbal task. These students were most likely attempting to relate the verbal demands of the task to the mental image strategy by using their fingers as physical objects on the tasks that did not provide concrete referents. In other words, the older children might have still needed to visualize the amount of each quantity in order to complete the task, but since the task did not afford mentally imagining objects, the children used their fingers instead.

Hughes (1981) also found similar results using a wider range of tasks in a sample of three to five year olds. In his study, children completed five quantitative tasks that varied on a continuum from concrete to abstract. In the most concrete task, box open, children viewed an open box in which bricks were either added or subtracted from the original set. Children could see the final set while answering. In the box closed condition, the children saw the original set but the box was closed before the experimenter performed the transformation, a condition similar to the nonverbal task used by Levine et al. (1992). In the hypothetical box condition, the experimenter read aloud items such as "If there was one brick in a box and two more were put in, how many would there be altogether?" The hypothetical shop condition was similar to the hypothetical box condition, but items stated "If there was one child in a sweetshop and two more went in, how many were in the sweetshop altogether?" Finally, the formal code condition consisted of items like Levine et al.'s (1992) number-fact problems, such as "What does one and two make?" Children answered all tasks with the number word for the final set.

An ANOVA found a main effect for the task, with abstract tasks significantly more difficult than concrete tasks. A significant interaction between task and set size was also reported. Tukey HSD follow-up tests confirmed that with small numbers less than five, the open and closed box tasks were not significantly different from each other and the two hypothetical tasks (box and shop) were not significantly different from each other. Therefore, the open and closed box tasks were easiest, the hypothetical box and shop tasks were of medium difficulty, and the formal code task was the most difficult. Since young children performed equally well on a task that required memory recall of the final set (box closed) and one that did not require memory recall (box open), memory did not appear to inhibit young children's performance on concrete items that included sets of five or less. With large numbers, however, the open and closed box tasks were significantly different from each other. Furthermore, performance in the closed box, hypothetical box, and hypothetical shop tasks was not significantly different. In other words, the open box task was the easiest; the closed box, hypothetical box, and hypothetical shop tasks were of medium difficulty; and the formal code task was the most difficult. Since the boxed-closed task was equally as difficult as the hypothetical tasks for large numbers, memory did appear to inhibit performance on concrete quantitative tasks with numbers greater than five for preschool children.

Using a different research paradigm, Starkey (1992) found evidence that young children could solve quantitative tasks without using verbal counting skills. In this experiment, children ages two through four saw a set amount of table tennis balls. Then the children placed each ball one at a time into a searchbox that hid the ball from view. Once the child had placed the last ball into the searchbox, the child watched the experimenter either add or remove balls from the searchbox. The experimenter then instructed child to take all of balls out of the searchbox. To prevent the child from feeling how many balls were left in the box, a hidden apparatus removed balls so only one ball was present in the searchbox at a time. Starkey found that children even as young as two years old could remove the exact amount of balls from the searchbox on problems that involved numerosities of three or less. However, performance on items that involved numerosities four or greater was not significantly better than chance even for four year olds.

Starkey asserted that these children did not solve the task using verbal counting procedures. First, the children rarely demonstrated overt counting procedures or nonverbal motor behaviors that preschool children tend to use when verbally counting. Moreover, even children who were unable to count, as evidenced both by parental report and performance on a counting task in the experiment, were successful in the searchbox task. Starkey therefore concluded that some early numerical reasoning abilities do not depend on the ability to count. Similar to Huttenlocher and colleagues, Starkey suggested that the children may have used mental imagery in order to create a numerically accurate representation of the original set with objects subsequently added or removed. However, Starkey cautioned that the mental imagery process would only be effective for small set sizes.

Using a paradigm similar to Levine et al. (1992), Jordan, Huttenlocher, and Levine (1994) examined the difficulty of calculation tasks for three to five year olds when children responded nonverbally, verbally, or simply recognized a nonverbal solution. The nonverbal condition was identical to the nonverbal task used by Levine et al. (1992) where children produced the number of chips hidden in a box after a transformation. The verbal condition used the same procedures as the nonverbal condition, but children responded with the appropriate number word after the transformation. In the nonverbal recognition condition, children chose the correct answer from among four options presented on an index card. In a sample of middle-income children, who tend to be more skilled in language, there were no differences in performance among the three response conditions. Jordan et al. (1994) concluded that young middle-income children could answer nonverbal calculation problems both verbally and nonverbally.

Jordan et al. (1994) extended their study to a sample of lower-income children who attended a Headstart program. Previous research indicated that lower-income students tend to perform more poorly than middle-income children on quantitative tasks that use verbal representations of quantity. If young children solved nonverbal quantitative tasks using a verbal strategy, then the sample of lower-income children would perform worse than middle-income children on all three conditions in the study. However, if young children solved nonverbal calculation tasks using an image-based mental model strategy, then the lower-income children would perform just as well as their middle-class peers in the nonverbal and nonverbal recognition conditions. In the lower-class sample, there was a main effect of response type whereby the verbal task was more difficult than both the nonverbal production and recognition task. Furthermore, the lower-income children performed as well as the middle-income children in the two nonverbal conditions, but significantly worse than the middle-income children in the verbal condition. These results provided additional evidence that young children could reason quantitatively without using verbal representations of quantity.

Interestingly, Jordan and colleagues reported an age by numerosity interaction. Performance on tasks with numbers 4 or less increased steadily from age 3 (with a mean of 2 of 6 items correct) through age 5 (with a mean of 5 of 6 items correct). However, on tasks with numerosities of 5 and above, performance remained poor with a mean of about 2 items correct until children reached 5.6 years old, when the mean increased to 4 items correct. Furthermore, children performed better on small number tasks that required a transformation than on a task with larger numbers that simply required children to recall the numerosity of the original set without a transformation. Whereas transformations on the small sets could be visualized using a mental-image representation, the mental-image representation did not even facilitate recall of large set sizes. When the size of the set inhibited the ability to visualize the set of objects, young children's performance on the quantitative tasks suffered.

In a similar study, Jordan, Huttenocher, and Levine (1992) assessed kindergarten children from middle and low-income families on a range of quantitative tasks. In addition to the nonverbal task, story problems, and number-fact problems used by Levine et al. (1992), a fourth condition consisted of word problems, such as "How much is one and two pennies?" This condition was designed to have similar referents as the story problems but with more decontextualized language.

Results confirmed a main effect of problem type. Follow-up analyses found that the nonverbal problems were significantly easier than all three verbal problem types. For addition items, there was no significant difference between the three verbal tasks. For subtraction items, story problems were significantly easier than number-fact items. A significant interaction of income level and problem type was also reported. Follow-up tests found a significant effect of income level on story problems, word problems, and number-fact problems, but not on the nonverbal problems. Therefore, the lower-income students performed worse than the middle-income students in all of the verbal conditions, but performed as well as the middle-income students in the nonverbal condition that afforded a mental-image representation. To determine whether the differences between income levels on the verbal items could be attributed to linguistic factors, an analysis of covariance (ANCOVA) was conducted using the verbal subtest of the Primary Test of Cognitive Skills as a covariate. The difference in performance between lower and middle class students was not significant for either story problems or word problems after verbal skills were statistically controlled. For number-facts, the difference between income levels was reduced but still significant.

To summarize, the mental model paradigm of quantitative reasoning reported that preschool children perform best on nonverbal measures of quantitative reasoning. Specifically, Huttenlocher et al. (1994) provided evidence that by 30 months of age, children had the ability to perform quantitative transformations on sets of concrete objects. However, children did not become equally proficient in solving story problems and number-fact problems with comparable mathematical demands until age five (Levine et al., 1992). Skill in solving nonverbal quantitative tasks therefore developed before skill in solving similar verbal quantitative tasks. Young children also performed better on story problems than on number-fact problems. This additionally supported the hypothesis that young children reason more effectively with conceptual referents that provide meaning to the quantities (Jordan et al., 1992). In other words, the number-fact problems were more difficult because the exclusive use of number words did not afford mentally visualizing a discrete set of objects that would be necessary for using an image-based mental model. On the other hand, the referents in the story problems (e.g. “two balls”) afforded envisioning the objects that enabled children to use a mental-image representation when solving the quantitative task.

Furthermore, although lower-income children performed just as well as middle-income children on nonverbal measures of quantitative reasoning, they performed significantly worse on verbal measures of quantitative reasoning or nonverbal tasks that required a verbal response. Since middle-income children tend to have stronger verbal skills and since the differences in performance on quantitative tasks were reduced or disappeared when verbal factors were taken into account, Huttenlocher and colleages concluded that verbal processes were most likely secondary to quantitative reasoning for young children (Huttenlocher et al., 1994). Likewise, other researchers have found that young children could solve quantitative tasks even when they had not developed the ability to count (Mix, 1999; Starkey, 1992).

To interpret these findings, Huttenlocher et al. (1994) and Mix et al. (2002) proposed a mental model theory of quantitative reasoning. According to this theory, young children reason quantitatively by constructing a mental representation of each discrete unit in a set. Quantitative transformations are then carried out by envisioning objects being added to or taken away from that set (Huttenlocher et al., 1994). As children develop their quantitative reasoning abilities, verbal counting skills merge with this mental model of quantitative reasoning (Mix et al., 2002).

*Comparison of Conceptual Structures and Mental Models*

In order to explain how children of various ages solved different types of problems within a domain, Case proposed a developmental theory of reasoning suggesting that children's reasoning qualitatively changes based on transformations to a central conceptual structure. According to his theory, the developed conceptual structure affects how a child will represent a problem that then influences the strategy that the child will choose to find a solution. The conceptual structure for quantitative reasoning develops from separate verbal counting and nonnumerical structures around age four to a mental counting line at age six to a completely integrated understanding of multiple counting lines by age ten. To summarize, Case proposed that a central conceptual structure influences how a problem is represented which then guides performance on virtually all tasks within a domain. On the other hand, Huttenlocher and colleagues' mental model theory described the patterns of preschoolers' performance on quantitative tasks. The theory hypothesized that preschoolers reason on quantitative tasks by constructing a mental representation of the quantity in a set.

Huttenlocher and colleagues' mental model theory complement Case's theory for understanding children's quantitative reasoning. Whereas Case applied a macroscopic lens to explain the development of reasoning across ages, Huttenlocher and colleagues applied a microscopic lens to explain quantitative reasoning in preschool children. Hence, Huttenlocher and colleagues' mental model theory provided a more detailed account of the nonnumerical central conceptual structure present in the *predimensional* level of four year olds. Case's theory has a wider application by explaining how the quantitative conceptual structure develops to an advanced structure that can solve complex quantitative tasks. Although Case and his colleagues have empirically established that the *unidimensional* conceptual structure of six year olds differs from the *bidimensional* conceptual structure of eight year olds and that this conceptual structure broadly affects a range of quantitative tasks, a more detailed account of the *unidimensional* conceptual structure was warranted.

### *Research Questions*

Additional research needed to be conducted to examine the conceptual structure that students from kindergarten through second grade use to represent quantitative reasoning tasks. Specifically, previous research has established that preschool children have two separate conceptual structures for interpreting quantity: a verbal counting structure and a mental-image structure. Case proposed that these two structures begin to merge in the early elementary grades. However, very little research specifically examined this process. Since most measures of quantitative reasoning for early elementary students assume that these two structures have merged and students use primarily a verbal structure of quantity, this research question is of great import.

The overall research question addressed in this study was, "What is the conceptual structure that kindergarten, first, and second grade students use on quantitative reasoning tasks?" I hypothesized that kindergarteners would tend to have distinct verbal and mental-image structures of quantity. Therefore, on most quantitative reasoning tasks they would tend to use their mental-image structure of quantity because it would enable them to make a direct evaluation of the quantities. First graders were hypothesized to have a more integrated structure of quantity where they would tend to associate verbal labels of quantity with a mental-image representation. However, this association still would be underdeveloped. Second graders were hypothesized to have successfully merged the two structures of quantity. Since the verbal representation of quantity is considerably more prevalent in educational contexts, second graders would tend to use a verbal representation of quantity.

This research question was addressed by five more specific research questions.

1. Do kindergarten, first, and second grade students perform better on a pictorial task of quantitative reasoning or on a matched verbal task of

quantitative reasoning? Is there an interaction with grade on performance in verbal and pictorial representations?

2. Can kindergarten, first, and second grade students move fluidly between pictorial and verbal representations of quantity? Is there an interaction with grade on whether students can move fluidly between representations of quantity?
3. Do kindergarten, first, and second grade students choose to use a verbal or a pictorial representation of quantity when solving a quantitative reasoning task? Is there an interaction with grade on choice of verbal or pictorial representations?
4. Do students exhibit a similar structure of quantity across conditions? Operationally, do students' choices of formats match their pattern of performance in pictorial and verbal quantitative reasoning tasks?
5. Does the pattern of findings differ when students are categorized by ability instead of grade?

Each of these five research questions concerned one aspect of the quantitative reasoning structure that children develop. When the results from each research question are integrated, a comprehensive account of the quantitative reasoning conceptual structure of early elementary students can be illustrated. These five questions were addressed by performance on two different quantitative reasoning tasks. The first task, Equivalence, required students to make equivalent quantities by combining quantitative sets. The second task, Number Series, required students to discern a pattern in a series of quantities and then provide the next quantity that continued the series. These two tasks were administered using verbal representations and pictorial representations of quantity.

The first research question asked whether kindergarten, first, and second grade students performed better on a pictorial task of quantitative reasoning or on a matched verbal task of quantitative reasoning. If students had a stronger verbal structure of quantity like most measures of quantitative reasoning assume, then they would perform better in the verbal condition. However, if students had a stronger mental-image structure of quantity, then they would perform better in the pictorial condition. If their two structures had merged, then they would perform equally as well in both conditions.

This question was answered by performance on both the Equivalence and Number Series tasks. Students completed both tasks using a verbal format with Arabic numerals (Numeral condition) and a pictorial format with a set of objects to represent the quantity (Pictorial condition). A purely verbal quantitative task with number words (e.g., “four”) requires students to remember the quantity. This would introduce memory confounds when compared to a pictorial representation of quantity. Therefore, this study used Arabic numerals to represent number words, assuming that kindergarteners could match the Arabic numeral with the appropriate number word. I hypothesized that kindergarteners would have better performance in the Pictorial condition when compared to the Numeral condition because they would tend to use a mental-image structure of quantity. First graders were hypothesized to have slightly higher performance in the Pictorial condition because they would have developed a more integrated structure of quantity. Second graders were hypothesized to have similar performance in the Numeral and Pictorial conditions because they would have integrated their verbal and mental image structures.

The second research question asked whether kindergarten, first, and second grade students could fluently move between pictorial and verbal representations of quantity. If students had merged their verbal and mental-image structures of quantity, then students would be as successful on a quantitative reasoning task using a combination of verbal and pictorial representations as they would be on a task exclusively using one representation. However, if students had not merged their verbal and mental-image structures of quantity, then performance would significantly decrease when solving a quantitative reasoning task with a combination of representations when compared to either representation alone.

A third condition in the Number Series task addressed this research question. In addition to the two baseline conditions (Numeral and Pictorial), an additional condition integrated both verbal and pictorial representations (Mixed condition). The two middle quantities in the series were pictures whereas the rest of the quantities were Arabic numerals. For example, on the item *1, 2, 3, 4*, the numbers two and three were represented with pictures and the numbers one and four with Arabic numerals. I hypothesized that scores on the Mixed condition would significantly decrease when compared to the Numeral and Pictorial conditions for kindergarteners, slightly decrease for first graders, and would not decrease for second graders.

