Grouping Gestures Promote Children's Effective Counting Strategies by Adding a Layer of Meaning Through Action

Azadeh Jamalian

Submitted in partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy
Under the Executive Committee
Of the Graduate School of Arts and Sciences
COLUMBIA UNIVERSITY
2014

ABSTRACT

Grouping Gestures Promote Children's Effective Counting Strategies by Adding a Layer of Meaning Through Action

Azadeh Jamalian

Preschoolers can often rattle off a long sequence of numbers in order, but have problems in reporting the exact number of objects even in a small set, and have trouble in comparing numerical relation of two sets that differ by exactly 1 item. The present study showed that representing and highlighting sets by showing a circular, enclosed diagram around them with or without a grouping gesture helps children to enhance their understanding of cardinality and to improve their overall math competence. Nighty-three preschool students, ages ranging from 3years-10 months to 4 years-9months (M= 51.82 months, SD= 3.56 months), from three public schools in Harlem, New York participated in this study. Children from each school were ranked based on their pre-test score on the Test of Early Mathematics Ability (TEMA-3), and were then assigned randomly to one of the three math comparison groups or the reading control group. Children in diagram-plus-gesture math group, were asked to draw a bubble by making a grouping gesture around each of the two sets on a touch screen device, indicate the number of fish in each bubble, and judge whether there were the same number of fish in each bubble, and in case the number was not the same, indicate which set had more fish. Children in the diagram only condition simply saw bubbles around sets without the need to do a grouping gesture around them. Children in the no diagram- no gesture condition neither saw a bubble nor did a grouping gesture. All participants played on the software for 4 sessions within a two-week time period and the data were examined microgenetically. Results showed that all children in the math comparison groups improved in their math scores during the

game-play and improved in their overall math competence from pre- to post-test, unlike the children in the reading control group. More importantly, children who saw the circular diagram (bubbles) around sets with or without the grouping gesture outperformed children who never saw bubbles nor made a grouping gesture in their accuracy, understanding of cardinality, and overall math competence from pre to post. Further, children with lower executive functioning skills benefitted from performing the grouping gesture in addition to seeing the circular diagram. Gestures can have the same form as diagrams, and hence, they may carry information that is redundant with diagrams. Such redundancy reinforces the message by presenting information in two modalities— a redundancy that may not be necessary for some, but beneficial to others (i.e. children with low executive functioning skills). Finally, over the course of game-play children who did the grouping gesture never counted the two sets together as one set when asked to compare their numerical relation— a mistake many preschoolers make; children in the other groups made that mistake occasionally. Because gestures are actions and dynamic by nature, they appear to be especially suited for changing actions and promoting early counting skills.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful to my advisers, Prof. Barbara Tversky and Prof. Herbert Ginsburg, for their invaluable guidance and unwavering supports. I like to thank Barbara for being a truly inspiring advisor, for continually conveying a spirit of adventure and excitement in my academic life, and for helping me grow as a researcher. I like to thank Herb for providing me with the wisdom and comfort I needed in the challenging times, for always showing me the best path to succeed, and for encouraging me to undertake new ventures.

I am also thankful for Prof. Matthew Johnson, who has always welcomed my visits to his office, has patiently listened to my questions, and has helped me to solve all those seemingly unsolvable statistics problems. And a special thank to my committee members, Prof. Catherine Chase and Prof. Koleen McCrink, for letting my defense to be an enjoyable experience, and for their insightful comments and suggestions. This work (and many more to follow) is a result of many insightful conversations I had with each one of them.

I would like to thank Institute of Education Sciences (IES), National Science Foundation (NSF), Cleveland Dodge fellowship, and Zankel fellowship programs for providing me with generous financial support throughout my doctorate degree. I also like to express appreciation to Teachers College for providing scholarships, fieldwork funding, and conference funding.

Grad school would have not been this enjoyable without the amazing Mac7 team, Esther Yoon, Kara Carpenter, Sam Creighan, Dana Pagar, Rachael Labrecque, Ama Awotwi, and Ben Friedman. I like to thank them for helping me to grow as part of our MathemAntics adventure, and for all their help in designing the study, collecting the data, and for the vibrant discussions about the findings. I also like to thank my great friend and game design instructor, Jessica Hammer, who inspired me to think of myself as a designer and encouraged me to take risks and to enjoy failure. And, a big thanks to Valeria Giardino, a true friend and a fantastic colleague who undertook with me the exciting research on gesture and diagrams, and helped me to think more deeply about those topics. I also like to thank my research assistants at Teachers College who helped me in collecting, cleaning, and analyzing data for this project.

I want to thank my amazing husband, Ala Alavi, who has shared with me all ups and downs of the doctoral journey, and has always been there for me with the warmest heart. I like to thank him for his constant love, for pushing me to the finish line, and for all the enjoyable memories along the way. I am grateful to him for making our home feel like a real home—somewhere we can both grow and enjoy life to the fullest.

I like to thank my wonderful brother, Arash Jamalian, for his incredible support throughout my life. It is remarkable how many times I went back to our childhood memories while writing this dissertation. Memories of playing games and coming up with random rules of play; memories of him totally understanding why 2 x 2 is 4 and at the same time 2 to the power of 2 is 4, and I be puzzled on how his little brain works! After all, he may be the reason I am obsessed with how kids learn math.

A special thank to my mother, Shabnam Pam, for not making a big deal out of my life accomplishments, and for always pushing me to do the next big thing, because in her eyes, I can always be better! “Great that you’re getting done with your PhD. How about a post-doc?” I want to thank her for being the kindest.

I like to thank my cheerful friends, Maryam Taghavi, Nazanin Sadr Azodi, Esther Yoon, Ali Emami, Jessica Mezei, and Morteza Dehghani, for their moral support that has made all the difference. Each one of them knows how they have brought meaning into my life, and how they have made it a happy one. A special thank to my lovely friend Maryam, who showed me how to push the norms.

I also like to thank my in-laws, Mojgan Mansoorivar, Fereidoon Alavi, and Alia Alavi for always encouraging me to pursue my dreams. They have shared with me all the stress I went through when collecting data and listened to all my complaints about the hurricane sandy.

Finally, I would like to thank my colleagues at Tiggly, Phyl Georgiou, Bart Clareman, Paul Zdanowicz, Michal Eliakim, Jenny Armstrong, Andrea Mascaro, and Marie Jeanne Lanum, for all their supports along the way, and for showing me that there is an exciting life awaiting me after the grad school.

To my dad

CHAPTER 1: THEORETICAL BACKGROUND

Introduction

Parents, and teachers too, are always surprised when their children can rattle off a long sequence of numbers in order, but have problems in reporting the number of objects even in a small set, like the number of fingers on one of their hands. Knowing the order of numbers is not the same as knowing how many, and knowing how many is not the same as knowing if two sets are equal in number. Children may first engage in counting as a rhythmic routine, more like a purposeless activity, and only gradually learn how it is related to enumeration and cardinality (Fuson, 1988, Wynn, 1990). It may even take longer for them to learn that counting is useful when comparing sets and when creating new sets to match existing ones (Michie, 1984; Sophian, 1987, 1995), and once they notice that counting is a useful strategy, it may be challenging for them to use it effectively. The present study investigated whether set representation through gestures and diagrams help children to relate counting to cardinality and to use their counting skills effectively when comparing numerical relation of two sets.

Counting Principles and Cardinality

An extensive body of literature (see for ex. Fuson, 1988; Wynn, 1990, 1992; Gallistel & Gelman, 1992; Gelman, 1993; Schaeffer, Eggleston, & Scott, 1974) has shown that the task of enumeration is substantially more complex than reciting the counting words and requires the understanding of several underlying principles. Gelman and Gallistel (1978) propose three principles that underline counting and enumeration: (1) stable order principle, stating that symbols have a consistent order across counting occurrences, (2) the one-to-one correspondence principle, stating that for every object in the counting set only one and only one counting symbol is applied and each counting symbol is only applied to one object, and (3) the cardinal principle, referring to the point that the last symbol of a count represents number of objects in the set that has been counted. Gelman (1993) argues that mastering these non-verbal counting principles precedes children's acquisition of verbal counting skills. Evidence for this proposal comes from the finding that children are often good at detecting a puppet's counting mistakes even when the puppet miscounts sets with cardinal values beyond their counting range (Gelman, 1993; Gelman and Meck, 1983, 1992). Further evidence for the "principles-before-skill" proposal comes from the observations that children sometimes come up with their own counting routines, different from the typical counting procedures adopted by their parents (Gelman & Gallistel, 1978). For example, they sometimes use lists other than numbers, such as alphabets, to count things, or they use a count list with numbers ordered differently (but with stable order) than the established count list. This evidence leads Gelman to argue that children may have difficulty in using counting effectively to recognize cardinal value of sets because of their limited understanding of what they are being asked to do (i.e. "utilization skills"), and/or their limited ability to plan and execute counting procedures to solve problems (i.e. "procedural skills") (Greeno, Riley, & Gelman, 1984).

However, there is a counterargument that children's difficulty in using counting to identify cardinality of sets is due to their inadequate knowledge of the actual meaning of the number words, as well as their limited conceptual understanding of counting as a possible strategy to solve problems. For example, Schaeffer, Eggleston, and Scott (1974) found that many children (mean age 3.5 years) who can successfully count arrays of 5 to 7 objects nevertheless failed to indicate the cardinal value of the set when the set was covered and asked how many objects there were hidden. Wynn (1990, 1992) also found that for about a year after children memorized a short count list, they rarely used counting to solve problems that required determination of the cardinality of a set, and that they only knew the exact meaning of a subset of the number words in their count list. For example, in Wynn's "Give a Number Task" (GN) children were asked to give a puppet one, two, three, five, or six toys from a pile. Wynn found that only older children (mean age 3:6) counted the objects and were able to give the puppet the exact number of toys he had asked for. Wynn called this group of children "Cardinality Principle knowers". Younger children, on the other hand, rarely used counting and were able to only give a subset of items in their counting list. Wynn called this group of children "sub-set knowers." In a more recent study, Le Corre, Van de Walle, Brannon, and Carey (2006), showed that "Cardinality Principle knowers" spontaneously counted more frequently than the sub-set knowers when asked to give a set of items, and that sub-set knowers were more likely to count when asked for small numbers than when asked for the larger numbers (4 or more). Interestingly, "one"- and "two"-knowers who sometimes counted never ended their counts at the target. Further, when asked to "fix" their sets with incorrect number of objects, "Cardinality Principle knowers" were able to do so by matching it with the exact requested number, while "sub-set knowers" sometimes left their set unchanged or changed them in the wrong direction. They also found similar results when they tested same children with a simpler task, namely "What's on this card?" (WOC) task adopted from Gelman (1993). In the WOC task children were probed to respond with both counting and a cardinal value when asked to indicate what they see on the shown card; of interest is whether children indicate the numerosity of the items on the card rather than other properties such as color. Gelman (1993) had previously argued that the WOC task is less taxing on children's performance skills and is more true to children's earlier numerical competence than Wynn's GN task. However, when Le Corre et al (2006), examined children's performance on the WOC task as a function of their Wynn stage, they found that although the

WOC task demanded less procedural skill, sub-set knowers still engaged in spontaneous counting less frequently than the cardinality principle knowers when shown cards with larger number of items.

Thus, there is evidence that children have difficulty in using counting for numerical reasoning such as responding with a cardinal answer in the WOC task, or producing a set with a given numerosity in the GN task; this difficulty may be due to children's limited understanding of counting as a useful strategy to evaluate cardinality of a set and/or due to their limited procedural skills. In a different paradigm, researchers have also found that children are often reluctant to count spontaneously when they compare the numerical relation of two sets. For example, Sophian (1987) found that 3.5 year-olds rarely counted spontaneously when they were shown a set of balloons and a set of clowns in separate trays and were asked, "Are there enough balloons for every clown to get one?" In a second experiment, similar hesitation to counting was found when 3.5 year-olds were asked to put just enough balloons onto a round tray so that every clown on a rectangular tray could get one, although they readily count when asked to determine how many objects there were on a tray. Sophian also noticed that use of counting increased significantly between 3.5 and 4.5 years. Similarly, Michie (1984) found that unless children were specifically asked to count, they rarely used counting to compare the numerical relation between two rows of red and blue dots, presented to children as buttons belonging to two dolls; instead they relied on perceptual cues such as relative length of the rows to make their judgment. However, when they were asked to count, both 3- and 4- year olds became more accurate in their judgments, implying that they actually used information derived from counting to compare numerical equivalence of the two sets. Similarly, Mix (1999) found that children are significantly less accurate in matching two equal but spatially different sets than matching two sets with congruent numerical and spatial properties. In fact, we see similar trend in Piaget's classical numerical conservation task. Four and even five year olds, including those who can count, typically fail in conserving number when the spatial arrangement of a set changes, even in the child's view (Piaget, 1952).

Michie (1984) argued that preschoolers' preference to rely on length cues as opposed to counting to compare numerical relation of two sets is due to their belief that length is a more reliable cue than counting, rather than their limited understanding of counting as a useful strategy for number comparison. She based her argument on the evidence that four year olds engaged in more spontaneous counting when they were given feedback on the correctness of their judgments. From Siegler's (1987) research on children's strategy use, we know that children adopt more efficient, faster strategies when they yield accurate solutions and only adopt slower strategies if they significantly increase accuracy. Since counting is a slower strategy than visual comparison, children may adopt the advanced but slower counting strategy only if they know counting actually increases their accuracy.