The third research question asked whether kindergarten, first, and second grade students preferred to use a verbal or a pictorial representation of quantity when solving a quantitative reasoning task. Presumably, students would choose to use the representation of quantity that more directly corresponded to the conceptual structure that they used to solve the task. Therefore, if students had a stronger mental-image structure of quantity, then they would choose to use the pictorial representation more often than the verbal representation. On the other hand, if students had a stronger verbal structure of quantity, then they would choose to use a verbal representation in a quantitative reasoning task. If students had merged their two structures, then they would most likely choose a verbal representation because that is the most common representation in an educational setting.

The third research question was addressed by the Equivalence task. In addition to the Numeral and Pictorial conditions, a third condition gave students the option of using either a verbal or pictorial representation to solve the task (Choice condition). I hypothesized that kindergarteners would choose to use the pictorial representation because they would be using a mental-image structure. First graders were hypothesized to equally choose the verbal and the pictorial representations because their structures would be beginning to merge. Second graders were hypothesized to prefer using a verbal representation because they would have successfully integrated the two structures.

Since conceptual structures were hypothesized to be the characteristic way that students represent quantitative tasks, the fourth research question examined whether students demonstrated a similar structure of quantity across conditions. If students had a stronger mental-image structure of quantity, then they would perform better on all pictorial tasks and choose the pictorial representation. Likewise, if students had a stronger verbal structure of quantity, then they would perform better on all verbal tasks and choose the verbal representation. If students had merged their structures, then they would perform similarly across conditions and choose the verbal representation since it is most frequently used in educational contexts.

To address the fourth research question, students were compared based on the format that they chose most frequently in the Choice condition for the Equivalence task. Performance in the Pictorial and Numeral conditions of both the Equivalence and Number Series tasks was then compared across the groups. Students who preferred a pictorial representation of quantity were hypothesized to have separate verbal and mental-image structures of quantity. Consequently, these students were hypothesized to perform better in both Pictorial conditions. Students who preferred the verbal representation of quantity were hypothesized to have merged their verbal and mental-image structures of quantity, so they were hypothesized to perform similarly in both conditions.

The fifth research question compared the conceptual structures that emerged by grade to the conceptual structures that emerged when students were grouped by ability. I hypothesized that the pattern of performance in the Numeral and Pictorial conditions would be similar for low, medium, and high ability students as the pattern of performance for kindergarten, first, and second grade. To examine this research question, students were grouped by ability on the complementary task. For example, when examining performance in the Equivalence task, students were divided into ability groups by performance on the Number Series task.

According to Case's theory of central conceptual structures, students in fifth grade have developed a sophisticated understanding of the mental number line. Not only should their verbal and mental-image quantitative structures be integrated, but they should also be able to relate multiple number lines to each other. Therefore, the quantitative structures of kindergarten, first, and second grade students were compared to a smaller sample of fifth grade students in order to examine how early elementary students' performance on quantitative reasoning tasks related to students who had developed a more advanced structure of quantity.

# CHAPTER 3

# METHODS

## *Tasks*

*Equivalence.* The Equivalence task was designed to be similar to the Number Operations Puzzle item format on the *Inview.* The materials for this task consisted of a foam board with two plates on opposite ends of the board and a cookie jar at the top (see Figure A.1). The plate in front of the experimenter had one strip of Velcro and the plate in front of the student had two strips of Velcro. The experimenter presented four cards, each with a representation of a quantity (cookies for the Pictorial condition and Arabic numerals for the Numeral condition). The cards were approximately 5.5 by 2.75 inches for the Pictorial condition and approximately 4.25 by 3.75 inches for the Numeral condition. Each card had a piece of Velcro on the back.

During the experiment, the experimenter placed one of the cards on the plate in front of her. Then the experimenter told the student that they had to put the same amount of cookies on their plate by sticking two cards on their plate and sticking the leftover card into the cookie jar.[1]

*Number Series.* The Number Series task was designed to be similar to the Number Series subtests on the CogAT, OLSAT-8, and WJ III ACH. A foam board was also the basis for the second task. This foam board had one long, horizontal strip of Velcro along the center (see Figure A.2). In this task, the experimenter attached one card, the stem, on the foam board with four to six quantities in a series. The stem card was approximately

---

[1] Occasionally, a student would insist that they only needed to use one card to match the experimenter's quantity. In this case, the experimenter allowed the student to use that one card and put two cards back into the cookie jar.

8.5 by 5.5 inches for the Pictorial and Mixed conditions and 11 by 2.5 inches for the Numeral condition. The student received a set of nine cards with the quantities one through nine, the distracters. The distracters were approximately 2 by 5.5 inches for the Pictorial condition and 4.25 by 3.75 inches for the Numeral and Mixed conditions.

During the task, the experimenter told the student that they were going to play a game that required them to determine what came next. The experimenter instructed the student to attach the quantity that continued the pattern to the Velcro strip at the end of the series.

*Design*

The design of this study consisted of two tasks with three conditions in each task (see Table 1). The Equivalence task had three conditions: Pictorial, Numeral, and Choice. In the Pictorial condition, the quantities were represented by pictures of cookies arranged in sets of five according to the pattern of dots on a domino. In the Numeral condition, students were given cards with Arabic numerals on them and told that those were the

Table 1

Design of the Study

| | Task | |
|---|---|---|
| Condition | Equivalence | Number Series |
| Pictorial | X | X |
| Numeral | X | X |
| Choice | X | |
| Mixed | | X |

numbers of cookies on that card. In the Choice condition, students were given both a set of cards with Arabic numerals and a set of cards with cookies. The experimenter explained that the two sets of cards had the same amounts of cookies. In half of the items in the Choice condition, the cards with Arabic numerals were placed in a set to the left of the cookie jar and the cards with the cookies were placed in a set to the right of the cookie jar. In the other half of the conditions, the locations of the cards were reversed. This order was counterbalanced across students.

The Number Series task also had three conditions: Pictorial, Numeral, and Mixed. In the Pictorial condition, all quantities were represented by a string of beads. The experimenter told the students that they were to figure out how many beads should come on the next string. In the Numeral condition, both the stem and the distracters used Arabic numerals. In the Mixed condition, the first quantity (for a series of four) or first two quantities (for a series of five or six) were Arabic numerals, as well as the last quantity (for a series of four or five) or two (for a series of six). The middle two quantities in the series were strings of beads. The distracters were all Arabic numerals in the Mixed condition.

The two tasks were administered in counterbalanced order with half of the students completing the Equivalence task first and the other half completing the Number Series task first. For both tasks, the Pictorial and Numeral conditions were always administered first with these two conditions administered in counterbalanced order. The order of the Pictorial and Numeral conditions was reversed for the task administered second. The Choice and Mixed conditions were always administered last.

For each task, a set of 8 items per condition was created for a total of 24 items per task. (See Appendix B for item specifications.) In similar studies, the number of items per condition has been 5 (Hughes, 1981), 6 (Levine et al., 1992), and 7 (Jordan et al, 1992; Jordan et al, 1994). Eight items per condition were chosen for this study because the item set could be divided evenly in half. This was important for appropriately counterbalancing the location of cards in the Choice condition. Each set of eight items was designed to be equivalent. The sets of items were counterbalanced across conditions so each item set was approximately equally represented in each condition. Some of the Number Series items were selected from a pilot study for a revision of the Primary Battery of the CogAT.

*Participants*

Participants in this study consisted of students enrolled in Tonganoxie Elementary School, a medium-sized public school district west of Kansas City, Kansas. Most of the students enrolled in this school district were Caucasian of middle socioeconomic status and monolingual English speakers. Table 2 shows the demographic characteristics of the samples at each grade. There were 149 total students who participated in this study: 44 kindergarteners, 52 first graders, 44 second graders, and 9 fifth graders. Some of the students did not complete all of the conditions on both tasks. In particular, a number of kindergarteners did not complete the Choice condition on the Equivalence task because they became restless on this longer task. Moreover, one first grader and one second grader were only able to complete one of the two tasks. Table 3 shows the number of students who completed each condition. Most of the kindergarten students attended school every other day.

Table 2

Demographic Characteristics of the Sample

| | Number | | Age in Years | | | Number |
|---|---|---|---|---|---|---|
| Grade | Male | Female | Mean | Minimum | Maximum | ELL |
| K | 24 | 20 | 5.11 | 5.4 | 6.11 | 1 |
| 1 | 27 | 25 | 6.11 | 6.4 | 7.9 | 0 |
| 2 | 17 | 27 | 7.11 | 6.10 | 8.7 | 2 |
| 5 | 4 | 5 | 10.10 | 10.4 | 11.4 | 1 |
| Total | 72 | 77 | 7.2 | 5.4 | 11.4 | 4 |

Table 3

Number of Students who Completed Each Condition

| | | Equivalence | | | Number Series | | | |
|---|---|---|---|---|---|---|---|---|
| Grade | Total | P | N | C | P | N | M | Listwise |
| K | 44 | 44 | 44 | 38 | 44 | 44 | 42 | 37 |
| 1 | 52 | 52 | 52 | 51 | 51 | 51 | 50 | 49 |
| 2 | 44 | 43 | 43 | 43 | 44 | 44 | 44 | 43 |
| 5 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

Note: P = Pictorial condition; N = Numeral condition; C = Choice condition; M = Mixed condition.

*Procedure*

Data were collected during the month of January. To avoid fatigue effects, the two tasks were individually administered on two separate days, typically within a week of each other. Students completed the tasks in a quiet room at Tonganoxie Elementary School. After the experimenter read the directions for the task (see Appendix C), students were given between two and five practice items to ensure that they understood the task. Most students only completed two practice items on the Equivalence task, while four or five practice items were necessary for the Number Series task. In the first practice item, the experimenter demonstrated the task for the student. In the second through fifth practice items, the experimenter guided the student through the task. The representation of quantity for the practice items was always identical to the representation that was used in the first condition of the task.

Students were given one point for each item that they correctly answered. Omits were scored as incorrect because students would omit a problem if they were not certain of the answer. For items in the Choice condition of the Equivalence task, the experimenter also recorded whether the student used numbers, pictures, or a combination of both.

*Analyses*

Performance across the conditions of the two tasks was analyzed using separate mixed design ANOVAs for each research question. In addition to significance tests, effect sizes were also reported. For ANOVA analyses, partial eta squared ($\eta_p^2$) statistics were calculated to represent the proportion of variance accounted for by the effect (Bakeman, 2006). Partial eta squared is calculated by dividing the sum of squares of the effect by the sum of the sum of squares for the effect plus the sum of squares for the error. For $\eta_p^2$, a large effect size is greater than .35, a medium effect size is greater than .15, and a small effect size is greater than .02 (Cohen, 1992). When comparing means with *t*-tests, Cohen's *d* is the appropriate index for calculating effect sizes. The *d* statistic was estimated by dividing the difference between the means by the pooled standard deviation (Cohen, 1988). Large, medium, and small effect sizes for Cohen's *d* are greater than .80, .50, and .20 respectively (Cohen, 1992).

For follow-up analyses, the Bonferroni correction was used to control for Type I error rates. According to the Bonferroni inequality, the probability that any given set of events occurs is less than or equal to the sum of their independent probabilities (Shaffer, 1995). In other words, the probability of making a Type I error on any set of analyses is equal to the sum of the probabilities of each separate analysis. Consequently, the significance level was adjusted for follow-up tests using the Bonferroni correction ($\alpha/n$ where $n$ is equal to the number of follow-up analyses).

# CHAPTER 4

# RESULTS

The purpose of this research study was to examine the conceptual structures that students in kindergarten through second grade use to represent quantitative reasoning tasks. To this end, performance on quantitative tasks using various combinations of pictures and Arabic numerals was compared. Whereas current measures of quantitative reasoning tend to assume that early elementary students use number words and Arabic numerals to represent and solve quantitative tasks, the hypothesis of this study was that early elementary students have two separate structures of quantity: an image-based mental structure and a verbal counting structure. Therefore, students in kindergarten, first, and second grade completed two quantitative reasoning tasks, Equivalence and Number Series, with both pictures and Arabic numerals. If early elementary students relied on a mental-image based conceptual structure to solve quantitative tasks, then performance on the Pictorial version of the quantitative reasoning tasks would exceed performance on the Numeral version. In addition to the identical Pictorial and Numeral conditions for both tasks, a different third condition was also administered (see Table 1). The Mixed condition of the Number Series task examined performance on items with a combination of pictures and Arabic numerals. The Choice condition on the Equivalence task enabled students to choose to use either pictures or Arabic numerals when solving the task.

Tables 4 and 5 show the means and standard deviations of the number of items correct for each condition on the Equivalence and Number Series tasks, respectively.

Table 4

Mean and Standard Deviation of Items Correct by Grade, Item Set, and Condition for Equivalence Task

| | Set A | Set B | Set C | **Mean** |
|---|---|---|---|---|
| | | Kindergarten (average n=14 per item set)[a] | | |
| Pictorial | 4.80 (1.74) | 4.00 (1.92) | 4.56 (1.71) | **4.48 (1.77)** |
| Numeral | 3.06 (2.41) | 3.40 (2.32) | 3.54 (2.44) | **3.32 (2.34)** |
| Choice | 4.10 (2.81) | 4.81 (2.07) | 5.42 (1.88) | **4.82 (2.23)** |
| | | First Grade (average n=17 per item set)[a] | | |
| Pictorial | 6.78 (1.26) | 6.29 (1.40) | 6.65 (1.32) | **6.58 (1.32)** |
| Numeral | 6.41 (1.62) | 6.50 (1.76) | 6.35 (1.77) | **6.42 (1.68)** |
| Choice | 6.65 (1.32) | 6.76 (1.44) | 7.29 (1.16) | **6.90 (1.32)** |
| | | Second Grade (average n=14 per item set)[a] | | |
| Pictorial | 7.38 (0.65) | 7.47 (0.83) | 7.33 (1.11) | **7.40 (0.88)** |
| Numeral | 7.53 (0.99) | 7.62 (0.51) | 7.67 (0.62) | **7.60 (0.73)** |
| Choice | 7.73 (0.46) | 7.40 (0.99) | 7.92 (0.28) | **7.67 (0.68)** |

Note: Standard deviations are in parentheses. Sets A, B, and C had eight items. All fifth graders received a score of 8 in all conditions, so their data was not included.

[a] The number of participants differed across item sets and conditions for two reasons. First, some students did not complete all conditions. Second, the number of students who participated was not divisible by three, so the number of students assigned to a particular order of item sets differed from one to two students.

Table 5

Mean and Standard Deviation of Items Correct by Grade, Item Set, and Condition for Number Series Task

| | Set A | Set B | Set C | **Mean** |
|---|---|---|---|---|
| | | Kindergarten (average n=14 per item set) [a] | | |
| Pictorial | 2.00 (1.31) | 1.58 (1.38) | 1.82 (1.19) | **1.82 (1.26)** |
| Numeral | 2.29 (1.26) | 2.40 (1.72) | 2.08 (1.24) | **2.27 (1.40)** |
| Mixed | 1.36 (1.03) | 1.18 (1.29) | 1.29 (1.73) | **1.26 (1.36)** |
| | | First Grade (average n=17 per item set) [a] | | |
| Pictorial | 2.53 (1.59) | 2.94 (1.78) | 3.76 (2.33) | **3.08 (1.96)** |
| Numeral | 3.35 (1.93) | 3.59 (1.50) | 4.06 (1.92) | **3.67 (1.79)** |
| Mixed | 3.00 (1.51) | 3.29 (1.99) | 2.76 (1.35) | **3.02 (1.62)** |
| | | Second Grade (average n=15 per item set) [a] | | |
| Pictorial | 3.64 (1.45) | 4.40 (2.16) | 5.33 (2.72) | **4.48 (2.25)** |
| Numeral | 5.40 (2.16) | 4.29 (1.49) | 5.80 (1.66) | **5.18 (1.87)** |
| Mixed | 5.27 (2.02) | 4.60 (1.88) | 3.79 (1.81) | **4.57 (1.96)** |
| | | Fifth Grade (n=3 per item set) | | |
| Pictorial | 5.67 (2.31) | 7.00 (0.00) | 7.67 (0.58) | **6.78 (1.48)** |
| Numeral | 7.67 (0.58) | 6.33 (2.08) | 6.33 (2.08) | **6.78 (1.64)** |
| Mixed | 4.67 (3.06) | 7.67 (0.58) | 6.00 (2.65) | **6.11 (2.42)** |

Note: Standard deviations are in parentheses. Sets A, B, and C had eight items.