However, we know children are also reluctant to count even when visual cues are absent, as is the case in the WOC task or the GN task; thus, it seems that a number of different mechanisms may underlie their failure to count when they compare sets: (1) limited conceptual understanding of exact numbers and of cardinality as a property of a set rather than an individual item, (2) ignorance of the link between counting and cardinality, (3) their limited procedural skills, (4) and their belief that spatial properties are better cues for comparison than counting.

Core Systems for Mathematical Cognition

An extensive body of literature (see Feigenson, Dehaene, & Spelke, 2004; Spelke, 2011 for reviews) suggests that the human numerical cognition is supported by two innate core systems: the Approximate Number System (ANS) for comparing and combining the analog magnitude or approximate cardinal values of sets, and the attentive object tracking system for tracking small numbers of distinct units. These core systems are innate, shared with our mammalian ancestors, and are automatically engaged in domain specific contexts— each with certain limitations.

Studies of common laboratory animals, such as the pigeon, the rat, and the monkey, suggest that these animals can estimate and remember continuous quantities by mental magnitudes that are not exact (King, McDonald, & Gallistel, 2001). Moyer and Landauer (1967; 1973) were the first to suggest that humans, similar to the laboratory animals, represent integers (discrete quantities) with mental magnitudes (continuous quantities). In a series of studies they asked adults to judge the numerical order of two Arabic numerals as fast as they can. They found that it takes more time for an adult to judge the order of two numbers, the less the distance between the two numbers is. For example, it takes more time to order 100 and 101 than 50 and 100. This effect has come to be known as the *symbolic distance effect* and has been replicated by many other researchers since then (Dehaene, 1997; Gallistel & Gelman, 2005). Studies on newborn infants also show that they can reliably discriminate between two sets that differ by a ratio of 3 but are far less accurate when comparing sets that differ by a ratio of 2 (Izard, Sann, Spelke, & Streri, 2009). Six-month infants are reliable in discriminating sets by a ratio of 2 (Lipton & Spelke, 2003; Lipton & Spelke, 2004; Xu, Spelke, & Goddard, 2005), and ten-month olds can discriminate sets by a ratio of 1.5 (Xu & Arriaga, 2007). Although this critical ratio drops as infants develop, it never reaches 1. Even when adults are asked to compare two large sets quickly without counting (Dehaene, 1997), their discrimination accuracy decreases as the numeric ratio between the two sets approaches 1, and the variability of their estimation of magnitude increases as the cardinality of the set increases. This is known as Weber's fraction limit.

Not only can infants compare the approximate magnitude of large numbers, but they can also combine and manipulate approximate quantities. In one study, McCrink and Wynn (2004) grouped nine-month old infants into two conditions, "addition" and "subtraction." Infants in the addition condition saw an array of 5 objects on screen, moving towards a white computerized occluder and covered behind it. They then saw another array of 5 objects appearing on screen and getting hidden behind the white occluder, which then moved off the screen to reveal either 5 or 10 objects. Infants in the subtraction condition saw an array of 10 objects being covered behind the white occluder, and then 5 objects emerged from behind the white occulder, which then moved off the screen to reveal either 5 or 10 objects. Interestingly, infants in the addition condition looked longer at the array of 5 objects (incorrect outcome) than at the array of 10 objects (correct income) and infants in the subtraction condition looked longer at the array of 10 objects (incorrect outcome) then at the array of 5 objects (correct income). These results indicate that nine-month old infants have some form of intuitive understanding of adding and subtracting approximate magnitudes.

The evidence thus suggests that an innate system that operates on approximate representation of quantities is responsible for comparing and combining large sets-- a system that is called the Approximate Number System (ANS). ANS represents and operates based on approximate numerical magnitude of sets, is subject to Weber's fraction limit, and fails to make explicit individual items within the set that it enumerates; thus it fails to support representations such as exactly six. In fact, research shows that when individual items are shown to an infant (instead of the whole set at once), their numerical discrimination is impaired or eliminated, perhaps because of inhibition of the operation of ANS (Feigenson, Carey, & Hauser, 2002). Feigenson et al (2002) had ten- and 12-month old infants see crackers sequentially placed into two containers and then crawl towards a container of their choice. Infants crawled towards the container containing 2 crackers but not 1, and towards the container with 3 crackers but not 2. However, they failed to discriminate between containers with 4 versus 3 or 6 versus 3 crackers. This finding is surprising given that even six-month olds are shown to be able to spontaneously discriminate between 6 and 3 (Lipton & Spelke, 2003). Feigenson et al explain that presenting items one by one in a sequential manner inhibits operation of the ANS and instead activates another core system that attends to individual items- a system called “parallel individuation” with a structure similar to working memory.

The “parallel individuation” system is exact and attends to the individual items within a set, but it only represents sets of up to three to four numerically distinct individuals (Carey & Xu, 2001; Feigenson & Carey, 2003, 2005; Hauser & Carey, 2003). Children and adults can recognize a set of 2 or 3 items (and in some cases sets of 4) without the need to count them, but when presented with a set of 5 or 6 items, to evaluate the exact cardinal value of the set they need to count the items in the set otherwise they can only approximate the cardinal value of sets; hence, engaging their ANS system. Thus, the exact “parallel individuation” system is activated under conditions complementary to the conditions that activate the ANS. In cases where small number of objects are attentively tracked, ANS is inhibited but “parallel individuation” system is activated, and cases in which sets of large items are presented at once (without the chance to count them) exhaust the size limitation of the “parallel individuation” system but elicit activity of the ANS system.

Conceptualization of Exact Large Numbers

If our mathematical cognition is based on these two core systems, one with a ratio limit on the representation of cardinal values and the other with a set size limit on the representation of individuals, then how can we conceptualize more advanced math concepts such as exact quantities of large sets or discriminating between sets with 100 and 101 items?

One view is that children acquire the counting principles (or cardinality) by gradually mapping numbers in their count list to their approximate analog magnitudes and hence associating numerical meaning to their otherwise meaningless count list (Dehaene, 1997; Wynn, 1992). Although the analog magnitudes are approximate, they can highlight an important property of the count list, namely that "later in the list means larger set" (Wynn, 1992) which could lead to the acquisition of the successor function-- any natural number, n, is followed by another natural number, n + 1. Evidence for this view comes from the observations that adults' estimation of large sets have the properties of the ANS system such as the Weber fraction limit and scalar variability (Dehaene, 1997; Izard & Dehaene, 2008; Whalen, Gallistel, & Gelman, 1999), and that even preschoolers who count up to 100, can also estimate the numerical value of sets of 100 objects, an estimate that linearly relates to the numerosity of the set, (Lipton & Spelke, 2005), confirming that they have mapped numbers in their count list to their approximate analog magnitudes.

Other scholars agree that we eventually map numbers to their analog magnitudes, but argue that such mapping is not the key in acquiring the counting principles (Carey, 2004), and some even argue that the mapping only comes *after* children become cardinality knowers (Le Corre & Carey, 2007). Carey (2004), for example, explains that children acquire counting principles by initially mapping numbers in their count list to the representations that arise from the "parallel individuation" core system and "set-based quantification." Based on this view, children first map number "one" to the representation of exactly 1 object, and then number "two" to the representation of any sets containing exactly 2 objects, and so on. In one study, Le Corre and Carey (2007) showed that children map numbers beyond 4 to their approximate magnitudes about six months after acquisition of counting principles. They asked 3 and 4 year olds to estimate the cardinality of sets of 1-10 items and realized that they could group cardinality knowers based on their accuracy of estimation into two groups: "non-mappers" who could successfully pass the Wynn task but could not map numbers beyond 4 to their analog magnitudes, and "mappers" who could reliably map numbers beyond 4 to their analog magnitudes. Mappers had an average age of 4.6, about six months older than the non-mappers. In another study, Le Corre (2013) directly asked children who have recently become cardinality knowers to compare numbers on the basis of verbal information alone. He asked these children, "Which box has more fish inside; this one with ten fish or this one with six fish?" and found out that "non-mappers" who have recently acquired counting principles can reliably compare one versus eight and two versus three but when comparing six versus ten or eight versus ten their accuracy was close to chance. On the other hand, "mappers" could reliably compare "eight" versus "ten" suggesting that they have mapped all numbers in their count list to analog magnitudes.

A third view (Feigenson, Dehaene, & Spelke, 2004; Hauser & Spelke, 2004; Spelke 2011) argues that counting principles arise when the representations from the two core systems could be productively combined. If children could systematically combine representations of sets and their cardinal values (representations from ANS system) with representations of numerically distinct individuals (representation from parallel individuation system) they could overcome the limitations of these two sets and hence be able to conceptualize ideas such as “exact quantities” and compare sets based on their exact cardinal values. Further, they argue that language plays a key role in combining the core systems to overcome their limitations. Spelke (2011), for example, explains that children combine the two systems when they begin to use number words in their everyday language and when they count. As suggested by many researchers (Fuson, 1988; Wynn, 1990, 1992; Sarnecka & Carey, 2008; Spelke, 2011), initially these number words seem to have little numerical meaning attached to them, but they are simply elicited by the presence of a collection of objects and are accompanied by pointing gestures to individual objects. Collection of objects activates representations of approximate cardinal values (i.e. ANS core system), and the pointing gestures are likely dependent on the attentive object tracking system.

For children who are learning to count, gestures play an important role in their counting proficiency. Preschoolers count objects most accurately when they gesture as they count (Carlson, Avraamides, Cary, & Strasberg, 2007; Fuson & Hall, 1983; Gelman & Meek, 1983; Saxe & Kaplan, 1981; Schaeffer, Eggleston, & Scott, 1974). For example, Alibali and DiRusso (1999) found that preschoolers were more accurate in counting sets of 7 to 17 objects when they gestured or when a puppet gestured for them, than when they were prohibited to gesture. Further, when children gestured, they were more likely to make errors in keeping track of the objects, but when the puppet gestured they made more errors in coordinating saying the number words with tagging the items to be counted. Thus, pointing gestures help children to limit two distinct difficulties associated with counting: keeping track of individual objects while counting them and coordinating saying the number words with the action of tagging items (Alibali & DiRusso, 1999).

However, it is possible that repeated pointing gestures accompanying counting impede children's understanding of the cardinality concept. Children typically produce pointing gestures to indicate objects in the environment (Bates, 1976; Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979) and there is a strong association between pointing gesture and acquisition of new vocabulary for objects (Iverson & Goldin-Meadow, 2005). In fact, preschoolers may interpret pointing gestures accompanying counting as suggesting the need to name some property of the individual objects rather than tagging them to evaluate the numerosity of the whole set. For example, a child counting five apples may point one by one to each of the apples, recite number words one to five, interpret the pointing gesture as a way of reciting property of the individual apples as one, two, three, four, five, and fail to realize that five refers to the cardinality of the whole set rather than a particular apple. Further, pointing gestures individuate items in a collection and are likely to be dependent on the attentive object tracking system (Spelke, 2011). Therefore, although pointing gestures may assure an accurate counting, they fail to represent sets as a collection with abstract properties such as a cardinal value.

Representing Sets Through Gestures and Diagrams

Feigenson (2011) argues that set representations play a critical role in representing the concept of cardinality, among many other numerical processes. Indeed, a recent study (Sarnecka, 2007) shows that children who learn languages that explicitly distinguish between singular and plurals seem to learn the cardinal meaning of the number word "one" much earlier than children who learn classifier languages such as Japanese. In fact, the current study makes an emphasis on *set representation* and argues that highlighting sets and at the same time attending to individual objects within that set help children to learn cardinality. More specifically, we investigate whether representing sets through gesture and diagram help children to learn about cardinality and to compare equivalence of two sets that are different by only 1.

While pointing gestures help keeping track of individual items, other gestures such as a single pointing gesture to the set or a grouping gesture around the whole set may help to represent sets and assist in conceptualizing cardinality of sets. Interestingly, Fuson (1988) used the grouping gesture as part of an intervention she did with children struggling to indicate "How Many" items are indicated by a count. After three trials of training, in which she instructed children saying, "When you count, the last word you say tells you how many things there are. Watch me, one, two, three, four, five, six. Six, there are six blocks" [she made a circling or grouping gesture when saying the final statement], 8 out of 13 children succeeded in answering the "how many" question. Although the results are promising, many confounding factors may contribute to that result. In fact, it is not clear whether the grouping gesture *per se* had a facilitating function in the development of children's understanding of cardinality. Instead, the first statement "When you count, the last word you say tells you how many things there are" may suggest that children attend to a rule rather than truly comprehend the meaning of cardinality. In other words, children may simply repeat the last word they counted when responding to the "how many" question without knowing that the last word actually indicated the cardinality of the whole set. Further, the modeling of accurate counting and repeating the last count word that accompanied the grouping gesture may have helped children with their knowledge of cardinality and not the grouping gesture itself.

In a more recent study, Suriyakham (2007) compared effects of the pointing + grouping gesture and pointing gesture alone on the development of children's cardinality knowledge. Surprisingly, she did not find any significant differences between the two kinds of gestures in improving children's performance on either the GN task or the WOC task, both measuring children's understanding of cardinality of a set. Different factors may have contributed to this finding. First, the variability in her dataset was large, and that may have impeded detecting a significant difference between the two conditions. Second, children in the pointing condition may have learned the abstract rule of repeating the last word of their count as a response to the cardinality tests without knowing the actual meaning of cardinality as a property of sets. Therefore, the measures used by Suriyakham (2007) may have not been sensitive enough to detect the benefits of the grouping gesture over the pointing gesture alone in facilitating children's understanding of cardinality as a feature of the set. Third, very few children engaged in overt counting behavior in Suriyakham's study, and as she suggested, strategy probes such as "How did you know to give Elmo X toys?" or "Show me how did you know when Elmo had X toys?" may have yielded more insight on children's mental representation of what the task entails (see Ginsburg, 1997, on the importance of follow up questions in the spirit of the clinical interview).

In a previous study (Jamalian, 2013), we asked children to show how they know their answers once they indicated if two sets with cardinal values between 5 to 9 were equal in number; however, even with the strategy probe, only half of the preschoolers (17 out of 36) used counting properly to compare the sets at pre-test. Other students either never counted or only counted one of the two sets or counted both sets together as one whole set. On the other hand (so to speak), three sessions of intervention in which children viewed a grouping gesture around the items in each set after repeated pointing gestures one by one to the items in the sets promoted effective use of counting. During the intervention, children were asked to compare the number of eggs of two hens displayed on a computer screen, and were given feedback on their responses. At the beginning of each trial, the experimenter asked the child if the two hens have the same number of eggs, and while asking, either pointed to the two hens (i.e. point to owner gesture), pointed to the eggs one by one, (i.e. pointing gesture only), or made a grouping gesture around each set after pointing to the eggs in each set one by one (i.e. pointing + grouping gesture). Children also saw the experimenter modeling two questions and performing the corresponding gestures at the beginning of the second and third sessions. The results demonstrated that children who saw the pointing + grouping gesture were more likely to use counting effectively when comparing cardinal values of two sets in a paper and pencil task at post-test than students who only saw the pointing gesture or point to the hens gesture. Children in the pointing + grouping gesture condition were also more likely to successfully use counting to make two sets equal at post-test.

Why is the grouping gesture effective? A grouping gesture has a different form than the counting pointing gestures, and this differentiation in form makes a distinction between counting as a repeated one-to-one tagging process and evaluating the cardinal value of the whole set, the group of items, as the result of that counting process. Importantly, the grouping gesture does not have an arbitrary form; it is an enclosed circle, like a container with an interior and an exterior putting a border around the items of the set. Lakoff and Nunez (2000) argue that we often conceptualize sets as containers, as if they contain their items *in* them. The “container” *image schema* is an everyday primitive spatial relation concept that links our everyday visual-motor experiences to the conceptualization of many mathematics concepts. For example, Tversky (2011) argues that bars in a bar graph are containers that contain similar instances, separating their content from dissimilar instances *in* other bars. Box tool in the MathemAntics’ “How Many” software activity are containers, holding counted animals inside them, separating them from the ones to be counted (Ginsburg, Carpenter, Labrecque, & Pagar, 2013). Classes are perceived as containers; sets are perceived as containers; the Cartesian plane is a container, containing figures, lines, and points *in* it. Talmy (1983) prefers to use the term “surround” to label the literal and metaphoric meaning of the linguistic proposition *in*, but agrees that that these fine linguistic structures, such as *in*, *on*, *out*, *over*, *across*, provide a systematic framework that schematizes space and uses spatial relations to convey meaning. Schematization, he explains, “is a process that involves the systematic *selection of certain aspects* of a referent scene to represent a whole, while *disregarding the remaining aspects*” (p.225, *italics* added). For example, we, as adults, select interior, exterior, and boundary aspects of a container and disregard its size or volume to conceptualize a group of items as a whole set, an object in itself that may have distinct properties such as a cardinal value. Like language, gestures schematize and abstract; unlike language, gestures represent and organize information directly through actions in space, in space, using relations in space. Natural mapping between meaning and the representation of the meaning promotes thought (Tversky, Kugelmass, and Winter, 1991; see also Tversky, 2011a). A grouping gesture represents sets by grouping things and putting a border around them separating them from other things that are not part of the set. The representation naturally maps to the container schema.

Diagrams too use space (of a paper or of the computer screen), position, and form in space to represent meaning, to “visualize thought” as Tversky (2011) puts it. For example, a circular diagram around items of a set represents the container schema, putting a boundary around members of the set, separating the interior from the exterior, and from other items that do not belong to the set. Tversky (2011) explains that diagrams can be seen as the consequence of gestures, the drawing actions. Enfield (2003) argues that gestures establish “virtual diagrams” in space. Gestures as virtual diagrams or diagrams as crystalized gestures, share representative features like position and form in space to convey meaning. However, diagrams are static, while gestures are dynamic; they have an added channel to represent meaning through action (Tversky, Jamalian, Giardino, Kang, & Kessel, 2013).

Present Study

Does this added layer of action help children to better learn about cardinality? In our previous study, we demonstrated that viewing a grouping gesture promotes effective counting strategy to compare numerical relation of two sets. Would viewing a circular diagram be just as effective? In the present study, we test this by having children either see a circular diagram around the sets or make a grouping gesture around the set to form and see a circular diagram when asked to compare the numerical relation between the two sets. Using pre-test post-test comparisons, we investigate whether either of these manipulations is more effective in enhancing children's understanding of cardinality. We use several measures: the GN task; their accuracy and effective use of counting when judging numerical equivalence of two sets, measured by an Equivalence Judgment (EJ) task; when producing a set equal in number with another give set, measured by the Give the Same (GS) task; and when making two sets equal by adding or removing items from them, measured by the Make the Same (GS) task.

In addition to enhancing meaning through action, gestures may play a special procedural role. In our previous study, we noticed that children who engaged in counting to compare numerical equivalence of two sets, often counted the two sets together as one whole set. Such error may be due to their limited understanding of the link between counting and evaluating cardinality of sets, and/or it may be due to their limited procedural skills. Children need to be able to count one set, stop, and start from the number 1 to count the other set; however, due to their limited inhibitory skills, children may find it difficult to interrupt their counting routine and start from number 1. Their failure to stop counting and start from 1 for the second set may also be due to their limited cognitive flexibility: they may find it difficult to think about the concept of sets while counting at the same time. In other words, they may be aware that counting is a good strategy to compare sets, but find it difficult to use their counting skills purposefully. A grouping gesture may help children to focus on items in only one set at a time and interrupts their counting routine by performing an action that has a form different than the repeated pointing gestured often used in the counting process. If that is the case, then the grouping gesture may be more beneficial to children with lower executive function skills than those with higher skills. We test this hypothesis by analyzing children's improvement in accuracy and strategy use as a function of their executive function skills, measured by the Head-Toes-Knees-Shoulder task (HTKS)- HTKS task (Ponitz, McClelland, Matthews, & Morrison, 2009).

As Tversky (2011) nicely points out, "before page [or a computer screen] there was space itself." We group similar things in space in close proximity to each other, away from dissimilar ones, a "separated spatial grouping" that indicates different categories and themes. For example, we separate the pile of unwashed laundry from the clean ones; the pile of read papers from the unread ones; and private things from public ones. Similarly, items in a set may be grouped by placing them close to each other further away from other things. By grouping items of a set in space using proximity, we reflect the Gestalt laws of perception, representing a whole, that is, the set, which is greater than or different from its parts, the members of the set. Thus, using spatial proximity may be just enough of a manipulation to help children to conceive items as a set. We test this by adding a third condition to the present study, in which children never see a circular diagram around the sets nor do they make a grouping gesture around the sets, but the items of each set are spatially arranged in close proximity of each other away from the items in the other set.

In summary, the present study tests whether the spatial schema of a container as a representation of sets, represented visually by a circular diagram around the set and tangibly by performing a grouping action around the items in the set, can help children to (1) conceive of cardinal values as an emergent property of a set rather than a property of the last counted item in the set, (2) perceive counting as a useful strategy when comparing sets and creating equal sets, and (3) use their counting skills more effectively to judge numerical equivalence of two sets, to create a set matching in number with a given set, and to make two sets equal in number by adding or removing items from them. We test whether the grouping gesture enhances effectiveness of the intervention beyond the visual circular diagrams around the sets by adding a layer of meaning through action and/or by enhancing children's procedural skills. Finally, we examine whether spatially grouping items in each set using proximity is as effective as enclosing them in a circular diagram with or without a grouping gesture.

Children often express ideas in their gestures prior to expressing them in their language (see Golden-Meadow & Alibali, 2013, and also McNeill, 1992), and those gestures may have a causal role in their language development (Rowe & Goldin-Meadow 2009; Rowe, Özçalişkan, Goldin-Meadow, 2008). If children can show ideas in their gestures before they are ready to express them in their speech, maybe bringing new gestures into their "gesture vocabulary" helps them to learn about complex concepts before they are ready to learn about them through language input. In fact, we may be able to introduce children to complex ideas through gesture before we can discuss those ideas with them in words. Thus, studying effects of gesture on thought is particularly important for early childhood education as it may open new possibilities for introducing concepts to young children at a stage when they are not ready yet to comprehend them through symbolic language, or perhaps through diagrams.

CHAPTER 2: METHODS

Participants

One hundred-twenty-three pre-school students (56 girls) mostly from low socioeconomic status (SES) African American and Latino families, ages ranging from 3years-10 months to 4 years-9 months (M= 51.82 months, SD= 3.56 months) were recruited to participate in the study. All children were recruited from three public pre-schools in Harlem, New York. Fifty-six students were recruited from school A, 48 from school B, and 19 from school C. A total of 21 students (8 from school A, 9 from school B, and 4 from school C) were screened out of the study due to language issues, behavior problems, or the child's resilience to participate. Moreover, 7 students (3 from school A, 1 from school B, and 3 from school C) had perfect scores for all the pre-test measures and did not participate in the intervention. Thus, in total 93 students participated in the intervention. Participants were recruited through fliers sent home by their teachers and through parent meetings at schools that researchers attended. As part of the consent forms, parents were asked to check off whether they consent to their children being video and audio taped, and whether their videos could be used in an educational setting.

Tasks

Pre-test Tasks

Head-Toes-Knees-Shoulder task (HTKS)- The HTKS task (Ponitz, McClelland, Matthews, & Morrison, 2009) was administered to measure children's behavioral regulations. The task has been shown to be a reliable and valid measure of children's behavioral regulation skills, involving multiple components of executive functions including attentional focusing, working memory, and inhibitory control (Ponitz et al., 2009). Two forms of the task, form A and B, were used.

Test of Early Mathematics Ability, Third Edition (TEMA-3) – The TEMA-3 (Ginsburg & Baroody, 2003) is a standardized test of children's general mathematical abilities, normed for use with children from 3 to 8 years of age. Two forms of the test, form A and B, were used. Stimuli included picture books, Examiner Record Booklets, Worksheets, 5" x 8" Cards, and 20 Blocks (see Figure 1).

Figure 1: Standard Test of Early Mathematics Ability, Third Edition (TEMA-3)

Give a Number (GN) task- The GN task, adapted from Wynn (1992), was administered to measure children's understanding of the cardinality principle. Stimuli consisted of a set of 15 blocks and a puppet to request "candies" (i.e. blocks).

Give the same (GS) task- The GS task was a researcher created measure. The task was administered to measure children's ability to produce a set of candies equal in number with another given set. Stimuli consisted of a set of 15 blocks and two cards. Two forms of the task were used. Form A consisted of a card with a picture of a boy with 1 to 7 candies printed on the card, and a card with a picture of a girl with no candies. Form B consisted of a card with a picture of a girl with 1 to 7 candies printed on the card, and a card with a picture of a boy with no candies. Figure 2 shows an example of the cards from form A in which the boy has 6 candies.

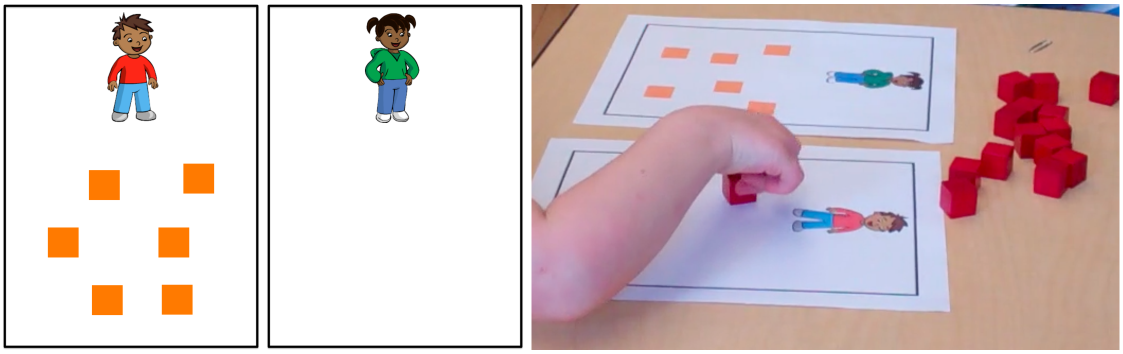

Figure 2: Give the Same Task

Make the same (MS) task-The MS task was a researcher created measure. The task was administered to measure children's ability to produce two equal sets by adding or removing items from the sets. Stimuli consisted of a set of 15 blocks, a card with a picture of a boy, and a picture of a girl (see Figure 3).