[a] The number of participants differed across item sets and conditions for two reasons. First, some students did not complete all conditions. Second, the number of students who participated was not divisible by three, so the number of students assigned to a particular order of item sets differed from one to two students.

Statistical tests confirmed that the three item sets behaved similarly across conditions. These tests simply confirmed the equivalence of item set difficulty. Since the items sets were counterbalanced across conditions, differential item set difficulty would not influence the results of the study. Scores by item set were combined for an overall mean score within each condition.

Previous research has suggested that there may be differences in performance between boys and girls in the mathematical domain (Geary, 1994). However, statistical tests confirmed that there were no sex differences on the two tasks in this study, so the data for males and females were combined for all of the analyses.

The overall research question of “What is the conceptual structure that kindergarten, first, and second grade students use on quantitative reasoning tasks?” was addressed by five more specific research questions. Each research question was addressed in turn.

*Comparison of Arabic Numerals and Pictures*

The main research question examined whether students performed better on quantitative reasoning tasks that used pictorial stimuli or numerical stimuli. To examine this research question, performance in the Pictorial and Numeral conditions was compared for both the Equivalence and Number Series tasks. A 3 x 2 x 2 mixed design ANOVA was conducted with number correct as the dependent variable. The between subjects factor in the ANOVA was grade with three levels: kindergarten, first, and second grade.[2] The two within subject factors were Task (Equivalence and Number Series) and Condition (Pictorial and Numeral). Table 6 shows the results of this ANOVA. The grade

---

[2] Due to the small sample size, fifth grade was not included in any of the statistical analyses. Instead, the general trend for fifth graders was compared to the statistical results for kindergarten, first, and second grade in Chapter 5.

Table 6

Analysis of Variance for Number Correct in Pictorial and Numeral Conditions in both the Equivalence and Number Series Tasks

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 76.88*** | .000 | .53 |
| Error | 135 | (6.00) | | |
| | | Within Subject | | |
| Task (T) | 1 | 264.55*** | .000 | .66 |
| Condition (C) | 1 | 2.03 | .157 | .02 |
| T x G | 2 | 5.67** | .004 | .08 |
| C x G | 2 | 9.01*** | .000 | .12 |
| T x C | 1 | 29.80*** | .000 | .18 |
| T x C x G | 2 | 3.63* | .029 | .05 |
| Error (T) | 135 | (3.30) | | |
| Error (C) | 135 | (0.86) | | |
| Error (T x C) | 135 | (1.04) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. The two tasks were Equivalence and Number Series. The two conditions were Pictorial and Numeral.

*$p$<.05. **$p$<.01. ***$p$<.001.

by condition by task interaction was significant ($p$<.05) with a small effect size ($\eta_p^2$=.05). Therefore, the pattern of performance in the Pictorial and Numeral conditions differed both by grade and by task (see Figure 2).

Due to the significant three-way interaction, further analyses examined each task separately. The follow-up analysis consisted of a 3 x 2 mixed design ANOVA for each task with number of items correct as the dependent variable. The between subjects factor was grade and the within subject factor was condition. Because there were two follow-up ANOVAs, the significance level was set at α=.025 (α=.05/2). Significant effects from these ANOVAs were followed up by *t*-tests comparing mean differences in the two conditions. There were four follow-up *t*-tests, so the significance level for the *t*-tests was set at α=.0125 (α=.05/4).

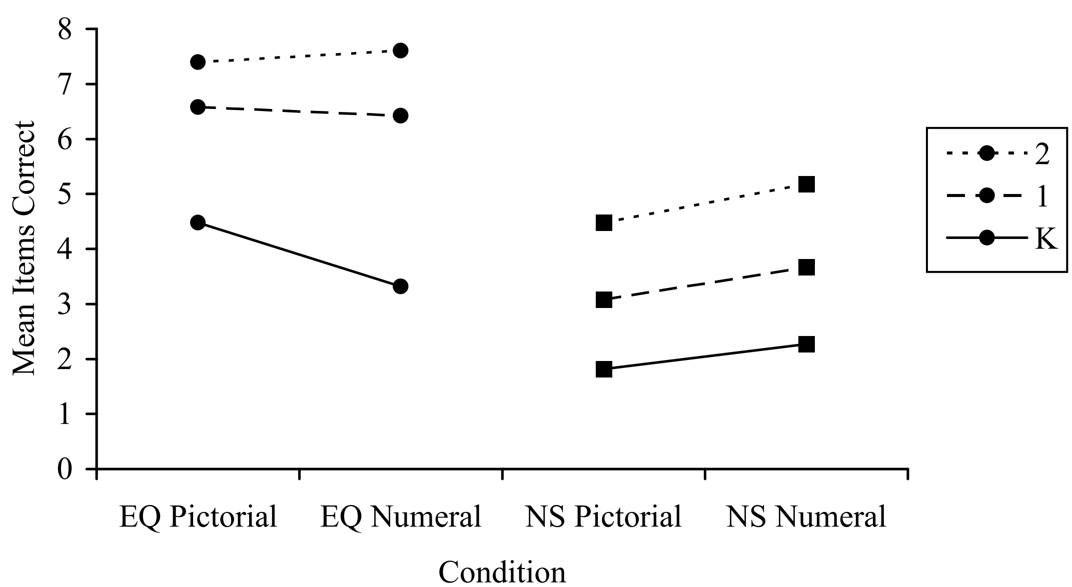



Figure 2. Performance in the Pictorial and Numeral conditions by grade and task. In the Equivalence (EQ) task, there were significant differences between conditions in kindergarten. All grades performed significantly better in the Numeral condition of the Number Series (NS) task.

*Equivalence.* The results of the follow-up ANOVA for the Equivalence task (see Table 7) demonstrated a significant condition by grade interaction ($p$<.001) with a medium effect size ($\eta_p^2$=.16). The pattern of performance between the Pictorial and Numeral conditions therefore varied by grade. The nonparallel slopes for the Equivalence task in Figure 2 illustrate this interaction. Consequently, performance in the two conditions was compared separately within each grade. In kindergarten, performance in the Pictorial condition was significantly better than the Numeral condition ($t$(43) = 4.61, $p$<.001; $d$=.60). There were no significant differences for either first grade ($t$(51) = 0.84,

Table 7

Follow-up Analysis of Variance for Number Correct in the Equivalence Task

| | df | $F$ | $p$ | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 75.96*** | .000 | .53 |
| Error | 136 | (4.00) | | |
| | | Within Subject | | |
| Condition (C) | 1 | 11.15*** | .001 | .08 |
| C x G | 2 | 13.21*** | .000 | .16 |
| Error | 136 | (0.84) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. The two conditions were Pictorial and Numeral.

*$p$<.025. **$p$<.01. ***$p$<.001.

*p*<.41) or second grade ($t(42) = 2.03$, $p<.05$, $d = .25$). Kindergarteners performed better when the problems were presented pictorially with a medium effect size whereas first second graders performed similarly in both conditions.

*Number Series.* A parallel ANOVA was conducted for the Number Series task (see Table 8). The interaction between grade and condition was not significant for this task. The main effect of condition was significant ($p<.001$) with a small effect size ($\eta_p^2$=.14). To follow-up the significant main effect, the *t*-test confirmed that students

Table 8

Follow-up Analysis of Variance for Number Correct in the Number Series Task

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 31.86*** | .000 | .32 |
| Error | 136 | (5.36) | | |
| | | Within Subject | | |
| Condition (C) | 1 | 22.28*** | .000 | .14 |
| C x G | 2 | 0.33 | .722 | .01 |
| Error | 136 | (1.05) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. The two conditions were Pictorial and Numeral.

*$p<.025$. **$p<.01$. ***$p<.001$.

performed significantly better in the Numeral condition than the Pictorial condition for all grade levels ($t(138) = 4.76$, $p<.001$; $d=.28$).

Interestingly, kindergarteners performed better on the Equivalence task when using pictures, but performed better in the Number Series task when using Arabic numerals. First and second graders performed better in the Number Series task when using Arabic numerals, but performed similarly in both conditions on the Equivalence task.

*Mixture of Arabic Numerals and Pictures*

Results of the first research question suggested that using pictures or Arabic numerals to solve a quantitative reasoning task influenced performance under certain conditions. However, the first research question only focused on the exclusive use of pictures or Arabic numerals. The second research question further explored this distinction by comparing performance in each condition to an additional condition that used a combination of pictures and Arabic numerals. To test this research question, the Mixed condition of the Number Series task combined beads and Arabic numerals in the stem of the task. Therefore, the second research question was addressed by comparing performance on the Mixed condition with performance on the Pictorial and Numeral conditions of the Number Series task.

A 3 x 3 mixed design ANOVA was conducted to answer the second research question with number of items correct as the dependent variable. The between subjects factor was grade and the within subject factor was condition (Pictorial, Numeral, and Mixed). The interaction between grade and condition was not significant (see Table 9).

Table 9

Analysis of Variance for Number Correct in Number Series Task with Mixed Condition

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 40.18*** | .000 | .38 |
| Error | 133 | (6.98) | | |
| | | Within Subject | | |
| Condition (C) | 2 | 19.31*** | .000 | .13 |
| C x G | 4 | 1.21 | .306 | .02 |
| Error | 266 | (1.12) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. The three conditions were Pictorial, Numeral, and Mixed.

*$p$<.05. **$p$<.01. ***$p$<.001.

All grades therefore had the same pattern of performance (see Figure 3). The main effect of condition was significant ($p$<.001) with a small effect size ($\eta_p^2$=.13).

Two follow-up *t*-tests compared performance in the Mixed condition to the Numeral and Pictorial conditions. The significance level for the follow-up *t*-tests was set at $\alpha$=.025 ($\alpha$=.05/2). Students performed significantly better in the Numeral condition than in the Mixed condition ($t(135) = 6.12$, $p$<.001; $d$=.35). However, students performed similarly in the Mixed and Pictorial conditions ($t(135) = 1.41$, $p$<.17). As with pictures,

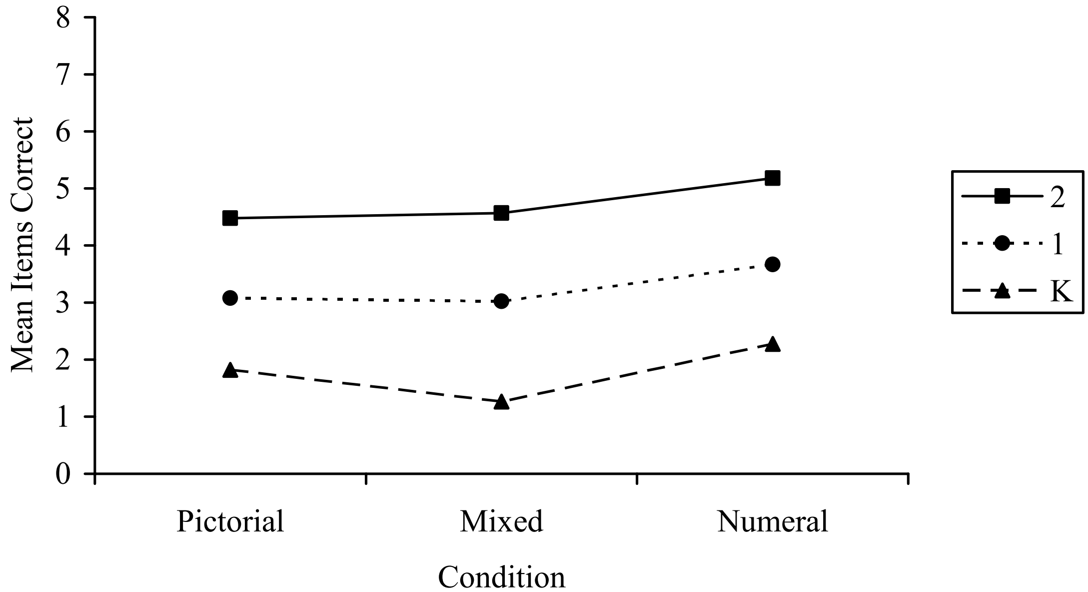



Figure 3. Performance in the Pictorial, Mixed, and Numeral conditions for the Number Series task. Students performed significantly better in the Numeral condition than the Mixed condition, but performed similarly in the Mixed and Pictorial conditions.

performance using a combination of pictures and Arabic numerals was significantly lower than performance with Arabic numerals only.

*Choice of Format*

Instead of number correct, the dependent variable for the third research question was the format that students chose to use when attempting the task. Recall that students were able to choose whether they wanted to use pictures or numerals to solve each item in the third condition of the Equivalence task, Choice. In addition to using only pictures or only numerals, many students chose to use a combination of pictures and numerals, representing one quantity with pictures and the other with numerals. This type of response was coded as using *both* pictures and numerals.[3] Table 10 shows the frequency that each format was chosen.

A 3 x 3 mixed design ANOVA was conducted with the frequency that each format was chosen as the dependent variable (see Table 11). The between subjects factor was grade and the within subject factor was format chosen (pictures, numerals, or both).

Table 10

Frequency of Format Chosen for Each Item in Choice Condition of the Equivalence Task

| Item | Kindergarten | | | First Grade | | | Second Grade | | |
|---|---|---|---|---|---|---|---|---|---|
| No. | P | N | B | P | N | B | P | N | B |
| 1 | 20 | 9 | 7 | 26 | 14 | 11 | 19 | 12 | 12 |
| 2 | 21 | 10 | 5 | 25 | 19 | 7 | 12 | 15 | 16 |
| 3 | 24 | 8 | 4 | 26 | 16 | 9 | 17 | 13 | 13 |
| 4 | 22 | 7 | 5 | 27 | 18 | 6 | 15 | 16 | 11 |
| 5 | 21 | 10 | 3 | 31 | 13 | 6 | 16 | 18 | 9 |
| 6 | 23 | 6 | 5 | 26 | 21 | 4 | 17 | 16 | 10 |
| 7 | 14 | 7 | 3 | 25 | 17 | 3 | 16 | 18 | 9 |
| 8 | 15 | 5 | 4 | 20 | 21 | 2 | 14 | 19 | 9 |
| Total | 160 | 62 | 36 | 206 | 139 | 48 | 126 | 127 | 89 |
| **Mean[a]** | **4.44** | **1.72** | **1.00** | **4.04** | **2.73** | **0.94** | **2.93** | **2.95** | **2.07** |

Note. P = Pictures. N = Numerals. B = Both pictures and numerals.

[a] Mean frequency across subjects.

[3] A few students chose to answer an item using a complete set of both cookies and numbers. This was recorded as either *cookies-numbers* or *numbers-cookies* depending on which format they responded with first. For data analysis, these students were categorized as choosing the first format that was used.

Table 11

Analysis of Variance for Frequency of Format Chosen in the Choice Condition of the Equivalence Task

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 9.08*** | .000 | .13 |
| Error | 127 | (0.23) | | |
| | | Within Subject | | |
| Format (Fo) | 2 | 21.25*** | .000 | .14 |
| Fo x G | 4 | 3.13* | .015 | .05 |
| Error | 254 | (9.15) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. Format compared choice of pictures, numerals, or both a picture and a numeral.