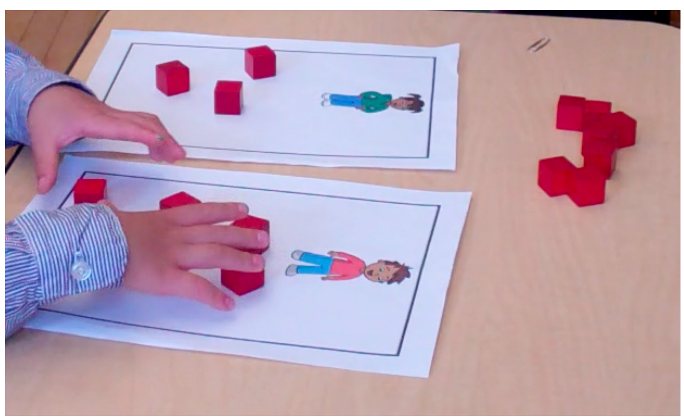

Figure 3: Make the Same Task

The MS task had a total of 6 trials in addition to a practice trial. Table 1 shows number of candies for the boy and for the girl in each the 6 trials of the MS task.

Table 1: Number of Candies for the Boy and the Girl in each the 6 trials for the MS task

Trial#	Boy's # of candies	Girl's # of candies
P	2	3
1	2	4
2	5	3
3	4	5
4	7	4
5	7	6

Equivalence Judgment (EJ) task- Stimuli consisted of 16 cards with a picture of a boy or a picture of a girl, each with some number of candies (see Figure 4).

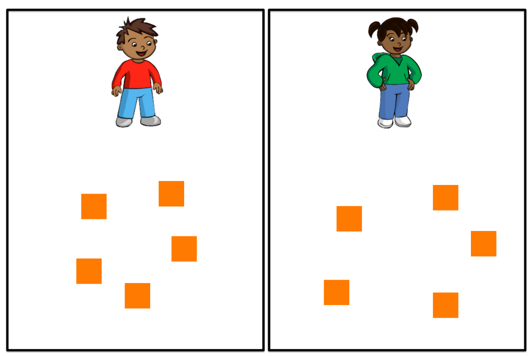

Figure 4: First card in the Judgment Task

The EJ task had total of 6 testing trials in addition to two practice trials. In each of the testing trials the number of candies for the boy and the girl ranged between 4 to 7. The candies were positioned randomly (but fixed) on the cards. The ratio between the boy's and the girl's set of candies was kept close to 1:1, and the relative area occupied by the candies in each set was incongruent with their relative numerical size; for example, if the boy had more candies than the girl, the girl's candies were more spread out on the card or were spread out equally the same as the boy's candies. The number of candies for the boy and the girl, as well as the relative area for the two sets in each of trial is shown in Table 2.

Table 2: Number of candies on each card in each trial for the EJ Task

Card#	Boy's # of candies	Girl's # of candies	Area
P1	2	2	Boy > Girl
P2	2	3	Boy = Girl
Form A			
1	4	4	Boy > Girl
2	4	5	Same
3	6	5	Boy < Girl
4	5	5	Boy < Girl
5	6	6	Boy > Girl
6	7	6	Same
Form B			
1	5	4	Same
2	4	4	Boy > Girl
3	5	5	Boy > Girl
4	6	5	Boy < Girl
5	6	7	Same
6	6	6	Boy < Girl

Training Tasks

Pop the Bubble activity- The Pop the Bubble activity is one of the MathemAntics software's activities for preschoolers. The activity is designed to help children learn about numerical relations between two sets while encouraging them to compare the sets by counting.

Noodle Words app- The Noodle Words app Version 1.3 is a published app on iTunes by NoodleWorks Interactive LLC. The app is designed to help children improve their language acquisition and word comprehension.

Post-test Tasks

Post-test measures were identical to the pre-test measures with the exception of the HSKT task, which was administered only at pre-test.

Design

The experiment used a between subject design with condition (diagram only, diagram +

gesture, no diagram/no gesture, and control) as between subject factor, and had three parts: (1) pre-test, (2) four sessions of training, (3) post-test.

Procedure

The study was conducted in pre-schools where students were attending. Children were pulled out of their classrooms and worked with trained examiners in one-on-one settings, away from their classroom in a hallway or a room assigned to the research group by the school administrators. Data for the study was collected during Fall 2012.

At pre-test, children were tested on six measures: Head Shoulders Knees Toes task (HSKT), Standard Test of Early Mathematics Ability (TEMA), Give a Number task (GN), Give the Same task (GS), Make the Same task (MS), and Equivalence Judgment task (EJ). In the first week of the study, students were tested on the HSKT task. In the second week of the study students were tested on the TEMA and the GN tasks, in a counterbalanced order. During the third week of the study students were tested on the GS, MS, and EJ tasks in a counterbalanced order in two sessions.

Four training sessions were conducted over two weeks, twice a week, for a total of 4 sessions for each child. Children were blocked by their normed pre-test TEMA results into 4 blocks (poor, below average, average, above average) and by their schools (A, B, C), and then randomly assigned to one of the four conditions: three math conditions and one reading condition (i.e. control). All students in the three math conditions— diagram only, diagram + gesture, and no diagram/no gesture— played the Pop the Bubble activity from the MathemAntics software. Students in the control reading condition played the Noodle Words app by NoodleWorks Interactive, LLC.

Post-test measures were conducted after completion of the four training sessions. In the first week of post-testing children were tested on the TEMA and the GN task in a counterbalanced order. During the second week of the post-testing students were tested on the GS, MS, and EJ tasks in a counterbalanced order over two sessions.

The pre-test, post-test, and intervention sessions were video recorded. In addition, video records of the computer screen and front views of participants were captured with Silverback © software. Figure 5 summarizes the overall design and procedure.

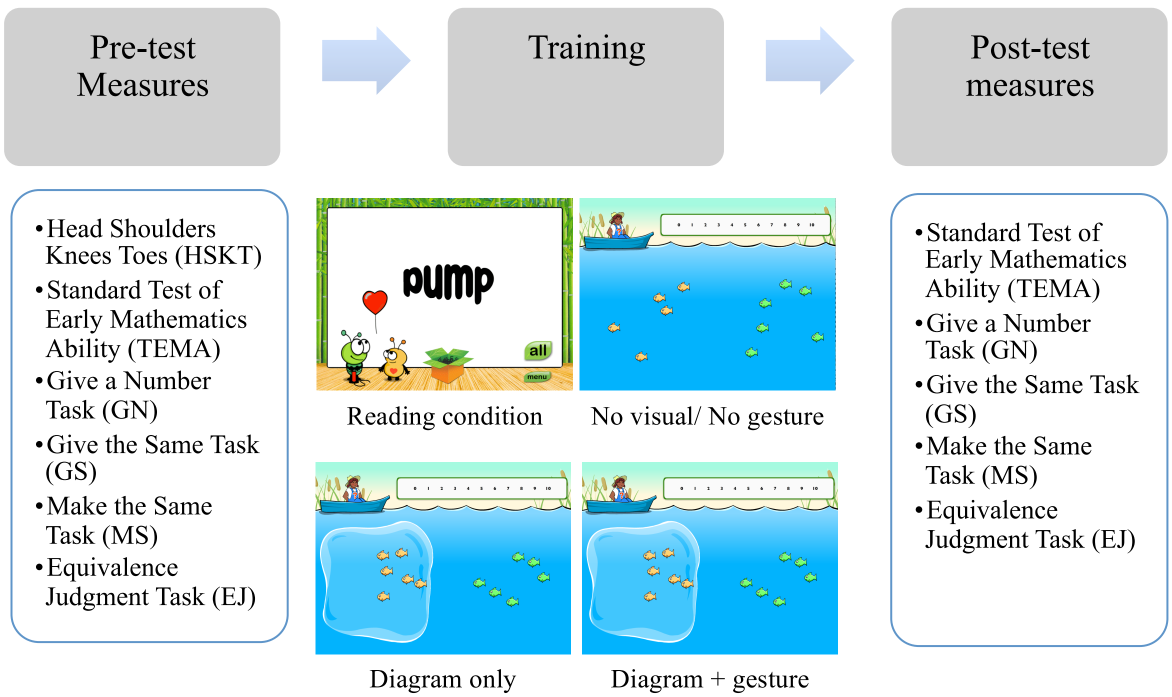

Figure 5: Design of the Study

Pre- and Post-test Tasks

Head-Toes-Knees-Shoulder task (HTKS)- Examiners followed original scripts used by Ponitz et al. (2009). The HTKS task included 20 test trials. The experimenter started by saying, “Now we’re going to play a game. The game has two parts. First, I want you to copy what I do. Touch your head”. She then waited for the child to put both his/her hands on head, and said, “Good! Now touch your toes.” Once the child put his/her hands on toes, experimenter said, “Good!”. The two commands were repeated with motions until the child imitated them correctly. The child was then told, “Now we’re going to be a little silly and do the opposite of what I say. When I say to touch your head, instead of touching your head, you touch your toes. When I say to touch your toes, you touch your head. So you’re doing something different from what I say. What do you do if I say touch your head?” Children who hesitated or responded incorrectly, were told, “Remember, when I say to touch your head, you touch your toes, so you are doing something different from what I say. Let’s try another one.” Children who responded correctly were told, “That’s exactly right”, and proceeded to the next trial. There were a total of two training trials and 4 practice trials. Experimenters could re-explain up to three times in the training and practice sections. If the child could not do the task after the third explanation, the test items were administered anyway. Part I of the testing included 10 trials. Experimenter started the test trials by saying, “We’re going to keep playing this game, and you keep doing the opposite of what I say.” Part II of the testing was administered only if child responded correctly (including self-corrects) to 5 or more items on Part I. An item was marked as “self-correct” if the child made *any discernible* motion toward the *incorrect* answer, but then changed his/her mind and made the correct response.

Experimenter started Part II by saying, “Ok, now that you’ve got that part, we’re going to add a part. Now, you’re going to touch your shoulders and your knees. First, touch your shoulders.” Part II had only one training question and children received only two explanations if they made mistakes— one explanation during training and one explanation during the practice. They, then, proceeded to 10 testing trials.

Correct responses earned 2 points; incorrect responses earned 0 points; self-correct responses earned 1 point. Scores ranged from 0 to 40. Commands were given in a consistent, nonrandom order following the ordered specified in the original scripts. Higher scores indicated higher levels of behavioral regulation. Two forms of the task were used, counterbalancing the commands used on the first 10 trials (heads/toes vs. knees/shoulders). No significant form differences in overall scores have emerged in previous studies (Ponitz et al., 2009). Children were randomly assigned to either of the two forms.

Test of Early Mathematics Ability (TEMA)- Examiners followed the exact script and procedures listed in the manual of the test. Examiners started with an item that was appropriate for the child’s age at the time of testing: three-years old started with item 1, and four-years old started with item 6. Examiners continued testing until child answered 5 consecutive questions incorrectly. Once ceiling point was established, examiners checked if child answered 5 consecutive questions correctly to establish Basal. If a child had not answered 5 consecutive questions correctly, examiners started at item 5 and went backwards until Basal was established or all items were tested. Children did not receive any feedback on their accuracy for any of the items. To encourage children to continue testing, examiners praised their effort by saying phrases like, “You’re working really hard!”, “Good listening", "You seemed to like that," or "I see you are trying really hard", but avoided comments that appeared to reflect on the accuracy of their responses, such as "very good" or "that's right".

At pre-test, children were randomly tested with either Form A or B of TEMA-3. At post-test the assignment switched.

Give a Number (GN) task- Children were asked to give 1, 2, 3, 4, 5, 6, or 7 candies to a puppet. The task was adopted from Wynn’s task but added numbers 4 and 7, which were not part of the original Wynn's task. To begin, experimenter placed 15 blocks on the table in front of child and said, “Let’s pretend these are candies [pointing to the blocks]”. She then introduced the child to a Puppet and said, ‘‘Puppy wants to have some number of candies; he’s going to ask you for the number he wants. You see if you can help him." Experimenter then made the puppet act and asked, "give me 1 candy. Put it right here, just one candy! [pointing to a place in front of puppet, within easy reach of the child]". If child did not provide a response or gave a number other than 1, experimenter prompted with phrases like, "he only wants 1 candy," and "can you pick out 1 candy for Puppy?". If child yet failed to give 1 candy, experimenter put 1 candy in front of the puppet and said, "Here is 1 candy for the puppy", then placed the candy back in the pile, and made the puppet to ask again, "give me 1 candy? Put it right here, just one candy!''. If child successfully placed 1 candy in front of the puppet, experimenter continued with the next trial; otherwise, stopped the task.

For the remaining trials, if child was successful in providing X candies, experimenter asked for X + 1 candies, and if child failed to give X candies, experimenter asked for X – 1 candies. In the event when child was successful in giving 7 candies, experimenter asked for 7 candies again until child successfully gave 7 candies twice out of three times. At the end of each trial, when child provided a set of any number of candies, puppet said, "Yay, I like my candies!", and experimenter put all the blocks together back in the pile.