*$p$<.05. **$p$<.01. ***$p$<.001.

The interaction between grade and format was significant ($p$<.05; $\eta_p^2$=.05), illustrated in Figure 4 by nonparallel slopes. Due to the significant interaction between grade and format, follow-up analyses examined each grade separately.

To determine whether there were significant differences in format chosen within each grade, within subject one-way ANOVAs were conducted within each grade. The within subject variable was format (pictures, numerals, and both) and the dependent variable was the frequency with which each format was chosen. Since there were three follow-up ANOVAs, the significance level was set at $\alpha$=.017 ($\alpha$=.05/3). Table 12 shows

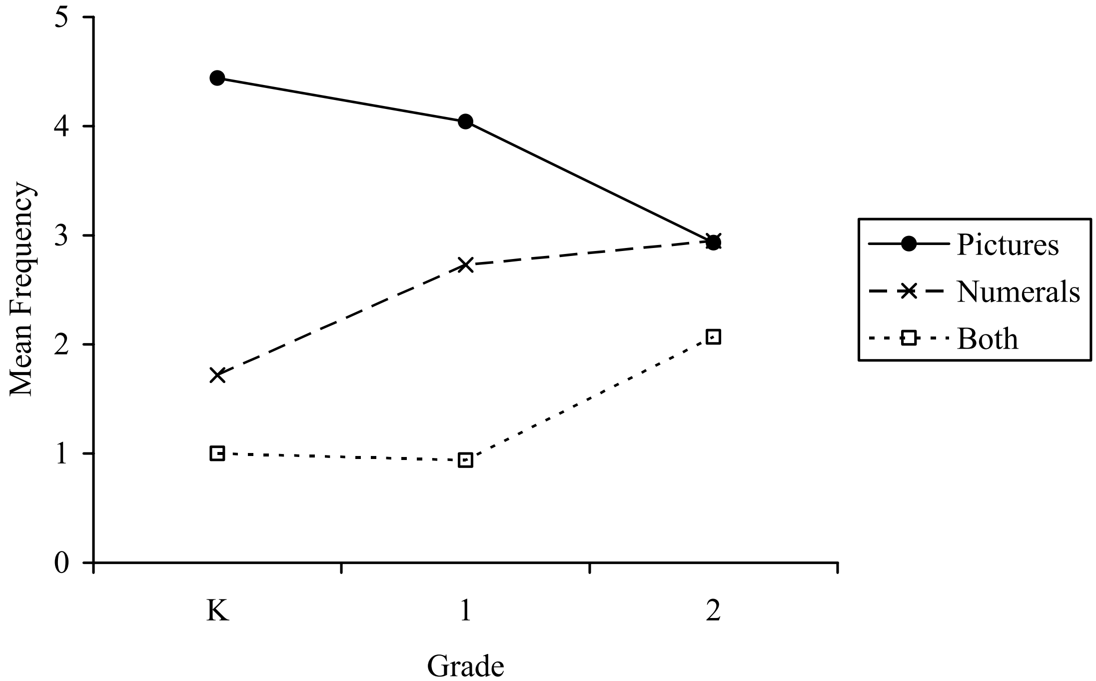



Figure 4. Frequency of the format that students chose to use in the Choice condition of the Equivalence task. The *both* format represents students who responded using one picture and one numeral. In kindergarten, pictures were chosen significantly more than numerals and both. In first grade, both pictures and numerals were chosen significantly more than both. There were no significant differences for second grade.

Table 12

Follow-up Analysis of Variance for Frequency of Format Chosen

in the Choice Condition of the Equivalence Task within Grade

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Kindergarten | | |
| Format | 2 | 13.24*** | .000 | .28 |
| Error | 70 | (8.97) | | |
| | | First Grade | | |
| Format | 2 | 13.63*** | .000 | .21 |
| Error | 100 | (9.05) | | |
| | | Second Grade | | |
| Format | 2 | 1.16 | .319 | .03 |
| Error | 84 | (9.42) | | |

Note. Values enclosed in parentheses represent mean square errors. Format compared choice of pictures, numerals, or both a picture and a numeral.

*$p$<.017. **$p$<.01. ***$p$<.001.

the follow-up results. A significant format effect was found for kindergarten and first grade ($p$<.001 for both grades) but not for second grade ($p$<.32).

Since there were significant differences in format chosen in kindergarten and first grade, post-hoc *t*-tests were conducted to compare the three formats. Six *t*-tests were conducted (three in kindergarten and three in first grade), and so the significance level was set at $\alpha$=.008 ($\alpha$=.05/6). Pictures were chosen significantly more than numerals for kindergarteners ($t(35) = 3.31$, $p<.008$; $d=1.00$) but not for first graders ($t(50) = 1.88$, $p<.07$). Numerals were chosen more frequently than both for first grade ($t(50) = 3.47$, $p<.001$; $d=.77$) but not for kindergarten ($t(35) = 1.42$, $p<.17$). Pictures were chosen significantly more frequently than both for kindergarten ($t(35) = 4.61$, $p<.001$; $d=1.34$) and first graders ($t(50) = 5.55$, $p<.001$; $d=1.30$).

To summarize, second graders chose each of the formats with comparable frequency. Kindergarteners chose pictures significantly more than numerals and both with large effect sizes, but there were no significant differences between numerals and both. There were no significant differences between pictures and numerals for first graders, but pictures and numerals were chosen significantly more than both, also with large effect sizes.

In addition to comparing the frequency that each format was chosen, an additional analysis compared performance in the condition where students could choose the format to the condition in which students were forced to use a specific format. To this end, performance in the Choice condition was compared to performance in the Pictorial and Numeral conditions (see Figure 5). A 3 by 3 mixed design ANOVA was conducted with items correct as the dependent variable. The between subjects variable was grade and the within subject variable was condition (Pictorial, Numeral, and Choice). The results are presented in Table 13. The grade by condition interaction was significant ($p<.001$). Therefore, follow-up analyses examined the pattern of performance separately for each grade.

As in the previous analysis, a one-way ANOVA for each grade was conducted to determine whether there were significant effects of condition at each grade level. Number

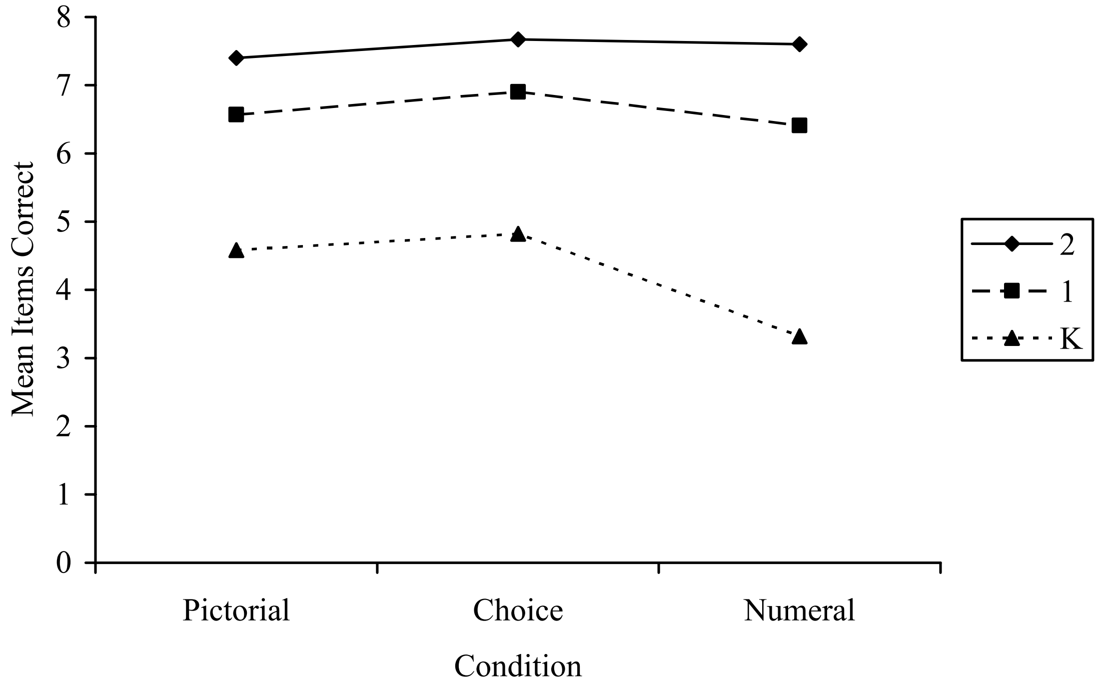



Figure 5. Performance in the Pictorial, Choice, and Numeral conditions for the Equivalence task. Both kindergarteners and first graders performed significantly better in the Choice condition than the Numeral condition.

Table 13

Analysis of Variance for Number Correct in the Equivalence Task including Choice Condition

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Grade (G) | 2 | 64.29*** | .000 | .50 |
| Error | 129 | (5.48) | | |
| | | Within Subject | | |
| Condition (C) | 2 | 21.29*** | .000 | .14 |
| C x G | 4 | 10.36*** | .000 | .14 |
| Error | 258 | (0.73) | | |

Note. Values enclosed in parentheses represent mean square errors. The grade factor included kindergarten, first, and second grades. The three conditions were Pictorial, Numeral, and Choice.

**p*<.05. ***p*<.01. ****p*<.001.

correct was the dependent variable and condition (Choice, Pictorial, and Numeral) was the independent variable. Since there were three follow-up ANOVAs, the significance level was set at $\alpha$=.017 ($\alpha$=.05/3). As shown in Table 14, all of the ANOVAs were significant (*p*<.001 for kindergarten; *p*<.017 for first and second grade), although the effect was larger at kindergarten ($\eta_p^2 = .33$) than at first and second grade ($\eta_p^2 = .08$ and $\eta_p^2 = .09$, respectively).

To compare performance in the Choice condition to the Numeral and Pictorial conditions, simple effects were examined with *t*-tests. There were six *t*-tests in all, so the

Table 14

Follow-up One-Way Analysis of Variance for Number Correct in the Equivalence Task including Choice Condition within Grade

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Kindergarten | | |
| Condition | 2 | 18.30*** | .000 | .33 |
| Error | 74 | (1.35) | | |
| | | First Grade | | |
| Condition | 2 | 4.55* | .013 | .08 |
| Error | 100 | (0.70) | | |
| | | Second Grade | | |
| Condition | 2 | 4.35* | .016 | .09 |
| Error | 84 | (0.21) | | |

Note. Values enclosed in parentheses represent mean square errors. The three conditions were Pictorial, Numeral, and Choice.

**p*<.017. ***p*<.01. ****p*<.001.

significance level was set at α=.008 (α=.05/6). Students in kindergarten and first grade performed significantly better in the Choice condition than in the Numeral condition ($t(37) = 4.73$, $p<.001$; $d=.66$ for kindergarteners; $t(50) = 3.55$, $p<.001$; $d=.32$ for first graders). However, there were no significant differences between the Choice condition and the Pictorial condition for kindergarteners and first graders ($t(37) = 1.16$, $p<.26$ for kindergarteners; $t(50) = 1.97$, $p<.06$ for first graders). Second graders, on the other hand, demonstrated the opposite pattern. Students performed better in the Choice condition than the Pictorial condition although the results did not reach the adjusted significance level ($t(42) = 2.61$, $p<.014$; $d=.34$). There was no difference between the Choice and the Numeral condition for second graders ($t(42) = 0.83$, $p<.42$). In other words, both kindergarteners and first graders performed better in the condition in which they were allowed to choose the problem format than they did when they were required to use Arabic numerals. Conversely, second graders performed better when they were allowed to choose the format than when they were required to use pictures.

*Consistency of Structure*

The fourth research question investigated whether students demonstrated a similar structure of quantity across conditions. To test this question, students were categorized according to the format that they chose to use on the majority of the items (five or more) in the Choice condition of the Equivalence task. Some students did not choose the same format in a majority of the items, so they were classified as *combined*. Accordingly, students were classified into one of four categories: cookies, Arabic numerals, both, and combined. (Recall that the *both* category represented students who chose one cookie and one Arabic numeral to solve an item.) Table 15 gives the percent of students who were classified in each category within each grade.

In order to compare performance of students according to the format that they preferred, a 4 x 2 x 2 mixed design ANOVA was conducted with number correct as the dependent variable. The two within-subject factors were task (Equivalence and Number Series) and condition (Pictorial and Numeral). The between subjects factor was format chosen (pictures, numerals, both, or combined). As shown in Table 16, the three-way interaction between format, task, and condition was not significant (see also Figure 6).

Table 15

Percent of Students Categorized in each Format

| | Pictures | Numerals | Both[a] | Combined[b] |
|---|---|---|---|---|
| K | 52 | 15 | 9 | 24 |
| 1 | 41 | 18 | 10 | 31 |
| 2 | 28 | 19 | 19 | 35 |
| Total | 39 | 17 | 13 | 31 |

Note. Totals may not sum up to 100 due to rounding. Students were categorized according to their choice on five or more items in the Choice condition of the Equivalence task.

[a] The both category consists of students who used one picture and one numeral on a majority of the items.

[b] The Combined category consists of students who did not choose one format on five or more items.

Table 16

Analysis of Variance for Number Correct in Pictorial and Numeral Conditions in both the Equivalence and Number Series Tasks According to Format Chosen

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Format (Fo) | 3 | 2.78* | .044 | .06 |
| Error | 122 | (11.04) | | |
| | | Within Subject | | |
| Task (T) | 1 | 215.10*** | .000 | .64 |
| Condition (C) | 1 | 7.60** | .007 | .06 |
| T x Fo | 3 | 1.66 | .179 | .04 |
| C x Fo | 3 | 3.95** | .010 | .09 |
| T x C | 1 | 15.81*** | .000 | .12 |
| T x C x Fo | 3 | 0.98 | .404 | .02 |
| Error (T) | 122 | (3.30) | | |
| Error (C) | 122 | (0.91) | | |
| Error (T x C) | 122 | (1.09) | | |

Note. Values enclosed in parentheses represent mean square errors. Format compared choice of pictures, numerals, both, or combined. The two tasks were Equivalence and Number Series. The two conditions were Pictorial and Numeral.

**p*<.05. ***p*<.01. ****p*<.001.

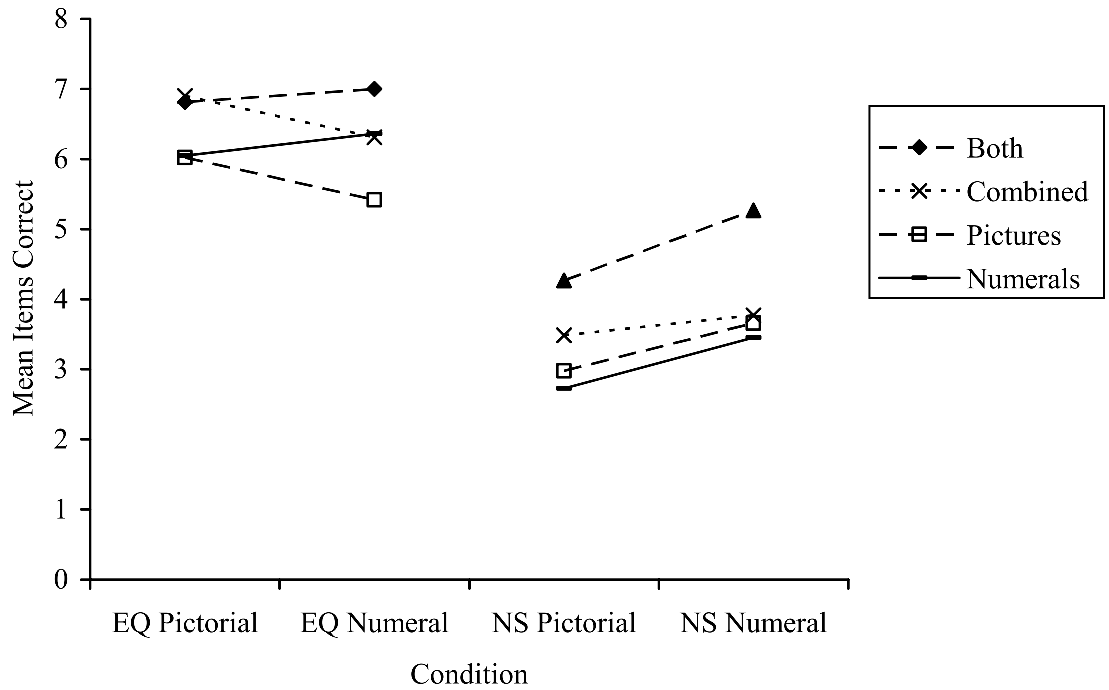



Figure 6. Performance in the Equivalence (EQ) and Number Series (NS) tasks based on the format that students chose to use in the Choice condition of the Equivalence task. Students were categorized as *both* if they used one picture and one numeral in a majority of the items. Students were categorized as *combined* if they did not use one format on a majority of the items. There was a significant format by condition interaction, as well as a significant task by condition interaction.

However, the format by condition interaction was significant ($p$<.01; $\eta_p^2$=.09), as was the task by condition interaction ($p$<.001; $\eta_p^2$=.12). Since the focus of this research question examined performance by the format chosen, the interaction between format and condition was further investigated.

To follow up the significant interaction between format and condition, total Pictorial and Numeral scores were calculated by summing up items correct across tasks (see Figure 7). Then four *t*-tests were conducted within each format comparing total Pictorial scores to total Numeral scores with the significance level set at α=.0125

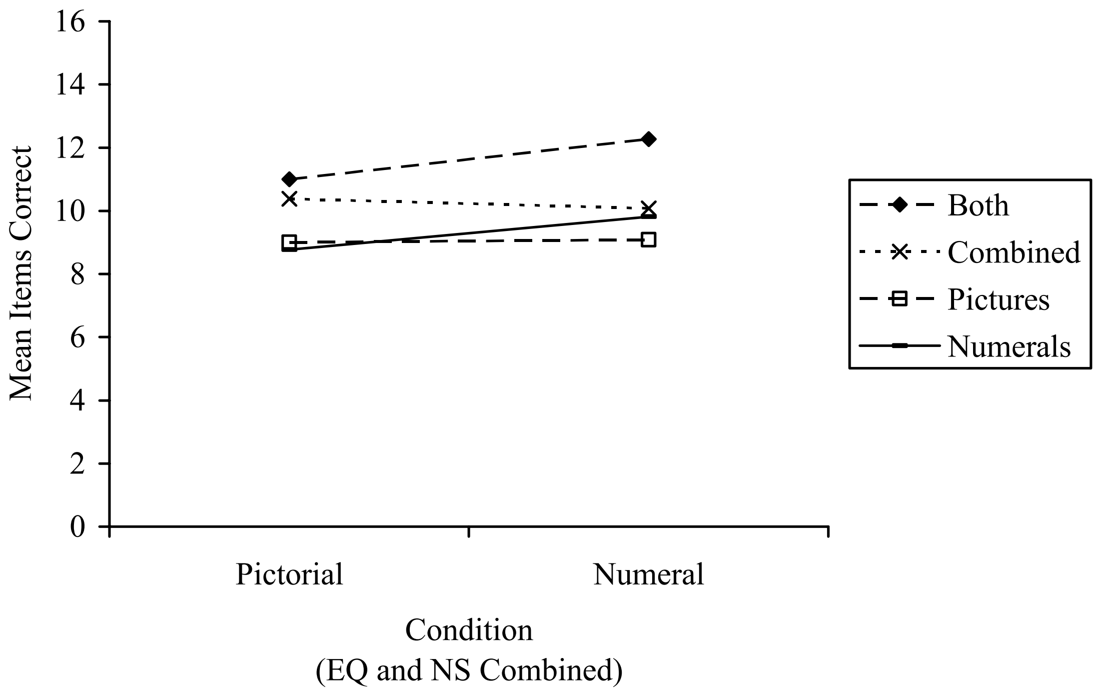



Figure 7. Performance by condition (Pictorial and Numeral) across tasks (Equivalence (EQ) and Number Series (NS)) and by the format that students chose to use in the Choice condition of the Equivalence task. Students were categorized as *both* if they used one picture and one numeral in a majority of the items. Students were categorized as *combined* if they did not use one format in a majority of the items. For students who chose numerals, the Numeral condition was significantly easier than the Pictorial condition. There were no significant differences for the other groups of students.

($\alpha$=.05/4). Students who chose numerals performed significantly better in the Numeral condition than in the Pictorial condition ($t(21) = 3.21$, $p<.01$; $d=.36$). There were no significant differences in performance for students who chose both ($t(14) = 2.43$, $p<.03$), cookies ($t(49) = .27$, $p<.79$) or a combination of formats ($t(38) = 1.10$, $p<.28$).

Although there were no significant differences between conditions for the students who chose pictures or a combination of formats, Figure 8 plotted performance for only the picture and combined groups. This figure suggested that there might be an interaction between task and condition for these two groups. Indeed, a significant task by

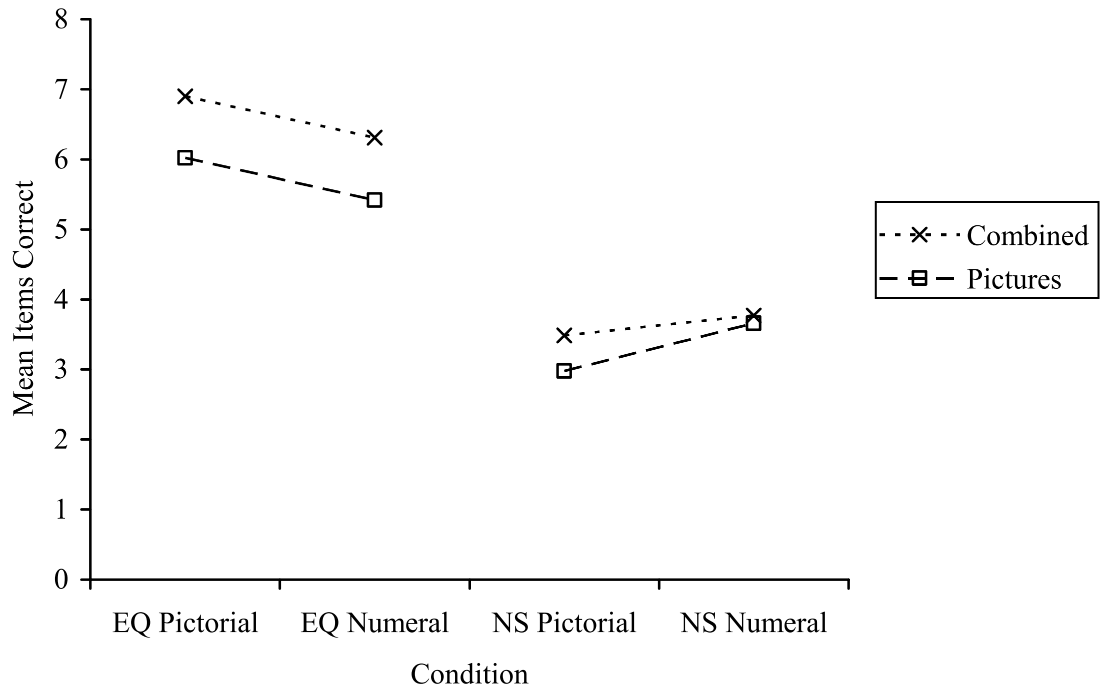



Figure 8. Performance in both the Equivalence (EQ) and Number Series (NS) tasks for students who chose to use Pictures and Combined in the Choice condition of the Equivalence task. Students were categorized as *combined* if they did not use one format on a majority of the items. Students who chose pictures performed significantly better in the Pictorial condition on the Equivalence task and in the Numeral condition on the Number Series task. Students who chose a combination of formats performed significantly better in the Pictorial condition on the Equivalence task, but there was no significant difference on the Number Series task.

condition interaction was found (see Table 16). The previous analysis that combined performance across tasks would have masked this interaction. Therefore, four *t*-tests were conducted to compare conditions within each task for the students who chose pictures and the students who chose a combination of formats. The significance level was set at $\alpha=.0125$ ($\alpha=.05/4$). On the Equivalence task, students who chose pictures performed significantly better in Pictorial condition than in the Numeral condition ($t(49) = 2.78$, $p<.01$; $d=.27$). On the other hand, these students performed significantly better in the Numeral condition on the Number Series task ($t(49) = 3.04$, $p<.01$; $d=.33$). Students who chose a combination of formats also performed significantly better in the Pictorial condition of the Equivalence task ($t(38) = 2.81$, $p<.01$; $d=.32$), but there was no significant difference between conditions on the Number Series task ($t(38) = 1.30$, $p<.21$).

To summarize, students who chose numerals performed significantly better in both of the Numeral conditions. Students who chose pictures performed better in the Pictorial condition of the Equivalence task and the Numeral condition of the Number Series task. Like those who chose pictures, students who chose a combination of formats performed better in the Pictorial condition of the Equivalence task, but they had comparable performed in both the Pictorial and Numeral conditions on the Number Series task. There were no differences in performance for students who chose both.

In addition to the significant interactions, the overall ANOVA (see Table 16) also indicated a main effect of format on performance. A one-way between subjects ANOVA was conducted with the format chosen as the independent variable and total items correct across all four conditions as the dependent variable (see Figure 9). Tukey HSD contrasts revealed that students who chose both performed significantly better than students who chose pictures ($p<.05$). There were no other significant differences between groups.

*Performance by Ability Level*

In addition to choice of format, performance in the Pictorial and Numeral conditions was also examined based on ability rather than on grade. To do this, students were categorized by their performance on the complementary task. In other words, performance in the Pictorial and Numeral conditions of the Equivalence task was examined as a function of performance on the Number Series task. A similar analysis was

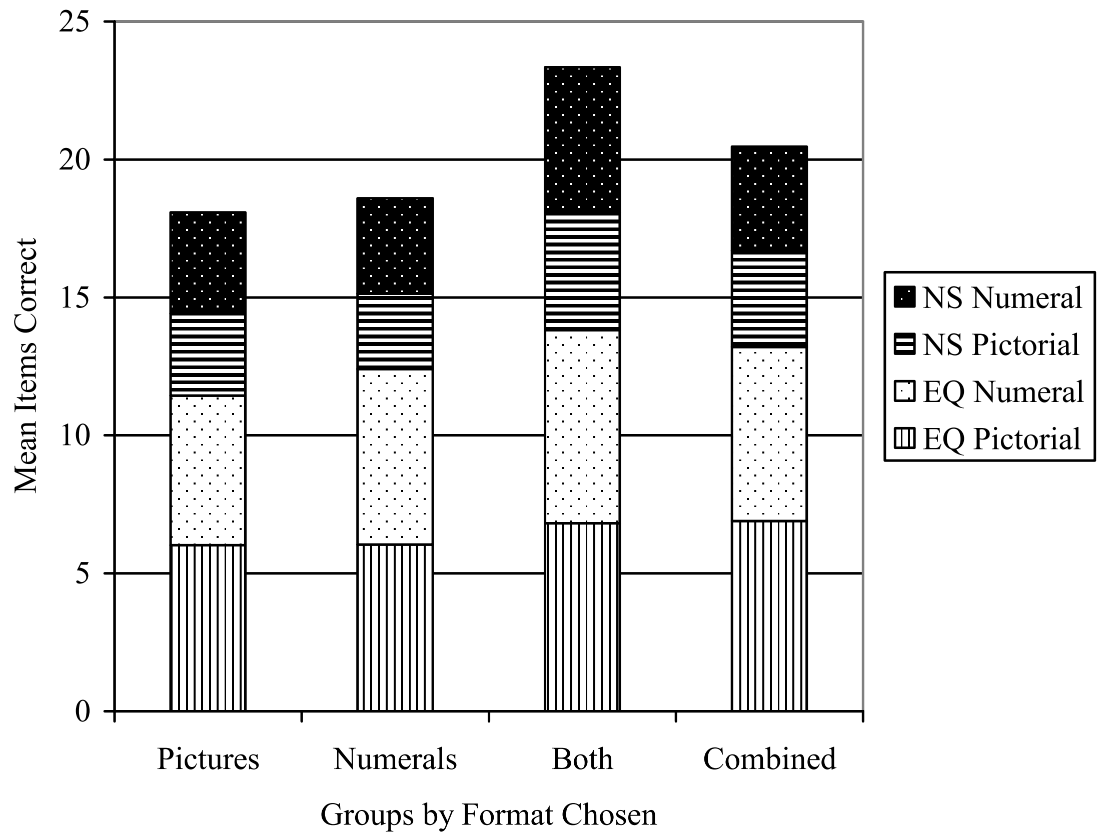



Figure 9. Total items correct in the Pictorial and Numeral conditions of the Equivalence (EQ) and Number Series (NS) task by the format that students chose to use in the Choice condition of the Equivalence task. Students were categorized as *both* if they used one picture and one numeral in a majority of the items. Students were categorized as *combined* if they did not use one format on a majority of the items. Students who chose both performed significantly better overall than students who chose pictures. There were no other significant differences.

also conducted for performance on the Number Series task based on scores from the Equivalence task. A 3 x 2 mixed design ANOVA was then conducted for each task with items correct as the dependent variable. The between subjects factor was ability level (high, medium, low), and the within subject factor was condition (Pictorial and Numeral). Since two ANOVAs were conducted, one for each task, the significance level was set at $\alpha=.025$ ($\alpha=.05/2$).

Table 17

Frequency of Students by Grade Categorized in each Ability Level based on Performance in the Number Series Task

| | Low[a] | Medium[b] | High[c] |
|---|---|---|---|
| K | 36 | 7 | 1 |
| 1 | 20 | 16 | 15 |
| 2 | 9 | 9 | 26 |
| Total | 65 | 32 | 42 |

[a] Low ability students answered three or less items correct.

[b] Medium ability students answered four items correct.

[c] High ability students answered five or more items correct.

*Equivalence.* Students were categorized into high ability, medium ability, and low ability groups based on their performance on the Numeral condition of the Number Series task. The average number of items correct in the Numeral condition was four. Therefore, students who answered 5 or more items correct were classified as high ability students, medium ability students answered 4 items correctly, and low ability students answered 3 or less items correctly. Table 17 shows the frequency of students in each grade by ability level. Performance was then compared in the Pictorial and Numeral conditions for the Equivalence task for each ability level (see Figure 10).