After responding, children were given "follow-up" questions. They were asked to "check and make sure" that they had given the correct number, and were reminded how many had been asked for. Any child who did not spontaneously count the objects, regardless of the number he/she had given, was prompted to count them, saying, ''Puppy wants X candies. Count and make sure you gave Puppy X candies?'' If children did not count in fixing the set, experimenter prompted them once more, saying "Count and make sure Puppy has X candies". Children who count and obtained a different number than what they had been asked for were then prompted, ''But Puppy wanted X candies—can you fix it so that there are X?'' The experimenter stopped prompting if child clearly had difficulty understanding the question or failed three times in providing a correct answer. Testing was stopped if child provided two incorrect responses for any number. Children were allowed a single counting error in each trial, so they were given credit even when they had actually given $X \pm 1$ but counted the number that they were asked for.

Give the same (GS) task- At pre-test half of the students were randomly tested with Form A, and the other half with Form B. At post-test the assignment switched. In what follows, we will outline the script used for Form A. To begin, experimenter placed the card with the picture of the Boy with 1 candy on the table, introduced the child to the Boy and said, “Here is Charlie and he has some candies [Pointed to the boy’s candies]. Then, child was introduced to the Girl, “Here is Sarah, Sarah wants to have the same number of candies as Charlie has; give her **the same** number of candies as Charlie has” [emphasis on “the same”]. Experimenter, then, placed 15 blocks on the table within reach of the child. Children who did not respond, were prompted by the phrases like “Sarah wants to have **the same** number of candies as Charlie has; can you give her **the same** number of candies”. If a child yet failed to give 1 candy to Sarah, experimenter put 1 candy on Sarah’s card and said, “Here is 1 candy for Sarah. Now they have the same. Charlie has 1 and Sarah has 1.” Then placed the candy back in the pile, and repeated the question, “give Sarah **the same** number of candies as Charlie has”. If child successfully gave 1 candy, the experimenter continued to the next trial; otherwise, the experimenter stopped testing.

For the remaining trials, if child was successful in providing X candies, experimenter asked for X + 1 candies (i.e. used the card with X + 1 candies), and if child failed to give X candies, experimenter asked for X – 1 candies. In the event when child was successful in giving 7 candies, experimenter asked for 7 candies again until child successfully gave 7 candies twice out of three times. At the end of each trial, when child provided a set of any number of candies, experimenter said, “Yay, Sarah likes her candies!”, and put all the blocks together back in the pile.

After responding, children were given “follow-up” questions. Children who counted both sets, but the last number of their count for the two sets did not match, were probed with, ‘‘But, Sarah wants to have the same number of candies— Fix it so that they have **the same**.’’ Children who counted only one of the two sets, or those who did not show any visible strategies (i.e. counting, matching, visual pattern) were probed with, “Check and make sure Sarah has the same number of candies as Charlie has.” Testing was stopped if child provided two incorrect responses for any number. Children were allowed a single counting error in each trial, and were given credit even when they had actually given X ± 1 but they have counted *both sets* and their count result indicated that the two sets are equal.

Make the same (MS) task- The experimenter started with one practice trial. In this practice trial, she first put 3 candies on girl’s card, and 2 candies on the boy’s card attending to the random spatial arrangement modeled on their recording sheets. To begin, experimenter introduced child to the boy and the girl, saying, “Here are Charlie and here is Sarah. They both have some candies [Pointed to the boy’s and the girl’s candies]. Charlie and Sarah want to have **the same** number of candies. Let’s see if you can help them. Make it so that they have **the same** number of candies [emphasis on the word “same”].” She then placed 11 blocks on the table, and gave scaffold, “You can give them candies [pointing to pile of extra candies on the table] or take away their candies [making a gesture of taking away], make them to have **the same** number of candies”. In response, if child was successful in making the two sets equal, experimenter said, “Excellent, they now both have X candies [X denotes number of candies in the two sets]. On the other hand, if child failed to make the two sets equal, experimenter said, “Let me show you how I do that”.

She then either removed one of Sarah's candies, or gave a candy to Charlie and said, "Now they both have X candies."

At the beginning of the testing trials, experimenter cleared the two cards and placed blocks (i.e. "candies") on each card in the order specified on their recording sheet, and then put 11 extra blocks on the table. Once, child indicated that she/he is done making changes, if it was not clear which candies belonged to which card, experimenter asked child to put Sarah's candies on Sarah's cards, and Charlie's candies on Charlie's card. Child who did not use any visible strategies (i.e. counting, matching, visual pattern, splitting in half), were probed with, ''Check and make sure Sarah and Charlie have the same number of candies". If child counted both sets, and the last number of their count for the two sets did not match, experimenter probed with, "But Sarah and Charlie want to have the same number of candies—fix it so that they have the same". If child counted only one of the two sets, experimenter probed with, "Check and make sure Sarah and Charlie have the same number of candies?". For any trial, if child was clearly playing around and not attending to the task, experimenter proceeded to the next trial. At the end of each trial, experimenter said, "Yay, Sarah and Charlie like their candies!" irrespective of the response accuracy.

While testing, experimenters recorded children's observed and expressed strategies as well as the type of mistakes they made on each trial on a recording sheet.

Equivalence Judgment (EJ) task- Experimenter started with two practice trials. In the first practice trial, she placed the card with the picture of the boy with 2 candies on the table in front of child and said, "Here is Charlie and he has some candies [pointed to the boy's candies one by one]". She then placed the card with the picture of the girl with 2 candies on the table and said, "Here is Sarah and she has some candies [pointed to the girl's candies one by one]. Check and tell me if they have **the same** number of candies [emphasis on the word “same”]. Children who counted both sets and said they have the same were praised, “Excellent! Charlie has 2 and Sarah has 2. They both have the same number of candies, let’s try another one.” Children who did not count and simply said they have the same, were encouraged to show their work, “Yes you’re right. But remember we have to check. Show me how you know they have the same?” If a child failed to do anything or did not provide a relevant answer, experimenter said, “ok! Let me check. Charlie has 1, 2; 2 candies all together, and Sarah has 1, 2; 2 candies all together [pointed to their candies one by one while counting them]. They both have the same number of candies, let’s try another one.” Experimenter, then, proceeded to the second practice questions. On the other hand, if child answered incorrectly (i.e. said they do not have the same), experimenter said, “Let me check. Actually, Charlie has 1, 2; 2 candies all together, and Sarah has 1, 2; 2 candies all together [pointed to their candies one by one while counting them]. They both have the same number of candies, let’s try another one.”

For the second practice trial, she showed the card with the picture of the girl with 3 candies and the card with the picture of the boy card with 2 candies and asked, “Check and tell me if they have **the same** number of candies [emphasized on the word “same”]. Children who counted both sets and said they don’t have the same, were praised, “Yes you’re right! Charlie has 2 and Sarah has 3. They don’t have the same number of candies. Now tell me, which one has more?” In response, if child said “Sarah” has more, experimenter said, “Great Job!”, but if child said “Charlie”, experimenter corrected him by saying, “hum… Actually, Sarah has 3 and Charlie has only 2 candies. Sarah has more”. Children who did not count the sets and just said they don’t have the same, were encouraged to show their work, “Yes you’re right! But remember we have to check! Show me how you do know they don’t have the same?”. If child yet failed to show his work or did not provide a relevant answer, experimenter said, “ok! Let me check. Sarah has 1, 2, 3; 3 candies all together, and Charlie has 1, 2; 2 candies all together [pointed to their candies one by one while counting them]. Sarah has 3 but Charlie has 2. They don’t have the same number of candies, Now tell me, which one has more?”. On the other hand, children who answered incorrectly (i.e. said they have the same), received scaffolding, “Actually, Sarah has 1, 2, 3; 3 candies all together, and Charlie has 1, 2; 2 candies all together [pointed to their candies one by one while counting them]. Sarah has 3 but Charlie has only 2 candies. They don’t both have the same, Sarah has more”.

For the test trials, experimenter showed the cards in the order specified in Table 2. At the beginning of each trial, experimenter asked, “Check and tell me if they have **the same** number of candies”. Children who answered with no visible strategy were prompted to check their answer and were told, “Remember, you don’t have to answer quickly. Check and then answer me. Check and tell me if they have **the same** number of candies”. For children who answered based on the color of candies or used relative words such as too much, a lot, a little, experimenter repeated the question and emphasized on “**the same number**”. Once children provided their answer, the experimenter asked, “Check, and show me how you know that?”, unless they have already counted the two sets separately *and* answered the question correctly. For the trials in which children said the boy and the girl do not have the same number of candies, experimenter asked, “which one has more?”

Previous studies suggest that preschoolers, when asked yes/no questions, have a “yes” bias in answering those questions (Steffensen, 1978) and that they tend to answer quickly without showing a strategy (Jamalian, 2013). To prevent these issues, in the current study, the experimenters avoided asking yes/no questions and encouraged children to “check and tell”.

While testing, experimenters recorded children's observed and expressed strategies as well as the type of mistakes they made on each trial on a recording sheet.

Training conditions

Pop the Bubble activity, _diagram + gesture_ condition- Experimenter started the session by saying, "Today, we are going to play the pond and the fish game." Children played the activity on an iPad. Activity started with a farmer greeting children, "Wow! What a nice day! Let's make bubbles with the fish in the pond. Oh! This will be fun". Farmer then asked the child to draw a bubble around the first set of fish, by saying, for example, "Make a bubble around the pink jellyfish". Children then drew a bubble around the set of fish by making a circular gesture around the set. If help was needed, experimenters explained to children to "Make a big circle around all the green fish" (for example) and if necessary they held the child's finger and drew bubbles together.

Once the bubble was made, farmer asked, "How many pink jellyfish are in the bubble?". Numbers from 0 to 10 appeared on the screen and child entered her answer by touching one of the numbers on the screen (see Figure 6). Experimenters provided scaffolding for finding numbers to children who had difficulty reading numbers; in these cases, they simply helped the child to find the number that he/she wanted to respond with, without making any correction in their counting, and reported in their recording sheet that the child needed help finding numbers. When a child counted above 10, the experimenter gave the scaffold: "… is too high. Try again. How many fish are in the bubble?"

Once children touched a number, the farmer repeated that number aloud and gave feedback on their accuracy of response. Children who accurately answered the how many question received feedback, "excellent counting", and the cardinality number appeared on top of the bubble and remained there for the remaining of the trial. They were then asked to make a bubble around the second set of fish and say how many fish are in that set. Children who made a mistake in answering the how many question were given a second chance and were told by the farmer, “Count again. How many yellow jellyfish are in the bubble?”. If child made a mistake in his second attempt, farmer corrected him by saying, for example, “That’s not right! Let me show you how I know that. There are 1, 2, 3, 4; 4 yellow jellyfish in this bubble.” Each fish got highlighted one by one when counted and remained highlighted to distinguish the counted items from uncounted ones. Once all items in the set were counted, their highlights were removed; farmer repeated the last counting word (ex. “4 pink jellyfish in this bubble”) and all the fish in the bubble got highlighted together.

When bubbles were made around the two sets and their cardinal values appeared on top of them, farmer asked the child, “do the two bubbles have the same number of fish in them?” A green checkmark and a red cross appeared on the screen. If needed, experimenter explained to the child that ✓ means “yes”, and ***x*** means “no”. Children received feedback on the accuracy of their response. Those who responded correctly were told, “Awesome! The two bubbles don’t have the same (or have the same) number of fish in them” and proceeded to the next question. Children who responded incorrectly were give instructive feedback, for example, “Oh! Oh! The two bubbles have the same (or don’t have the same) number of fish in them. Let me show you how I know that. There are 1, 2, 3; 3 yellow fish in the bubble. And there are 1, 2, 3; 3 yellow clams in the other bubble. The two bubbles have the same number of fish in them.” Children played a total of 5 trials in each session. Number of fish in each set in each trail for each session is reported in Table 3. Order of questions was set randomly by computer for each child in each session. The two sets of fish were spatially separated, and the items is each set was arranged unsystematic and randomly by computer. The particular type of fish in each set was also chosen randomly by the computer from possible choices of Yellow Fish, Green Fish, Pink Jellyfish, and Yellow Clam. All the items in each set were similar in type and color, but items in the two sets were always different from each other.

Table 3: Session Trials

Session #	First Set Cardinal Value	Second Set Cardinal Value
1	3	3
1	2	3
1	5	5
1	4	5
1	4	3
2	5	5
2	5	6
2	6	6
2	6	7
2	7	7
3, 4	5	6
3, 4	6	6
3, 4	6	7
3, 4	7	7
3, 4	7	8

In sessions 3 and 4, if the two sets were not the same in their cardinal value, children were asked to indicate which set has more fish in it. After child had responded whether the two bubbles have the same number of fish in them and had received feedback on the accuracy of her response, Farmer asked the child, “Pop the Bubble that has more fish in it”. If child chose the correct bubble, the bubble popped and farmer said, “You are right! You popped the bubble that had more fish in it”. If child chose the wrong bubble, the bubble changed color to orange, and farmer counted the two sets and explained the other bubble has more fish in it.

During the session and while children were playing, experimenters recorded children’s observed strategies as well as the type of mistakes they made on each trial on a recording sheet.

On average each trial lasted 118.83 seconds (SD= 56.35). The exact duration of trial depended on how fast children drew the bubbles, how fast they responded to the "how many" and equivalence judgment questions, and whether they made a mistake in their responses in which case counting feedback was provided to them.

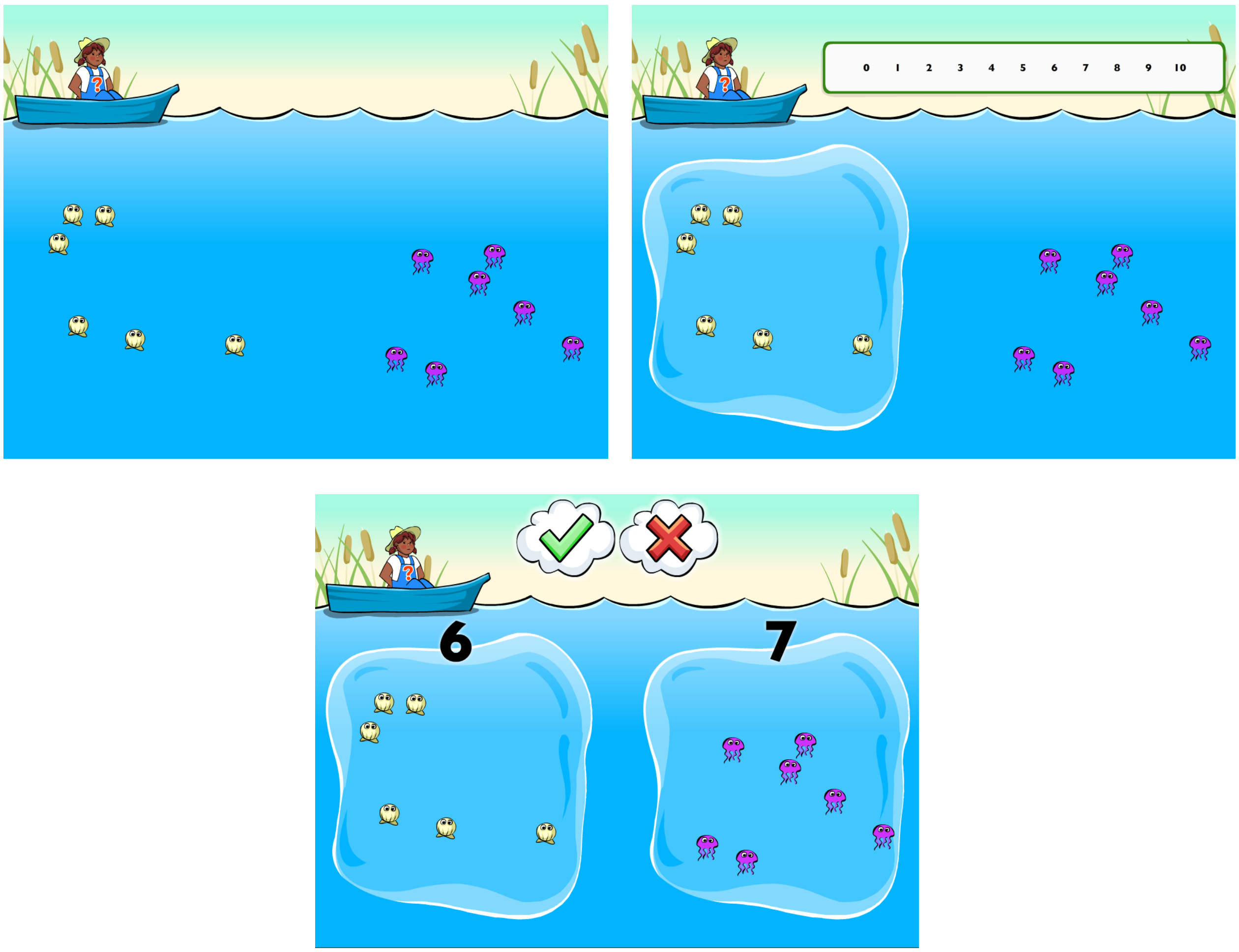

Figure 6: Pop the Bubble Activity with Bubbles

Pop the Bubble activity, <u>diagram only</u> condition- The only difference between this condition and the diagram + gesture condition was that bubbles appeared automatically around the two sets and children were not asked to make bubbles.

On average each trial lasted 84.70 seconds (SD= 58.57). The exact duration of trial depended on how fast children responded to the "how many" and equivalence judgment questions, and whether they made a mistake in their responses in which case counting feedback was provided to them.

Pop the Bubble activity, no diagram/ no gesture condition- Children in this condition never saw bubbles nor were asked to make bubbles nor heard the word "bubble". Instead of asking, "How many yellow clams are in the bubble?", the farmer asked, "How many clams are there?". Instead of asking, "do the two bubbles have the same number of fish in them?", the farmer asked, "Are there the same number of yellow clams as there are pink jellyfish?". In sessions 3 and 4, instead of asking "Pop the Bubble that has more fish in it," the farmer asked, "show me which group has more." Figure 7 shows two screenshots of the game in this condition.

On average each trial lasted 86.90 seconds (SD= 43.08). The exact duration of trial depended on how fast children responded to the "how many" and equivalence judgment questions, and whether they made a mistake in their responses in which case counting feedback was provided to them.

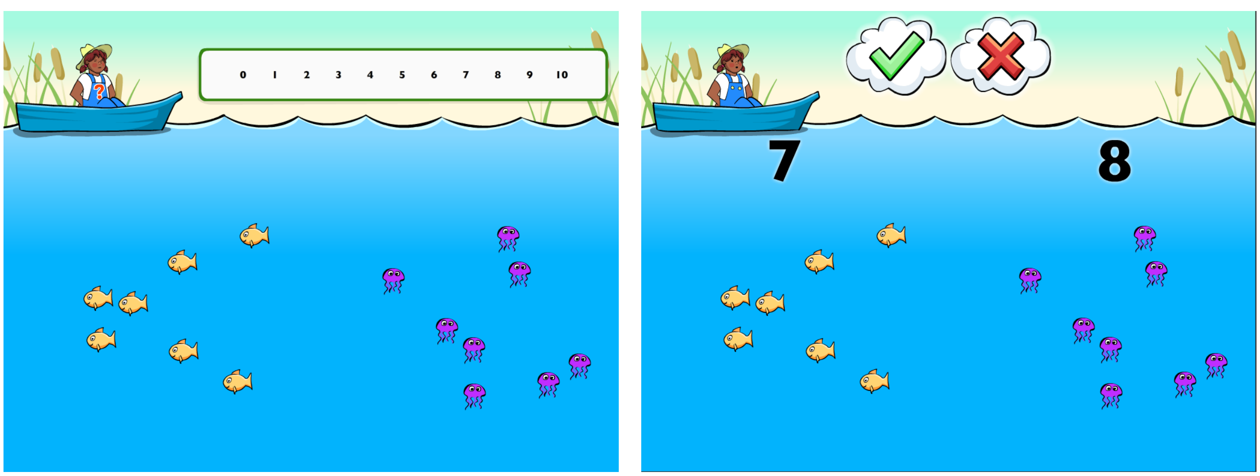

Figure 7: Pop the Bubble Activity with no Bubbles

Reading condition- Experimenter started the session by saying, "Today, we are going to play a reading game." Children played the NoodleWords app on an iPad. The game started with a bee saying, “Hi! Welcome to Noodle Words! These are my friends. This is “Stretch” and that’s “Squish” [referring to two bugs on the screen]. Here is how you play. We’ve got a box of words for you. Tap on the box to get a word, you can tap the word too, and you can press and hold to move it around. Tap the bugs, tap the words, and anywhere else you want.” When child tapped on the box, a new action verb from a set of 18 possible choices (listed in Table 4) would appear in the middle of the screen, and the two bugs acted the action through an animation (see Figure 8). Children could also tap on the word to see the word in animation, for example, by tapping on the word “spin” the word would spin and a narrator would read out the word “spin”.

Experimenters recorded the verbs children saw when interacting with the app on a recording sheet.

Table 4: Set of Verbs Available in the Noodle Words App

Set of Verbs in the Noodle Words App
Spin Sparkle Stretch Laugh Pop Drop Dance Grow Blow Bubble Bump Jump Pump Surprise Shake Eat Run Wave

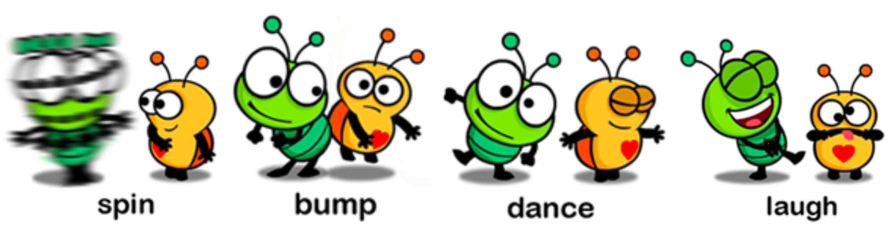

Figure 8: Examples of animations in the Noodle Words App

CHAPTER 3: DATA ANALYSIS

The results are reported in three main sections. In the first section, pre-test performance will be described and changes from pre- to post-test will be analyzed. The goal is to provide an overview of the amount of learning that occurred as a result of the intervention— learning both in terms of accuracy and in terms of strategy use. Another goal is to examine the paths that led to some children's learning and improvement of strategy use and others not, that is, to examine individual differences in the overall amount of learning or achievement. In the second section, children's changing performance and their strategy use during the four sessions of the math intervention will be analyzed. Here, the first goal is to provide a precise depiction of children's progress in learning and strategy use throughout the sessions. The second goal is to analyze the individual differences within this progression— some children may learn more quickly than others or show different learning progressions to the same end-state. In the third section, children's performance in the post-test standardized test will be compared to their performance during the last session of the intervention. The goal is to examine transfer of skills from the intervention to the post-test standardized test.

Changes from Pre-test to Post-test

Head-Toes-Knees-Shoulder task (HTKS)- To determine any possible group differences in executive functioning skills, a 3 x 2 x 3 three-way analysis of covariance (ANCOVA) with condition (diagram only, diagram + gesture, no diagram/no gesture, reading group), gender, and school as between subject variables, age (number of months as a continuous variable) as a covariate, and HTKS scores as dependent variable, was conducted. Results confirmed that there were no significant group differences in children's executive functioning skills, F (3, 92)= 0.731, p= 0.537. There were also no significant differences between girls and boys in their executive functioning skills, F (1, 93)= 0.244, p= 0.623.

Test of Early Mathematics Ability (TEMA)- To examine group differences in TEMA scores at pre-test, a 3x2x3x2 univariate analysis of covariance (ANCOVA) with condition, gender, and school as between subject variables, HTKS scores and age (number of months as a continuous variable) as a covariate variables, and TEMA pre-test scores as dependent measures was conducted. Results confirmed that there were no significant group differences in TEMA scores at pre-test, F(3, 93)= 1.978, p= 0.125). However, there was a significant difference between the three schools in their pre-test TEMA scores, F(2, 92)= 5.401, p= 0.007. Post hoc Tukey's Honestly Significant Difference (HSD) pairwise comparison showed that on average children from school C (M= 13.929, SE= 1.657) outperformed children from school A (M= 8.023, SE= 0.755), p= 0.005, and children from school B (M= 8.346, SE= 0.836), p= 0.029. There was no significant difference between school A and B on TEMA performance.

To examine effects of condition (diagram only, diagram + gesture, no diagram/no gesture, reading group) on TEMA score gains, a 3x2x3x2 repeated measure of covariance with condition, gender, and school as between subject variables, HTKS scores and age (number of months as a continuous variable) as covariate variables, time of testing (pre- test vs. post-test) as within subject variable, and TEMA raw scores as dependent measures, was conducted. All the non-significant interactions were removed from the model.

Results confirmed that there was a significant interaction effect of condition and time of testing on TEMA raw scores, F(3, 92)= 2.949; p= 0.037. Post-hoc sequential Bonferroni pairwise analysis with time as the contrast field, showed that on average children in the three math conditions improved in their TEMA scores from pre- to post-test, while children in the reading condition did not improve significantly from pre- to post-test. Children in the gesture + diagram condition had an average score of 14.241 (SE= 1.27) at post-test which was significantly more than their average at pre-test (M= 10.341, SE= 1.039), t(175)= 2.676, p= 0.008. Similarly, on average children in the diagram only condition scored higher at post-test (M= 15.861, SE= 1.202) than at pre-test, (M= 10.409, SE= 0.984), t(175)= 3.770, p<0.001; and children in the no-diagram/no-gesture condition had a higher average at post-test (M= 12.151; SE= 1.202), than at pre-test (M= 8.559, SE= 1.04), t(175)= 2.469, p= 0.014. On the other hand children in the reading condition did not significantly improve in their TEMA scores at post-test (M=12.270, SE= 1.258), comparing to their pre-test (M= 10.344, SE= 1.03), t(175)= 1.323, p=0.188. Further, post-hoc sequential Bonferroni pairwise analysis with condition as the contrast field, showed that children in the diagram-only condition outperformed children in the no-gesture/no-diagram condition at post-test, t(175)= 2.506, p= 0.013, but not at pre-test, t(175)= 1.277, p= 0.203, and children in the reading condition at post-test, t(175)= 2.435, p= 0.016, but not at pre-test, t(175)= 0.052, p= 0.958. No other pairwise comparisons were significant. Table 5 summarizes pre- and post-test performances in TEMA for each of the three conditions.