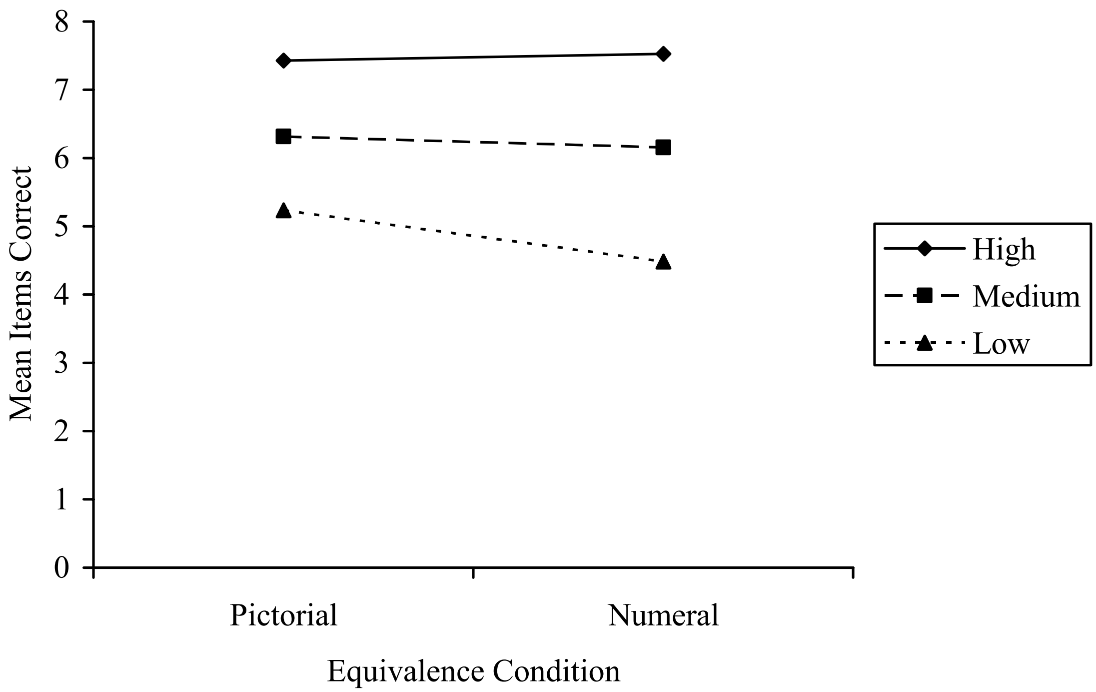



Figure 10. Performance in the Pictorial and Numeral conditions of the Equivalence task by ability (low, medium, or high) on the Numeral condition of the Number Series task. There was a significant difference between the Pictorial and Numeral conditions for the low ability students, but not for the medium or high ability students.

There was a significant interaction between ability and condition (see Table 18; $p$<.01; $\eta_p^2$=.07). Follow-up analyses consisted of *t*-tests within each ability level comparing performance in the Pictorial and Numeral conditions. Since there were three *t*-tests, the significance level was set at α=.017 (α=.05/3). Low ability students performed significantly better in the Pictorial condition than the Numeral condition ($t(63) = 3.79$, $p$<.001; $d$=.33). There were no significant differences for medium and high ability students ($t(31) = 0.63$, $p$<.54 for medium ability; $t(41) = 0.68$, $p$<.50 for high ability). The same results were found when the sample was divided based on performance in the Pictorial condition of the Number Series task. To summarize, low ability students

Table 18

Analysis of Variance for Number Correct in the Equivalence Task by Ability based on the Numeral Condition of the Number Series Task

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Ability (A) | 2 | 29.95*** | .000 | .31 |
| Error | 135 | (5.89) | | |
| | | Within Subject | | |
| Condition (C) | 1 | 4.99* | .027 | .04 |
| C x A | 2 | 5.30** | .006 | .07 |
| Error | 135 | (0.93) | | |

Note. Values enclosed in parentheses represent mean square errors. Students were categorized according to low, medium, and high ability on the Number Series task. The two conditions were Pictorial and Numeral.

*$p$<.025. **$p$<.01. ***$p$<.001.

performed better with pictures on the Equivalence task than with Arabic numerals. However, there were no significant differences for medium and high ability students.

*Number Series.* Similar to the previous analysis, students were grouped according to their ability on the Numeral condition of the Equivalence task. The average performance in this condition was 6, so students who answered 5 or less items correct were categorized as low ability students, medium ability students answered 6 items correctly, and high ability students answered 7 or more items correctly. The frequency of

Table 19

Frequency of Students by Grade Categorized in each Ability Level based on Performance in the Equivalence Task

| | Low[a] | Medium[b] | High[c] |
|---|---|---|---|
| K | 34 | 4 | 6 |
| 1 | 13 | 11 | 28 |
| 2 | 1 | 3 | 39 |
| Total | 48 | 18 | 73 |

[a] Low ability students answered five or less items correct.

[b] Medium ability students answered six items correct.

[c] High ability students answered seven or more items correct.

students in grade by ability level is presented in Table 19. Performance was then compared for each ability level in the Pictorial and Numeral conditions for the Number Series task (see Figure 11). As shown in Table 20, there was no significant interaction between ability level and performance ($p<.38$). Similarly, no significant interaction was found when ability level was defined by performance in the Pictorial condition of the Equivalence task. The significant main effect of condition duplicated the significant main effect found when comparing performance by grade. Consequently, no follow-up tests were conducted.

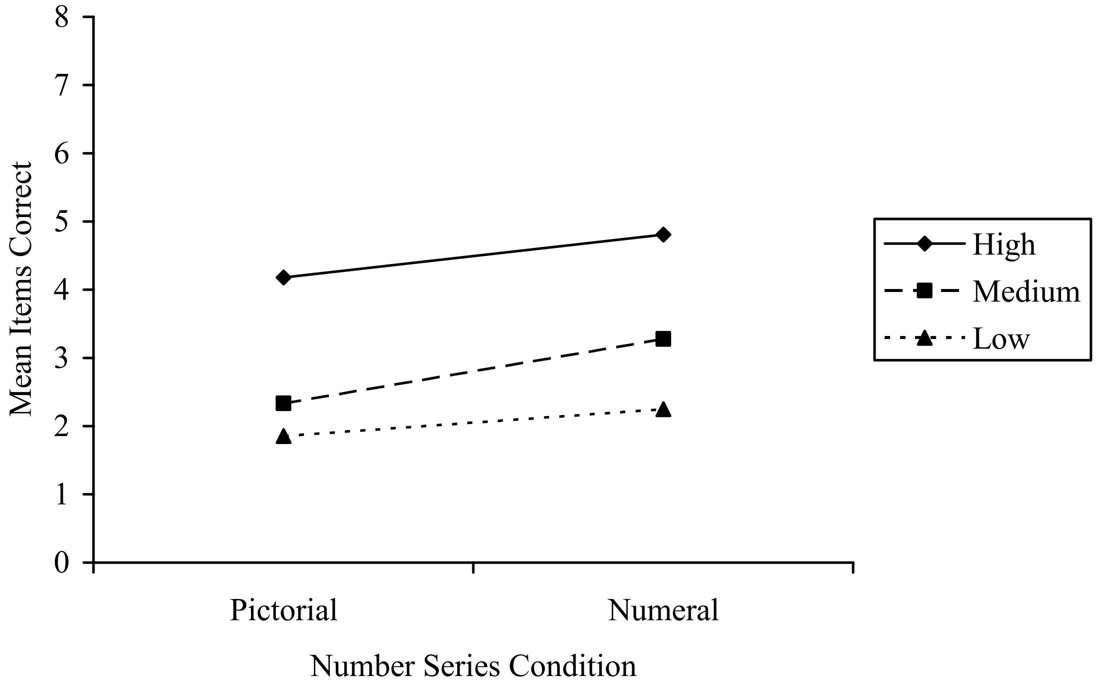



Figure 11. Performance in the Pictorial and Numeral conditions of the Number Series task by ability level (low, medium, or high) on the Numeral condition of the Equivalence task. Students performed significantly better in the Numeral condition than in the Pictorial condition.

Table 20

Analysis of Variance for Number Correct in the Number Series Task by Ability based on the Numeral Condition of the Equivalence Task

| | df | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|
| | | Between Subjects | | |
| Ability (A) | 2 | 34.28*** | .000 | .34 |
| Error | 135 | (1.05) | | |
| | | Within Subject | | |
| Condition (C) | 1 | 20.39*** | .000 | .13 |
| C x A | 2 | 0.99 | .374 | .01 |
| Error | 135 | (1.05) | | |

Note. Values enclosed in parentheses represent mean square errors. Students were categorized according to low, medium, and high ability on the Equivalence task. The two conditions were Pictorial and Numeral.

*$p$<.025. **$p$<.01. ***$p$<.001.

# CHAPTER 5

# DISCUSSION

The two most popular theories of quantitative reasoning in young children are currently Case's developmental theory of reasoning based on central conceptual structures and Huttenlocher and colleagues' descriptive theory of mental models. Case proposed that children develop central conceptual structures that guide reasoning and that these conceptual structures mature in qualitatively distinct stages (Case & Okamoto, 1996). In the first stage, children have separate verbal and mental-image structures for quantity. In the second stage, students develop a mental number line that merges the verbal and mental-image structures. As children's conceptual structures mature, they are able to integrate multiple counting lines when solving quantitative tasks. According to Huttenlocher and colleagues, preschool children reason quantitatively by constructing a mental representation of the critical quantitative features of a situation. Quantitative transformations are then mentally visualized (Huttenlocher et al., 1994; Mix, 1999; Mix et al., 2002).

The purpose of this study was to investigate these two theories in kindergarten through second grade students. Huttenlocher and colleagues' theory of quantitative reasoning in preschool children provided a detailed description of the mental-image structure of quantity that, according to Case, children developed in the first stage. However, little research has examined the conceptual structures of kindergarten through second grade students.

To assess the quantitative structures of kindergarten through second grade students, two quantitative reasoning tasks were administered using both pictures and Arabic numerals. The first research question simply asked which format resulted in the best performance. The second research question compared performance in conditions that exclusively used pictures or Arabic numerals to a condition with a mixture of both formats. These two research questions assumed that students would perform better in the condition that more closely matched the conceptual structure that they used to solve the task. Whereas the first two research questions examined performance, the third research question asked which format students preferred to use. This research question assumed that students would choose to use the format that matched the conceptual structure that they used to solve the quantitative reasoning task. In contrast to the first three research questions that categorized students by grade, the fourth and fifth research questions categorized students according to their performance on the quantitative reasoning tasks. The fourth research question examined patterns of performance by classifying students according to their format preference on the Equivalence task. The fifth research question examined performance when students were grouped according to quantitative reasoning ability.

In addition to the early elementary students who were the focus of this study, a smaller sample of 9 fifth grade students also completed the quantitative reasoning tasks. Fifth graders were tested to compare early elementary students' conceptual structures to students who had developed a more mature understanding of the mental number line. Since the sample of fifth graders was small, their results were not entered into the statistical analyses in Chapter 4. Instead, the general trends for fifth grade students were compared to the statistical results for the kindergarten through second grade students in the following discussion.

*Comparison of Arabic Numerals and Pictures*

If early elementary students used a verbal structure of quantity as most measures of quantitative reasoning assume, then they would perform better in the Numeral conditions on both tasks. However, if young students used the mental-image structure of quantity that Huttenlocher and colleagues proposed, then early elementary students would perform better in the Pictorial condition on both tasks. However, the analysis revealed an unexpected result: a significant interaction between grade, task, and condition indicated that students' pattern of performance differed across the tasks.

Performance on the Equivalence task, where students produced an equivalent quantitative set, supported the hypothesis that early elementary students had a mental-image structure of quantity. Specifically, kindergarten students performed better in the Pictorial condition than in the Numeral condition. First and second graders, on the other hand, demonstrated that their verbal and mental-image structures of quantity were equally useful on the Equivalence task. Even though second graders performed better in the Numeral condition, the difference did not exceed the adjusted significance level. A ceiling effect might have masked differences in performance for the second grade sample. Unfortunately, the entire fifth grade sample scored at the ceiling on the Equivalence task in both conditions. Therefore, these results could not conclusively suggest whether students with a mature quantitative structure performed better on the Equivalence task using a verbal or mental-image structure. However, these results provided convincing evidence that kindergarteners also possess the mental-image structure of quantity suggested by Huttenlocher and colleagues.

On the Number Series task, the fifth grade sample performed as well in both conditions (mean=6.78 for both). Therefore, a mature structure of quantity on this task with small quantities produced no difference in performance between pictures and Arabic numerals. On the other hand, students in kindergarten through second grade performed better in Numeral condition than the Pictorial condition. Alone, these results would support the assumption that early elementary students were proficient at reasoning with Arabic numerals. However, by comparing performance on the Number Series task to performance on the Equivalence task, the results actually supported Case's theory. More specifically, Case proposed that in the *predimensional* stage of reasoning, children had two independent conceptions of quantity: a verbally counting ability and another nonverbal quantitative ability that included determining which set of objects had more and less (Case & Okamoto, 1996).

Most of the items on the Number Series task were effectively solved by applying the counting schema. For example, the item *3, 3, 4, 4, 5* could be solved by realizing that the pattern required counting up and repeating the digits. In the Pictorial condition of this task, instead of looking at the whole picture to determine a pattern, most students attempted to count the number of beads on each string. Many students then became frustrated by the large amount of beads that had to be counted. On the other hand, most students were able to quickly and accurately label the Arabic numerals. Recognizing the Arabic numerals enabled them to discern the pattern much more readily.

Alternatively, the Equivalence task assessed the ability to apply a part-whole schema to equivalent sets. The part-whole schema consists of understanding that sets are additive by combining two quantities to make a larger quantity (Resnick, 1989). Resnick proposed that the part-whole schema initially develops as a protoquantitative structure allowing young children to make quantitative judgments perceptually. On the Equivalence task, students had to understand that two quantities could be combined to match the experimenter's quantity. The additive nature of the parts (the students' cookies) and wholes (the experimenter's cookies) was more apparent in the Pictorial condition because the students could actively manipulate the sets of cookies. In contrast, since Arabic numerals symbolize quantitative sets, the additive nature of Arabic numerals was not as evident. Three cookies and four cookies obviously combined to make seven cookies. Without understanding the quantitative sets that the Arabic numerals 3 and 4 symbolize, 3 and 4 could not logically combine to make 7. Consequently, without understanding the symbolic nature of Arabic numerals, successful performance in the Numeral condition became a matter of chance.

Therefore, the two quantitative tasks in this research study assessed different structures of quantity. Both the Equivalence task and the nonverbal task that Huttenlocher and colleagues administered to preschool students measured the protoquantitative part-whole schema. The Number Series task, on the other hand, assessed the verbal counting structure. Kindergarteners showed evidence of two separate quantitative structures because they performed better in the Pictorial condition on the task that assessed the protoquantitative structure but performed better in the Numeral condition on the task that assessed the verbal counting structure. First and second graders performed better in the

Numeral condition on the task that assessed the counting schema, but performed comparably in both conditions on the task that assessed their protoquantitative structure. This supported Case's proposition that students begin to merge their counting and nonverbal reasoning structures at approximately first grade.

*Mixture of Arabic Numerals and Pictures*

In addition to examining how early elementary students performed on tasks that used exclusively pictures or Arabic numerals, the second research question examined their performance on a task with a mixture of pictures and Arabic numerals. This research question assumed that if students had merged their verbal and mental-image structures of quantity, then they would perform as well on a task with a mixture of Arabic numerals and pictures, the Mixed condition of the Number Series task, as on a task that used either exclusively. Performance in the Mixed condition was thus compared to performance in the Pictorial and Numeral conditions.

While students in all grades performed significantly worse in the Mixed condition than they did in the Numeral condition, there was no difference in performance between the Mixed and Pictorial conditions. Because the Numeral condition of the Number Series task evoked students' counting schemas, the presence of any pictures appeared to thwart successful application of the counting schema. In the Mixed condition, two quantities were always represented by pictures and two to four quantities were represented by Arabic numerals (see Figure A2). Students also responded with an Arabic numeral. Therefore, the only difference between the Numeral and Mixed conditions was replacement of two Arabic numerals with pictures. Substituting any pictorial representations of quantity in a string of four to six Arabic numerals was therefore similar to substituting all Arabic numerals with pictorial representations on the Number Series task.

The fifth grade sample also had a slightly lower mean score in the Mixed condition (mean number correct = 6.11) than in both the Pictorial and Numeral conditions (mean number correct = 6.78 for both). Therefore, the assumption that students would demonstrate comparable performance in the Mixed condition if they had merged their verbal and mental-image structures of quantity seemed to be incorrect for the Number Series task. Unfortunately, a similar evaluation could not be made for the Equivalence task that measured the part-whole schema. Indeed, when given the option of choosing whether to use pictures or Arabic numerals on the Equivalence task, some students chose to use a combination of both. This combination would be comparable to the Mixed condition of the Number Series task. Therefore, the results for this research question would most likely be different on a measure of the part-whole schema.

*Choice of Format*

The third research question focused on students' choice of pictures or Arabic numerals, assuming that students would choose to use the format that was more closely related to the structure that they used to solve the quantitative task. To assess this research question, students were given the option of using either pictures or Arabic numerals in the Choice condition of the Equivalence task. Students in each grade showed a different pattern of choice. Kindergarteners used pictures significantly more frequently than all other formats. Although first graders tended to use pictures most frequently, they only used pictures significantly more frequently than both. Second graders used pictures, Arabic numerals, and both with approximately equal frequency. The observed developmental trend for the Equivalence task was that early elementary students initially chose pictures, and then began to prefer using Arabic numerals as their quantitative reasoning structure developed. The data from fifth graders confirmed this trend. They chose Arabic numerals the most (33 total items), then both (23 total items), and cookies least frequently (16 total items).

These results supported the theoretical conclusion derived from students' performance in the Pictorial and Numeral conditions of this task. Kindergarteners performed better on the Equivalence task when they used pictures and they also chose to use pictures more frequently. These two findings supported the claim that kindergarteners used mental-image representations of quantity on tasks that assessed their part-whole schema. Although first graders performed similarly in the Pictorial and Numeral conditions, they chose pictures more frequently than Arabic numerals even though the result did not exceed the significance level. Even though first graders seemed to be transitioning to a merged structure of quantitative reasoning, they still provided some evidence of retaining a separate mental-image representation of quantity by still preferring to use pictures. Second graders had approximately equivalent performance in the Pictorial and Numeral conditions and they also chose to use pictures and Arabic numerals with equal frequency. As a group, second graders demonstrated a more merged structure of quantity. However, note that second graders still preferred to use pictures on about one third of the items.

In addition to preference of formats, performance in the Choice condition was also compared to the two conditions where students were forced to use either Arabic numerals or pictures. Both kindergarteners and first graders performed better in the

Choice condition than they did in the Numeral condition. Second graders, on the other hand, performed better in the Choice condition than they did in the Pictorial condition with marginal statistical significance. Caution should be taken in interpreting the results of this analysis because all students did not use the same format in the Choice condition. Regardless, when students were allowed to choose their strategy, kindergarteners and first graders performed better than when they were forced to use Arabic numerals. This suggested that these students still had difficulty using Arabic numerals to solve part-whole tasks. When second graders were allowed to choose their strategy, they performed better than when they were forced to use pictures. This suggested that students with a mature quantitative conceptual structure might actually perform better when using Arabic numerals on a part-whole task.

Overall, these results provided evidence that most kindergarteners and first graders and several second graders preferred to use pictures when solving part-whole quantitative reasoning tasks. Therefore, measures of quantitative reasoning should provide the option for early elementary to use pictures on tasks that evoke the part-whole schema.

*Consistency of Structure*

The central tenet of Case's theory was that central conceptual structures influence performance on a broad range of problems within a particular domain (Case, 1993). The fourth research question addressed this proposition by grouping students according to the format that they chose the most frequently in the Choice condition of the Equivalence task. Overall, the results supported Case's conception of a central conceptual structure.

Students who chose both formats on the same item performed the best overall. These students demonstrated no differences in performance across conditions, although the small number of students in this group and a ceiling effect on the Equivalence task might have masked differences. In general, these students demonstrated a merged structure of quantity. Whereas students who had not merged their structures of quantity performed better in the Pictorial condition of the Equivalence task, these students performed similarly in the Pictorial and Numeral conditions. Therefore, students who chose both formats demonstrated an integrated part-whole schema and counting schema that enabled them to solve part-whole tasks using both pictures and Arabic numerals. In fact, their structures had merged so well that they chose to solve the Equivalence task with a combination of pictures and Arabic numerals.

By the same logic, students who chose Arabic numerals on the Equivalence task had a slightly less mature quantitative structure than the students who chose both. That they chose to use Arabic numerals on the Equivalence task suggested that they had integrated their counting schema with their part-whole schema. They demonstrated better performance overall in the Numeral conditions, although this was most likely due to much better performance in the Numeral condition of the Number Series task. Their less mature quantitative conceptual structure therefore inhibited their performance in the counting task when it was presented using pictures.

Students who chose a combination of formats demonstrated the opposite pattern from students who chose Arabic numerals. Comparable to the students who chose both, there was no difference in performance between the conditions in the Number Series task. However, students who chose a combination of formats performed better in the Pictorial condition of the Equivalence task. Although these students seemed to have developed the ability to transfer their mental-image structure to a task that afforded a counting schema, they were unable to transfer their counting schema to the task that afforded a part-whole schema.

Finally, students who chose pictures had two distinct quantitative structures. These students performed better when using pictures in the Equivalence task, but they performed better using Arabic numerals in the Number Series task. Therefore, these students used the structure of quantity that the task directly afforded.

*Performance by Ability Level*

Since grade level is a proxy for overall cognitive development or ability, the first three research questions categorized students according to their grade. However, the final research question examined the pattern of performance by ability directly by grouping students according to their performance on the complementary task. The low ability groups consisted mostly of kindergarteners, the medium ability groups consisted mostly of first graders, and the high ability groups consisted mostly of second graders. However, there were some kindergarteners who were in the high ability groups and some second graders who were in the low ability groups.

The pattern of performance based on ability was similar to the pattern of performance based on grade. On the Equivalence task, the low ability group, as with kindergarteners, performed significantly better in the Pictorial condition. Like first graders, the medium ability group performed equally as well in both conditions. Both the second graders and the high ability group also had similar performance in these two conditions, although this was most likely caused by a ceiling effect on this task.

Similarly, when grouped either by age or by ability, only a main effect of condition was found for the Number Series task.

*Summary*

The overall research question that guided this study was “What is the conceptual structure that kindergarten, first, and second grade students use on quantitative reasoning tasks?” Kindergarteners were hypothesized to have distinct verbal and mental-image structures of quantity. This hypothesis was supported by the results. On the Equivalence task, which assessed a protoquantitative part-whole schema, kindergarteners performed better in the Pictorial condition than in the Numeral condition and when given the option, chose to use pictures more frequently than Arabic numerals. On the other hand, kindergarteners performed better in the Numeral condition on the Number Series task that assessed the verbal counting structure. Consequently, kindergarteners’ mental-image structure of quantity was applied to their performance on the Equivalence task and their verbal counting structure was applied to their performance on the Number Series task.

In contrast to the kindergarteners, second graders were hypothesized to have merged their verbal and mental-image structures of quantity. The results provided tentative support for this hypothesis. Based on the results from the small fifth grade sample, a mature quantitative conceptual structure produced similar performance when the Number Series task was presented with pictures and Arabic numerals. However, second graders still performed better using Arabic numerals on the Number Series task. Furthermore, second graders performed slightly better in the Numeral condition of the Equivalence task. Due to a ceiling effect on this task, additional research will be necessary to determine how students with a mature quantitative conceptual structure perform on the Equivalence task.

Finally, first graders were hypothesized to be in the process of merging the two structures of quantity. As expected, first graders performed similarly to kindergarteners in some instances and similarly to second graders in other instances. Like both kindergarteners and second graders, first graders performed better in the Numeral condition of the Number Series task. On the Equivalence task, they performed midway between the pattern for kindergarteners and for second graders with equal proficiency in both conditions. While cookies were chosen most frequently in the Equivalence task, more first graders chose to use Arabic numerals than kindergarteners.

Contrary to expectation, the Equivalence and Number Series tasks displayed different patterns of results. The tasks were designed with the assumption that they would produce similar patterns of performance across conditions. However, re-consideration of the demands of the tasks suggested that the two tasks actually assessed different structures of quantity. This unexpectedly enriched the study.

The results that bore on the second research question were the only other outcome that did not confirm the hypotheses of the study. Results for the Mixed condition of the Number Series task presented a more complex picture of the structure of quantitative reasoning than originally hypothesized. These results demonstrated that substituting just two pictures for a series of Arabic numerals significantly lowered performance to levels similar to those observed when the task was presented with only pictures, even for students who otherwise demonstrated a merged structure of quantity. This result undermined the assumption that students would be able to perform equally as well on all tasks using both pictures and Arabic numerals when they have merged their two structures of quantity. Indeed, Arabic numerals tend to offer a more powerful method of reasoning quantitatively (Mix et al., 2002). Using Arabic numerals reduces memory burdens and decreases the likelihood for errors when counting the stimuli. Although these results did not support the original hypothesis of this study, reconsideration of the assumptions of this research question revealed why. This research question was based on the inaccurate assumption that both the Equivalence and Number Series tasks would measure the same general quantitative reasoning structures. Since the Number Series task afforded the verbal structure of quantity, the Mixed condition unnecessarily evoked the mental-image structure, thereby complicating performance on this task. Consequently, the mental-image representation of quantity was only helpful on certain tasks, particularly those that afforded a part-whole schema like the Equivalence task.

*Limitations*

One of the limitations of this research study was the constitution of the sample. First, students were not randomly selected from a larger population to participate in the study. Information about the study was sent home with all of the students at Tonganoxie Elementary School. The parents then had to sign and return an informed consent document. This might have biased the sample. However, the teachers at Tonganoxie Elementary School commented that students from a range of ability levels did participate in the study. The sample was also relatively homogeneous with mostly Caucasian middle class students who lived in a relatively rural community. Therefore, the results might not generalize to other student populations, particularly those in an urban community or those with greater diversity in ethnicity and social class.

Since the experimenter in the study also authored the paper, the experimenter might have biased students' responses during the task. However, this conclusion was unlikely since the results for the Number Series task were contrary to the original hypothesis. Regardless, a blind administration of the experiment would have been desirable.

A ceiling effect also most likely influenced performance on the Equivalence task. This effect might have masked differences in performance between the Pictorial and Numeral conditions, particularly in the high ability and second grade samples.

*Future Directions*

This study should be replicated using different task stimuli to determine whether these results generalize to the universe of pictorial objects. The Pictorial conditions in both tasks used round objects (beads and cookies) and were both presented individually using a foam board. Therefore, results may differ for different types of pictorial objects, such as for rectangular boxes or triangular pyramids. Furthermore, the foam board format did not allow students to work out the item solutions using paper and pencil as many quantitative tasks allow. Results, particularly for the Number Series task, may have differed if students had the opportunity to write Arabic numerals below the string of beads.

The data from this study compared performance on quantitative reasoning tasks using pictorial and verbal stimuli. To make further distinctions of the quantitative reasoning conceptual structures of early elementary students, additional research will be necessary using different research paradigms. Other methods, such as think-aloud procedures, might enable researchers to understand students' thought processes as they are solving quantitative reasoning tasks using pictures and numerals.

Previous research has also established that preschoolers have the ability to count in their verbal structure (Briars & Siegler, 1984; Gelman & Gallistel, 1978). This ability was assessed by performance on the Number Series task. Secondly, preschoolers have the ability to understand equivalence though part-whole relationships in their mental-image structure (Resnick, 1989). This ability was assessed by performance on the Equivalence task. Finally, preschoolers have the ability to determine *more* and *less* in multiple sets of objects (Barth et al., 2005; Huntley-Fenner & Cannon, 2000; Siegel, 1974). However, this research study did not have a measure related to the ability to determine more or less. Performance in Pictorial and Numeral conditions of a more/less reasoning task would be informative. Students might solve a more/less task using a mental-image structure in order to compare sets in a one-to-one fashion, so performance on this task might be similar to the Equivalence task. On the other hand, students might understand that further along the counting list represents *more*. This would result in similar performance as the Number Series task.

The experiment was conducted in the middle of the academic year. Future studies should be conducted in the beginning and end of the academic year to determine how structures of quantitative reasoning change through instruction. Indeed, first graders in the beginning of the academic year would likely have a similar structure of quantitative reasoning as the kindergarteners in this study. Likewise, kindergarteners' structure of quantitative reasoning at the end of the academic year could possibly be more similar to first graders.

This experiment could also be conducted with samples of low socioeconomic status students. Previous research has established that low socioeconomic preschoolers and kindergarteners tend to perform as well as middle socioeconomic status students on nonverbal measures of quantitative reasoning, but significantly worse on verbal measures of quantitative reasoning (Jordan et al., 1992; Jordan et al., 1994). Therefore, low socioeconomic students in first and second grade might have similar structures of quantitative reasoning as middle socioeconomic status kindergarteners.

*Implications*

Current measures of cognitive abilities tend to use number words and Arabic numerals to assess quantitative reasoning. The purpose of this study was to examine the fundamental assumption that early elementary students use a verbal structure of quantity to reason quantitatively. This assumption was supported in testing situations that evoke the verbal counting structure of quantity, but not in testing situations that evoke nonnumerical quantitative reasoning abilities such as the part-whole schema. Number Series was an appropriate task to use Arabic numerals with early elementary students, but the Equivalence task was an inappropriate use of Arabic numerals. Since the Equivalence task afforded the application of a part-whole schema, a mental-image structure of quantity was more beneficial when solving the task. Early elementary students had not yet merged their mental-image structure with the verbal counting structure, so these students were less proficient on the Equivalence task with Arabic numerals.

As a result, test developers should closely examine their assessments to determine whether their quantitative reasoning tasks afford a counting schema or nonnumerical quantitative reasoning abilities. Tasks that use a counting schema should exclusively use

Arabic numerals (for all students who can read Arabic numerals) or counting words because the inclusion of pictures inhibits performance. On the other hand, tasks that afford nonnumerical quantitative reasoning abilities should be presented in a format that either exclusively uses pictures or supplements Arabic numerals with pictures. Pictures should be used at least through second grade, and further research will have to be conducted to determine exactly how long pictures should be available.

The results of study also provided additional support for current theories of quantitative reasoning in children. In contrast with most other studies in the literature, this study applied a microscopic lens to the quantitative reasoning abilities of early elementary students. The results extended the work by Huttenlocher and colleagues by showing that kindergarteners have a similar structure of quantitative reasoning as preschool students. Furthermore, this study provided additional confirmation for Case's theory of central conceptual structures. Specifically, by closely examining the quantitative reasoning structures in kindergarten, first, and second grade students, this study supported Case's untested hypothesis that the verbal and mental-image quantitative reasoning structures begin to merge when students are about six years of age.