Table 5: Pre- and Post-test TEMA results

Condition	Time	Mean	Standard Error
Gesture + Diagram	Pre-test	10.341	1.039
	Post-test	14.241	1.27
Diagram only	Pre-test	10.409	0.984
	Post-test	15.861	1.202
No-gesture/no-diagram	Pre-test	8.559	1.04
	Post-test	12.151	1.271
Reading	Pre-test	10.344	1.03
	Post-test	12.270	1.258

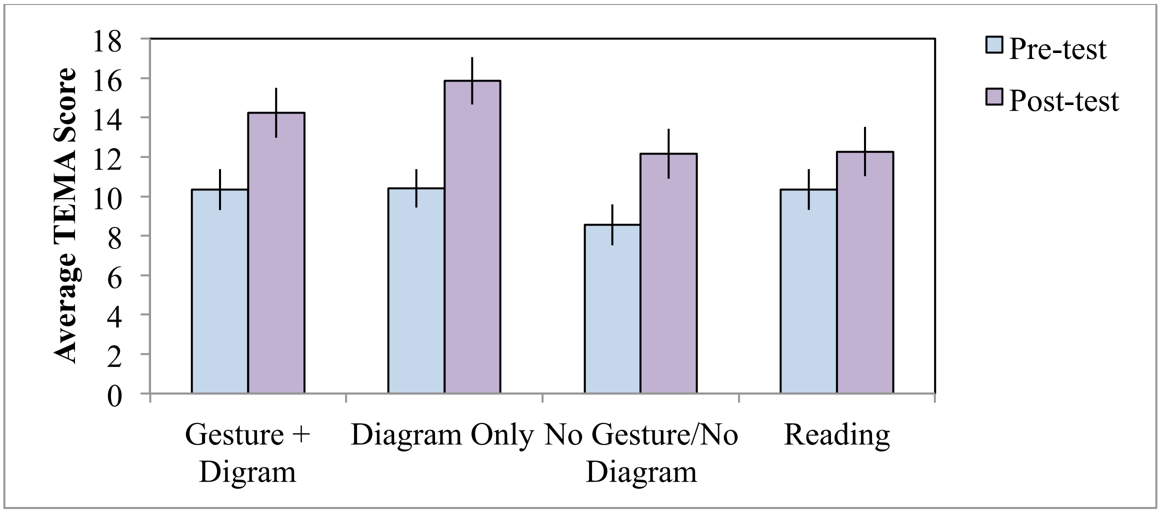

Figure 9: Average TEMA Scores Across the Four Condition at Pre- and Post-test

Age (number of months as a continuous variable), F(1, 93)= 20.514, p<0.001, and HTKS scores, F(1, 93)= 3.747, p= 0.056, explained a significant proportion of variance in TEMA scores. Age was significantly correlated with TEMA scores at pre-test, r(92)= 0.422, p<0.000, and at post-test, r(92)= 0.445, p< 0.001. Similarly, HTKS scores were positively correlated with pre-test TEMA scores, r(92)= 0.349, p= 0.001, and marginally correlated with post-test TEMA scores, r(92)= 0.198, p= 0.056.

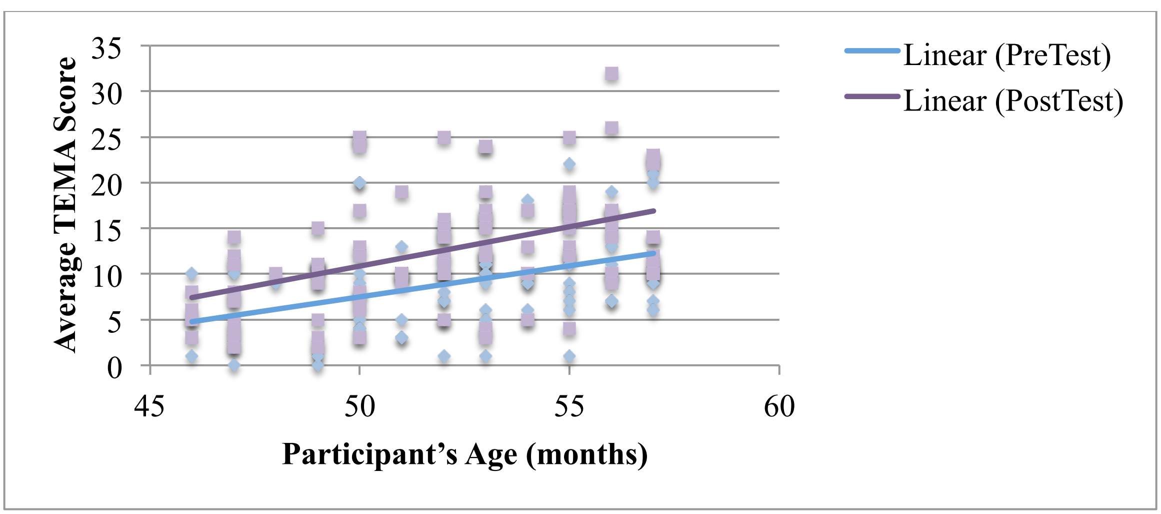

Figure 10: Average TEMA scores versus Participant's age

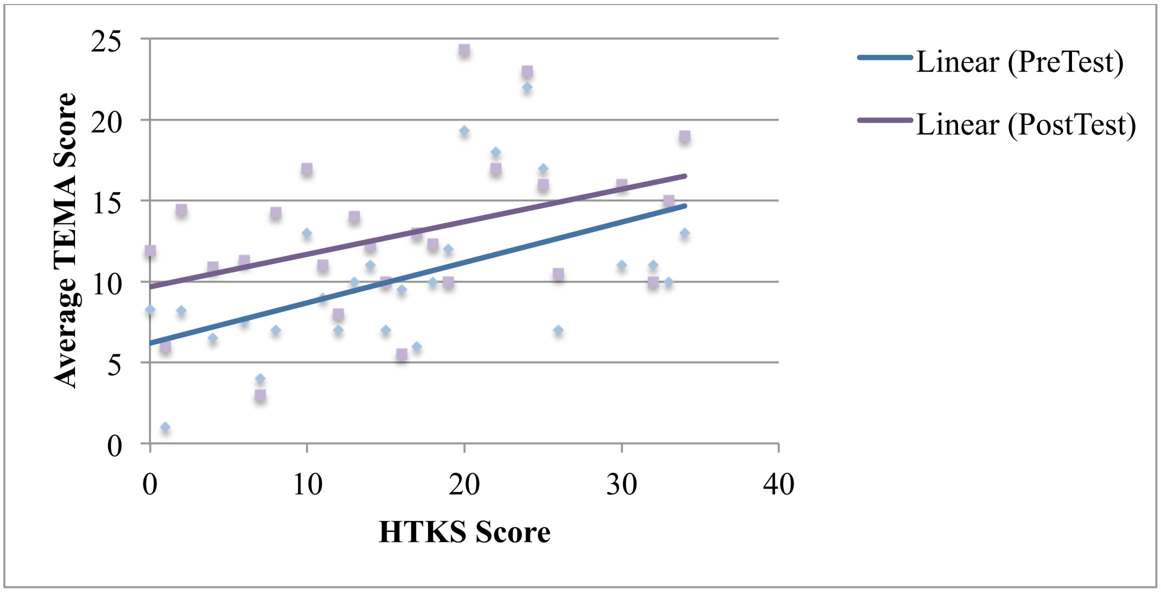

Figure 11: Average TEMA Scores versus HTKS (Measure of Executive Functioning Skills)

Further, there was a significant main effect of school on TEMA scores, $F(2, 93)= 3.568$, $p=0.033$. Post hoc Tukey's Honestly Significant Difference (HSD) pairwise comparison showed that on average children from school C ($M= 15.857$, $SE= 1.771$), outperformed children from school A ($M= 9.880$, $SE= 0.807$), $p= 0.010$, and children from school B ($M= 10.193$, $SE= 0.894$), $p= 0.034$. There was no significant difference between school A and B on TEMA performance.

Table 6 summarizes results from the repeated measure ANCOVA analysis. In summary, children in the three mathematics comparison groups significantly improved in TEMA scores from pre-test to post-test, while children in the reading condition did not improve significantly. Children in the diagram-only condition outperformed children in the no-gesture/no-diagram condition and children in the reading condition. There was no significant difference between children in the gesture + diagram condition and children in the diagram-only condition. Further, age and executive functioning skills explained a significant proportion of variance in TEMA scores.

Table 6: Summary of ANCOVA analysis on TEMA scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Between Subject Effect					
Intercept	425.191	1	425.191	9.885	0.002
HKTS	161.177	1	161.177	3.747	0.056
Age	882.359	1	882.359	20.514	0.000
Condition	199.686	3	66.562	1.548	0.208
School	306.972	2	153.486	3.568	0.033
Gender	4.053	1	4.053	0.094	0.760
Within Subject Effects					
Time	13.939	1	13.939	1.694	0.197
Time * HSKT	19.412	1	19.412	2.359	0.128
Time * Age	30.23	1	30.23	3.673	0.059
Time * Condition	72.81	3	24.27	2.949	0.037
Time * School	1.21	2	0.605	0.073	0.929
Time * Gender	2.829	1	2.829	0.344	0.559

Next, we focused our analysis on the specific TEMA items that measured skills directly related to the intervention. Table 7 lists the items along with a description for each. These items measured children's skills in recognizing written numerals, reciting the counting numbers, enumerating sets of objects, and comparing quantities.

For each item, a score of 1 was assigned for correct response, and score of 0 was assigned for incorrect responses as determined by the criteria given in the TEMA handbook (also outlined in Table 7). To examine whether condition had an effect on children's likelihood of correct responses to these items, a linear binary logistic regression mixed model analysis with score of each item as the dependent variable, condition, gender, and school as between subject variables, item number as within subject variables, and HTKS score and age as covariant variables was conducted.

Table 7: TEMA Items Directly Related to the Math Intervention

Item number	Item Name	Stimulus
4	Perception of More: Up to 10 Items	Which side has more? Form A- p: 10 or 2; a: 7 or 3, b: 2 or 8; c: 1 or 6; d: 9 or 4 Form B- p: 9 or 1; a: 8 or 2, b: 3 or 7; c: 2 or 6; d: 9 or 3
5	Nonverbal Production: 1 to 4 items	Make yours just like mine. Form A- a: 2; b: 4; c: 3 Form B- a: 3; b: 2; c: 4
6	Enumeration: 1 to 5 items	You count the stars. Form A- p: 2; a: 4; b: 5 Form B- p: 2; a: 3; b: 5
7	Cardinality Rule	How many stars did you count? Form A- p: 2; a: 4; b: 5 Form B- p: 2; a: 3; b: 5
9	Number Constancy	How many tokens are there? Form A- a: 3; b: 5; c: 4 Form B- a: 3; b: 5; c: 4
10	Produce sets up to 5 items	Give me … tokens. Form A- a: 3; b: 5 Form B- a: 2; b: 5
12	Verbal counting by ones: 1 to 10	1, 2, 3, now count by yourself.
14	Reading numerals: Single digit numbers	What number is this? Form A- a: 2; b: 5; c: 6 Form B- a: 3; b: 7; c: 9

Results of the analysis on specific TEMA items were fairly similar to those on the overall TEMA scores. There was a significant interaction effect of condition and time of testing on the likelihood of correct responses to the specific TEMA items, $F(3, 1452)= 4.647$, $p= 0.003$. Post-hoc sequential Bonferroni pairwise analysis contrasting with respect to time, showed that on average, children in the gesture + diagram condition were more likely to respond correctly to these items at post-test ($M= 0.848$, $SE= 0.848$) than at pre-test ($M= 0.669$, $SE= 0.32$), $t(1452)=$

4.299, $p< 0.001$; similarly, on average children in the diagram only condition were more likely to respond correctly to these items at post-test (M= 0.859, SE= 0.175) than at pre-test (M= 0.612, SE= 0.343), $t(1452)= 5.501$, $p<0.001$; and children in the no-diagram/no-gesture condition had a higher likelihood of correct responses at post-test (M= 0.707; SE= 0.299), than at pre-test (M= 0.581, SE= 0.352), $t(1452)= 2.306$, $p= 0.021$. On the other hand, children in the reading condition were not significantly more likely to respond correctly at post-test (M= 0.702, SE= 0.302) as they were in pre-test (M= 0.668, SE= 0.32), $t(1452)= 0.659$, $p= 0.510$. Further, post-hoc sequential Bonferroni pairwise analysis contrasting with condition showed that children in the diagram + gesture condition significantly outperformed children in the no-gesture/no-diagram condition at post-test, $t(1452)= 3.425$, $p= 0.001$, but not at pre-test, $t(1452)= 1.548$, $p= 0.122$, and children in the reading condition at post-test, $t(1452)= 3.596$, $p< 0.001$, but not at pre-test, $t(1452)= 0.022$, $p= 0.983$. Similarly, children in the diagram-only condition significantly outperformed children in the no-gesture/no-diagram condition at post-test, $t(1452)= 3.797$, $p< 0.001$, but not at pre-test, $t(1452)= 0.526$, $p= 0.599$, and children in the reading condition at post-test, $t(1452)= 3.897$, $p< 0.001$, but not at pre-test, $t(1452)= -1.001$, $p= 0.317$. No other pairwise comparisons were significant. Table 8 summarizes average likelihood of correct responses to the specific TEMA items in each condition at pre- and post-test.