# APPENDIX A

## FORMAT OF STUDY MATERIALS

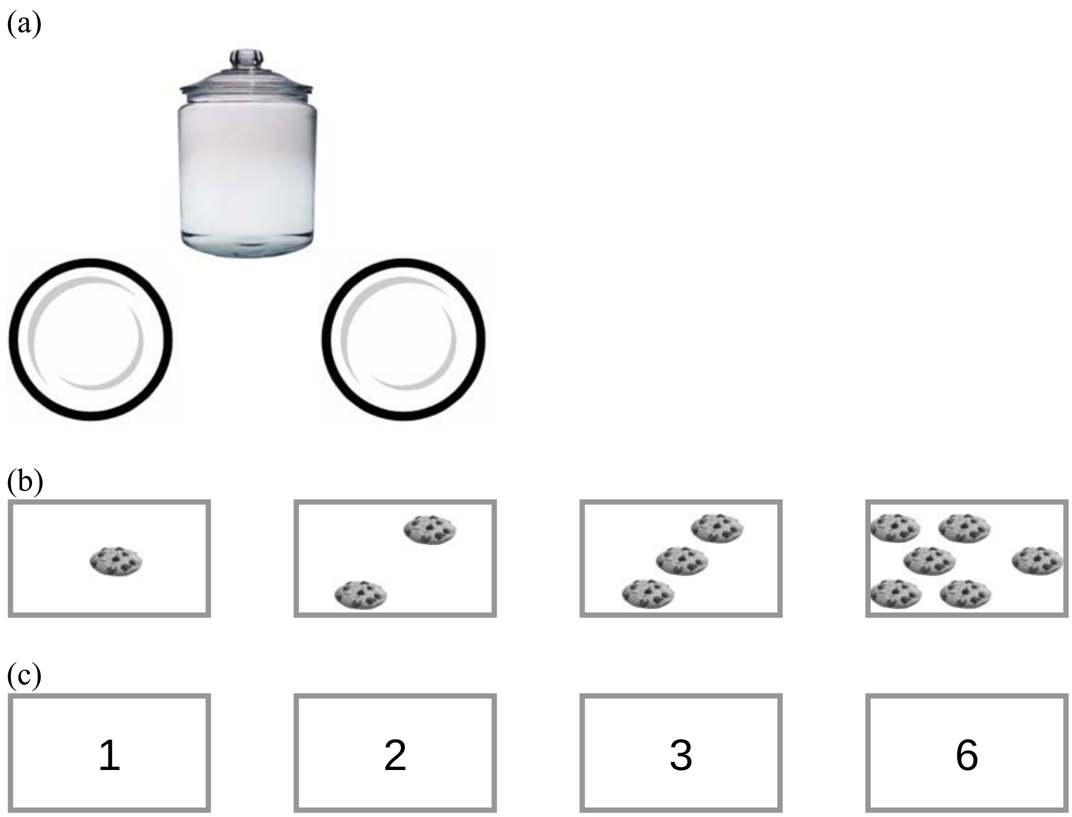
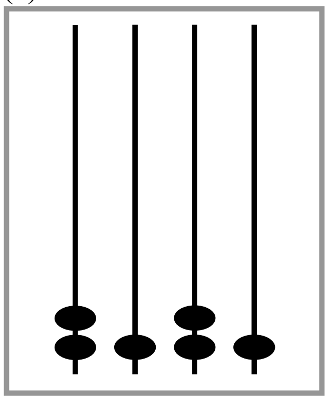
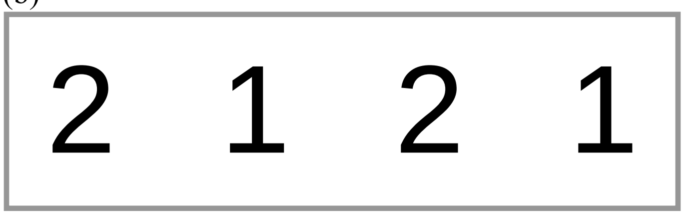



Figure A1. Equivalence display. Row *(a)* illustrates the stem of the item. Row *(b)* illustrates the cards in the Pictorial condition. Row *(c)* illustrates the cards in the Numeral condition.

(a)



(b)




(c)

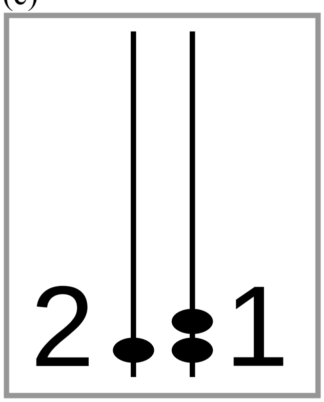



Figure A2. Number Series display. Row *(a)* illustrates a sample item in the Pictorial condition. Row *(b)* illustrates a sample item in the Numeral condition. Row *(c)* illustrates a sample item in the Mixed condition.

# APPENDIX B

# ITEM SPECIFICATIONS

Table B1

Equivalence Item Specifications

| Item | Item Set | | |
| --- | --- | --- | --- |
| Number | Set A | Set B | Set C |
| 1. | 4: 1, 3, 6 | 4: 1, 3, 8 | 4: 1, 3, 9 |
| 2. | 6: 1, 5, 8 | 7: 1, 6, 9 | 5: 1, 4, 7 |
| 3. | 9: 1, 6, 8 | 7: 1, 4, 6 | 8: 1, 5, 7 |
| 4. | 5: 2, 3, 4 | 5: 2, 3, 6 | 5: 2, 3, 9 |
| 5. | 7: 2, 5, 6 | 9: 2, 7, 8 | 6: 2, 4, 5 |
| 6. | 9: 3, 6, 8 | 7: 3, 4, 6 | 8: 3, 5, 7, |
| 7. | 3: 4, 5, -2 | 4: 5, 7, -3 | 2: 3, 6, -4 |
| 8. | 2: 8, -5, -6 | 3: 9, -5, -6 | 2: 7, -4, -5 |

Note. The quantity on the experimenter’s plate is listed first. The quantities that the student received to solve the task are listed after the colon.

Table B2

Number Series Item Specifications

| Item | Item Set | | |
|---|---|---|---|
| Number | Set A | Set B | Set C |
| 1. | 1 3 1 3 | 2 4 2 4 | 3 5 3 5 |
| 2. | 4 4 5 5 6 | 5 5 6 6 7 | 3 3 4 4 5 |
| 3. | 5 6 7 5 6 7 | 2 3 4 2 3 4 | 3 4 5 3 4 5 |
| 4. | 1 1 3 3 5 5 | 2 2 4 4 6 6 | 3 3 5 5 7 7 |
| 5. | 8 7 6 8 7 | 5 4 3 5 4 | 6 5 4 6 5 |
| 6. | 2 1 4 1 6 1 | 3 1 5 1 7 1 | 4 1 5 1 6 1 |
| 7. | 2 1 3 2 4 | 2 4 3 5 4 | 1 5 2 6 3 |
| 8. | 5 6 1 7 8 1 | 2 3 1 4 5 1 | 4 5 1 6 7 1 |

Note. Students received quantities 1 through 9 as the distractors of the task.

# APPENDIX C

# TASK DIRECTIONS AND PRACTICE ITEMS

*Number Series*

Have you ever made a bracelet or necklace out of beads? We are going to play a new game that has strings of beads like that. *Numeral Condition: In this game, there will be some numbers that show how many beads are on a string. Pictorial Condition: In this game, there will be some beads on a string.* To play the game, you need to figure out how many beads should come on the next string. Once you figure out how many beads should come on the next string, find the card that shows that amount and stick it on the board.

I will show you how to do the first one. *Numeral Condition: See, there is one, then two, then three. Pictorial condition: See, there is one bead, then two beads, then three beads.* When we count, we go 1, 2, 3, and 4 comes next. So the next string of beads needs to have four beads. Let's find a card that has four beads. Why don't you try the next one?

**Switching from Numeral to Pictorial:** Now we are going to do the same thing, but the cards will look a little different. Try these cards that have pictures of the beads.

**Switching from Pictorial to Numeral:** Now we are going to do the same thing, but the cards will look a little different. Try these cards that show the number of beads.

**Introducing the Mixed condition:** Now we are going to do the same thing, but the cards will look a little different. There will be both numbers and pictures of beads. Let me show you how to do the first one. There is the number one, then one, two beads, then one, two, three beads, then the number four. So it is one, two, three, four. What comes after four? Five comes after four. Here is another one to show you. There is the number two, then one bead, then two beads, then one. So it is two, one, two, one, and what comes next? Two beads come next.

**Practice Items:**

1, 2, 3<br>
2, 1, 2, 1<br>
4, 3, 2<br>
2, 2, 1, 2, 2<br>
1, 2, 3, 1, 2

*Equivalence*

Do you like to eat cookies? We are going to play a new math game with cookies. Here are two plates. This is my plate and this is your plate. Both plates have to have the same amount of cookies, but we have to put some extra cookies back into the cookie jar over here. *Numeral condition: Here are some cards that show the number of cookies. Pictorial condition: Here are some cards with cookies on them.*

I will show you how to do the first one. Here are four cards. This card has 1 cookie, this card has 2 cookies, this card has 3 cookies, and this card has 6 cookies. These cookies go on my plate. There are one, two, three cookies on my plate. Then I am going to see if I can combine other cards to make 3 cookies on your plate like I have three cookies on my plate. If I put these together, 6 and 2 make 8. Eight is not the same as 3. I'll try to put these together. Six and 1 make 7. Seven is not the same as 3. Is there any other way that we can make 3 cookies? Let's combine one and two cookies. Now there are three cookies! So you see that I have three cookies and you have three cookies. They are the same. Here is another one. Why don't you try this one?

**At the end of the second practice item:** *Pictorial condition: See how this card has four cookies crossed out? Numeral condition: See how this card has the number four crossed out?* If you see a card like that, it means that you take cookies away from the other amount of cookies. So if we would put this card with the card that has 5 cookies, there would only be 1 cookie on that plate because 5 take away 4 is 1.

**Switching from Numeral to Pictorial:** Now we are going to do the same thing, but your cards will look a little different. Try these cards with some cookies on them.

**Switching from Pictorial to Numeral:** Now we are going to do the same thing, but your cards will look a little different. Try these cards that show the number of cookies.

**Introducing the Choice Condition:** Now we are going to do the same thing, but you will get to pick which cards you want to use. The cards over here have the same amounts of cookies as the cards over there.

**Practice Items:**

3: 1, 2, 6
5: 2, 3, -4
3: 1, 2, 8
3: 1, 2, 7
5: 2, 3, -6

REFERENCES


American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Arendasy, M., & Sommer, M. (2005). Psychometric approaches help resolve competing cognitive models: When less is more than it seems. *Cognition and Instruction, 23,* 503-521.

Bakeman, R. (2006). VII. The practical importance of findings. *Monographs of the Society for Research in Child Development, 71*, 127-145.

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America, 102,* 14116-14121.

Benbow, C. P. (1992). Academic achievement in mathematics and science of students between the ages 13 and 23: Are there differences among students in the top one percent of mathematical ability? *Journal of Educational Psychology, 84,* 51-61.

Briars, D., & Siegler, R. S. (1984). A feature analysis of preschoolers' counting knowledge. *Developmental Psychology, 20,* 607-618.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15,* 179-202.

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

Case, R. (1978). Intellectual development from birth to adulthood: A neo-Piagetian interpretation. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 37-71). Hillsdale, NJ: Lawrence Erlbaum Associates.

Case, R. (1993). Theories of learning and theories of development. *Educational Psychologist, 28,* 219-233.

Case, R., & Griffin, S. (1990). Child cognitive development: The role of central conceptual structures in the development of scientific and social thought. *Developmental psychology: Cognitive, perceptuo-motor, and neuropsychological perspectives.* Amsterdam, The Netherlands, Elsevier Science.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61,* (1-2, Serial No. 246).

Case, R., & Sandieson, R. (1992). Testing for the presence of a central quantitative structure: Use of the transfer paradigm. In R. Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge* (pp. 117-132). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

CTB/McGraw-Hill. (2001). *Inview.* Monterey, CA: Author.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179-197.

Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology principles. *Journal of Educational Measurement, 38,* 343-368.

Floyd, R. G. (2005). Information-processing approaches to interpretation of contemporary intellectual assessment instruments. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 203-233). New York: The Guilford Press.

Fuson, K. C. (1988). *Children's counting and concepts of number.* New York: Springer-Verlag.

Geary, D. C. (1994). *Children's mathematical development: Research and practical applications.* Washington, D.C.: American Psychological Association.

Gelman, R. (1978). Counting in the preschooler: What does and does not develop. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 213-241). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number.* Cambridge, MA: Harvard University Press.

Gelman, R. & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13,* 343-359.

Griffin, S. (2003). The development of math competence in the preschool and early school years. In J. M. Royer (Ed.), *Mathematical cognition* (pp. 1-32). Greenwich, CT: Information Age.

Griffin, S. (2004). Building sense with Number Worlds: A mathematics program for young children. *Early Childhood Research Quarterly, 19,* 173-180.

Griffin, S., Case, R., & Sandieson, R. (1992). Synchrony and asynchrony in the acquisition of children's everyday mathematical knowledge. In R.Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge* (pp. 75-97). Hillsdale, NJ: Lawrence Erlbaum Associates.

Griffin, S., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25-49). Cambridge, MA: The MIT Press.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high stakes testing. *Educational Measurement: Issues and Practice, 23,* (17-27).

Halford, G. S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology, 26*, 339-366.

Hudson, T. (1983). Correspondences and numerical differences between disjoint sets. *Child Development, 54,* 84-90.

Hughes, M. (1981). Can preschool children add and subtract? *Educational Psychology, 1,* 207-219.

Huntley-Fenner, G., & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science, 11,* 147-152.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123,* 284-296.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness.* Cambridge, MA: Harvard University Press.

Jordan, N. C., Huttenocher, J., & Levine, S. C. (1992). Differential calculation abilities in young children from middle- and low-income families. *Developmental Psychology, 28,* 644-653.

Jordan, N. C., Huttenocher, J., & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences, 6,* 413-432.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (3rd ed., pp. 17-64). Westport, CT: Praeger.

Keith, T. Z. & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly, 12*(2), 89-107.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127,* 162-181.

Leighton, J. P., & Sternberg, R. J. (2003). Reasoning and problem solving. In A. F. Healy & R. W. Proctor (Eds.), *Handbook of psychology: Experimental psychology* (Vol. 4, pp. 623-648). Hoboken, NJ: John Wiley & Sons.

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology, 53,* 72-103.

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test Form 6.* Itasca, IL: Riverside.

Marini, Z. (1992). Synchrony and asynchrony in the development of children's scientific reasoning. In R.Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge* (pp. 55-73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Marini, Z., & Case, R. (1994). The development of abstract reasoning about the physical and social world. *Child Development, 65,* 147-159.

McGrew, K. S. (2005). The Catell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-181). New York: The Guilford Press.

McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20,* 93-108.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3-62.

Mix, K. S. (1999). Preschoolers' recognition of numerical equivalence: Sequential sets. *Journal of Experimental Child Psychology, 74,* 309-332.

Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood.* New York: Oxford University Press.

National Association for the Education of Young Children & National Council for Teachers of Mathematics. (2002). *Early childhood mathematics: Promoting good beginnings*. Retrieved November 17, 2006, from http://www.naeyc.org/about/positions/pdf/psmath.pdf

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Research Council. (2001). *Classroom assessment and the national science education standards.* Washington, D.C.: National Academy Press

Norman, D. A. (1983). Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp 7-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Okamoto, Y. (1996). Modeling children's understanding of quantitative relations in texts: A developmental perspective. *Cognition and Instruction, 14,* 409-440.

Otis, A. S., & Lennon, R. T. (2003). *Otis-Lennon School Ability Test-Eighth Edition.* Houghton Mifflin Company.

Palmer, S. (2000). Working memory: A developmental study of phonological recoding. *Memory, 8,* 179-193.

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist, 44,* 162-169.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction, 5,* 49-101.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales-Fifth Edition*. Itasca, IL: Riverside Publishing.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46,* 561-584,

Siegel, L. S. (1974). Development of number concepts: Ordering and correspondence operations and the role of length cues. *Developmental Psychology, 10,* 907-912.

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 241-312). New York: Academic Press.

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills: The eighteenth annual Carnegie Symposium on cognition* (pp. 229-293). Hillsdale, NJ: Lawrence Erlbaum Associates.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43,* 93-126.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology, 97,* 484-492.

Wang, M. C., Resnick, L. B., & Boozer, R. F. (1971). The sequence of development of some early mathematics behaviors. *Child Development, 42,* 1767-1778.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content.* Cambridge, MA: Harvard University Press.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition.* San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.