Table 8: Likelihood of Correct Response to Specific TEMA Scores Across the Four Conditions

Condition	Time	Mean	Standard Error
Gesture + Diagram	Pre-test	0.669	0.32
	Post-test	0.848	0.186
Diagram only	Pre-test	0.612	0.343
	Post-test	0.859	0.175
No-gesture/no-diagram	Pre-test	0.581	0.352
	Post-test	0.707	0.299
Reading	Pre-test	0.668	0.32
	Post-test	0.702	0.302

To summarize, post-hoc analysis showed that on average children in the three math conditions were significantly more likely to respond correctly to the specific TEMA items at post-test than at pre-test, but children in the reading condition were not significantly more likely to respond correctly at post-test as they were in pre-test. Further, at post-test children in the diagram + gesture and in diagram-only conditions were more likely to respond correctly to the specific TEMA items than children in the no-gesture/no-diagram and reading conditions. There was no significant difference between children in the gesture + diagram condition and children in the diagram only conditions.

Give a Number (GN) task- Children were categorized based on the highest number of objects they could give. I adopted Wynn's criteria to determine the highest number word "N" each child knew the exact meaning of:

1) Give N objects at least two out of three times when asked for than number,
2) Give N objects no more than half as often when asked for a different number,
3) Satisfy conditions 1 and 2 for all numbers less than N.

Following Wynn's criteria, children who made one counting mistake were considered to be accurate. For example, a child who gave 5 candies when asked for 6 candies but counted the set as 6 was considered to be correct. Based on the above criteria children were categorized as 1-knower, 2-knowers, 3-knowers, 4-knowers, 5-knowers, 6-knowers, and 7-knowers. Following Wynn (1990, 1992), children who knew exact meaning of number 5 and above were categorized at cardinality-knowers. To reduce number of categories for simpler interpretation, one-, two-, three-, and four-knowers were combined as sub-knowers.

At pre-test, twelve children in the gesture + diagram condition (total n= 24), sixteen children in the diagram-only condition (total n= 24), eleven children in the no-gesture/no-diagram condition (total n= 23), and ten children in the reading condition (total n= 23) already knew cardinality. Number of cardinality-knowers was not significantly different in each of the four conditions at pre-test, X^2 (3, 94)= 2.903, p= 0.553.

At post-test 73% of children in the gesture + diagram condition (8 out of 11), 75% of children in the diagram condition (6 out of 8), and 58% of children in the no-gesture/no diagram condition who did not know cardinality at pre-test, became cardinality-knowers at post-test, while only 20% of children in the reading condition who did not know cardinality at pre-test, became cardinality-knowers at post-test. To examine whether the math intervention, and in particular performing grouping gesture and observing enclosed diagram around sets, promoted understanding of the cardinality principle, a 3x2 likelihood ratio chi square was conducted with condition (diagram only, diagram + gesture, no diagram/no gesture, reading group) and cardinality category (cardinality-knower and sub-knowers) as main effect, and number of children in each category at post-test as the dependent variable. Children who already knew cardinality at pre-test were excluded from the analysis. Analysis showed significant interaction effect between condition and cardinality groups, X^2 (3, 43)= 8.617, p= 0.035. Children who were sub-knowers at pre-test were more likely to become cardinality-knowers at post-test if they were in the gesture + diagram condition, X^2 (1, 43)= 61.66, p= 0.013, or if they were in the diagram-only condition, X^2 (1, 43)= 5.725, p= 0.017, than if they were in the reading condition. Table 9 shows number of sub-knowers and cardinality knowers in each of the conditions at pre- and post-test.

Table 9: Sub-knowers and Cardinality-knowers in Each of the Four Conditions

Condition	Time	#of Sub-knowers	#of cardinality knowers
Gesture + Diagram	Pre-test	11	12
	Post-test	2	21
Diagram only	Pre-test	8	16
	Post-test	2	22
No-gesture/no-diagram	Pre-test	12	11
	Post-test	5	18
Reading	Pre-test	10	10
	Post-test	8	12

Therefore, we can conclude that observing the enclosed diagram around the sets with or without the grouping gesture helped children to improve their understanding of cardinality.

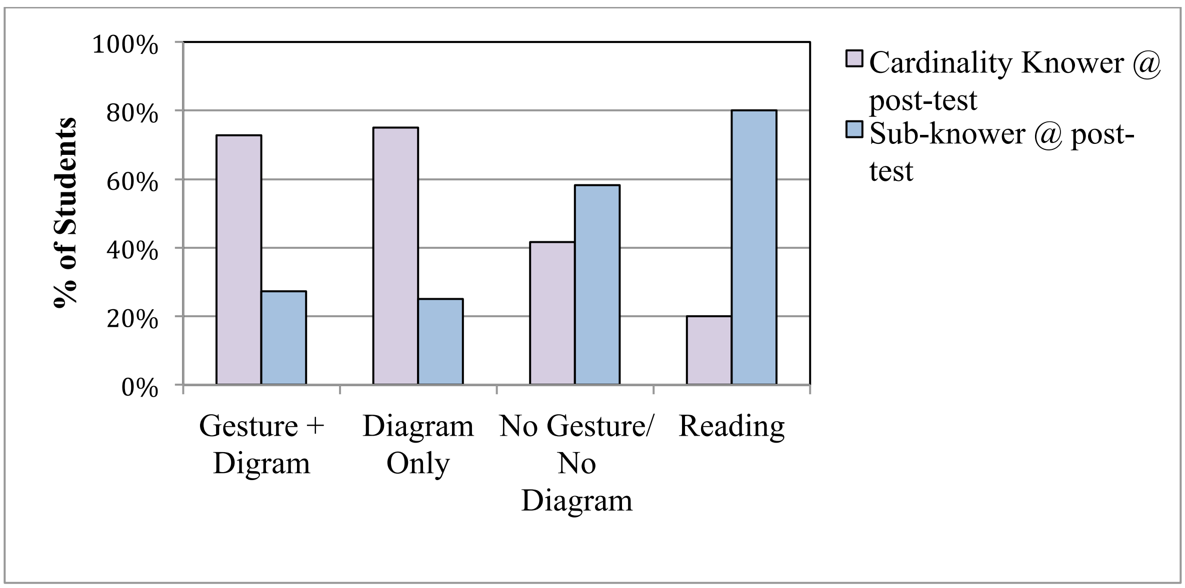

Figure 12: % of Sub-knowers and Cardinality Knowers in Each of the Four Conditions at Post-test (excluding children who already knew cardinality at pre-test)

Give the same (GS) task- Similar to the GN task, in GS task students were asked to produce sets; however, the two tasks differed in two major ways. (1) In GN task students were told to give "N candies", so they needed to know the meaning of each word number to succeed in the task. On the other hand, in GS task they were asked to produce a set with the **same number** of candies comparing to another set. Thus, in order to succeed in the task, children did not need to know the meaning of the number words but they needed to know the meaning of the word “same”. (2) To solve problems in the GS task, children could engage in visual inspection or matching strategy in addition to the counting strategy, but the only applicable strategy to solve GN problems was counting.

Children were categorized based on the highest set number they could produce. I followed similar criteria to the Wynn’s criteria to determine the highest set number “N” each child could produce:

1) Give N objects at least two out of three times when asked to match that number,
2) Give N objects no more than half as often when asked for a different number,
3) Satisfy conditions 1 and 2 for all numbers less than N.

Following Wynn’s criteria, children were allowed to make one counting mistake to be considered accurate. For example, a child who gave 5 candies to Sarah when shown the card with Charlie’s picture with 6 candies, but counted **both** sets as 6 was considered to be correct. Children were also given credit if they solved the problems with strategies other than counting—strategies such as one-to-one matching or visual matching. Based on the above criteria children were categorized as 1-matcher, 2-matcher, 3-matcher, 4-matcher, 5-matcher, 6-matcher, and 7-matcher. To reduce number of categories for simpler interpretation, one-, two-, three- and four-matchers were combined as <u>sub-matchers</u>, five-, six-, and seven-matchers were combined as <u>matchers</u>. Table 10 shows number of matcher and non-matchers in each condition at pre- and post-test.

Table 10: Matchers and Non-matchers in Give the Same Task across the Four Conditions

Condition	Time of testing	# of matchers	# of non-matcher
Gesture + Diagram	Pre-test	17	6
	Post-test	19	4
Diagram only	Pre-test	11	13
	Post-test	17	7
No-gesture/ No-diagram	Pre-test	11	12
	Post-test	15	8
Reading	Pre-test	12	11
	Post-test	19	4

To examine whether the math intervention, and in particular performing grouping gesture and observing enclosed diagram around sets, helped children's performance in this task, a 3x2x2 likelihood ratio chi square was conducted with condition (diagram only, diagram + gesture, no diagram/no gesture, reading group), time of testing (pre- and post-test), and matching category (matcher versus non-matcher) as main effect, and number of children in each category as the dependent variable. Result showed that significantly more children were able to match a set with the given set at post-test than at pre-test $X^2(1, 93)= 8.411$, $p= 0.004$. Condition did not have an effect on children's performance.

Equivalence Judgment (EJ) task- Following, children's <u>accuracy</u> and <u>strategy</u> use when comparing numerical relation of two sets will be analyzed.

<u>Accuracy of Equivalence Judgment</u>: A binary score for every trial at pre- and post-tests was given: children received a score of 1 for every trial in which they compared the numerical relation of two sets correctly, and a score of 0 for the trials in which they provided an incorrect answer. Next, an overall score was computed for each child at pre-test and at post-test. Four children from gesture + diagram condition, three from diagram-only condition, one from no-

gesture/no-diagram condition, and two from the reading condition scored perfectly at pretest and were removed from the analysis. A 2x4x2x3x2 repeated analysis of variance (repeated ANOVA) with time of testing (pre- or post-test) as the within subject variable; condition, gender, school, and cardinality category at pre-test (cardinality-knower or subset-knower) as fixed factors; age and HTKS scores as covariates; and average accuracy as the dependent variable was conducted. On average, children significantly improved from pre-test to post-test in comparing numerical relation of two sets, $F(1, 83) = 5.258$, $p = 0.025$. Further, on average children who knew cardinality at pre-test scored significantly higher than did sub-knowers, $F(1, 83) = 9.861$, $p = 0.002$, and they improved more significantly from pre-test to post-test than did sub-knowers, $F(1, 83) = 3.833$, $p = 0.054$. Neither condition, nor its interaction with the time of testing, had a significant effect on Equivalence Judgment (EJ) task scores. Table 11 summarizes children's performance at pre- and post-test.

Table 11: Performance of Sub-knowers and Cardinality-knowers in Equivalence Judgment Task

Cardinality category (at pre-test)	Time of testing	Mean of average EJ scores (max=6)	Standard Error
Cardinality-knower	Pre-test	3.535	0.143
	Post-test	4.302	0.229
Sub-knower	Pre-test	3.000	0.131
	Post-test	3.098	0.134

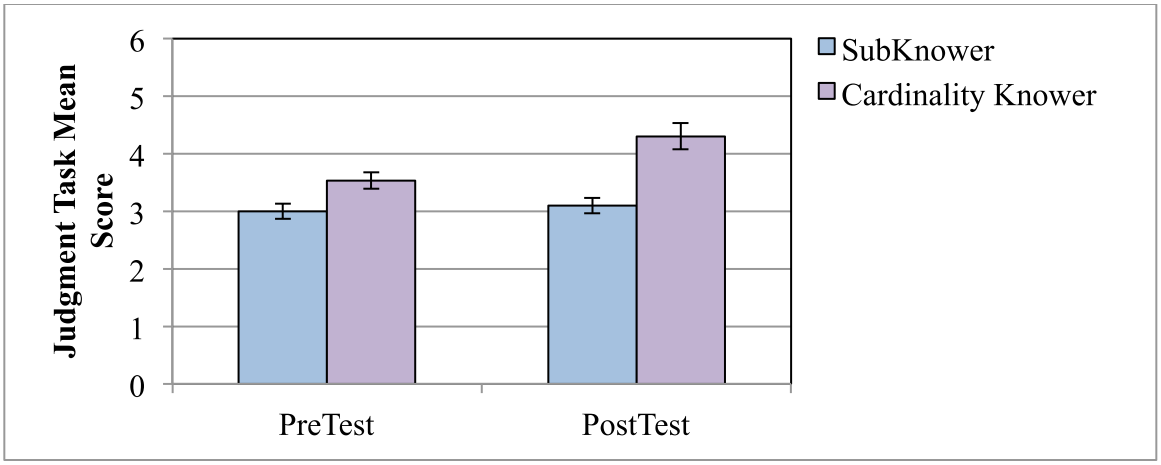

Figure 13: Average Equivalence Judgment Task Scores at Pre- and Post-test for Sub-knowers and Cardinality knowers

Since the two sets to be compared differed in only one item in each trial, to be correct, children had to compare the numerical relation between the two sets based on their exact cardinal values. Therefore, it is not surprising that children who knew cardinality at pre-test were significantly more accurate in judging the numerical equivalence of the two sets in the EJ task.

Next, children were categorized as "good judges" and "bad judges" based on their average EJ score. Median for pre-test scores and post-test scores was 3.00. Children who scored above median were categorized as "good judges" and those who scored equal or below the median were categorized as "bad judges". Table 12 shows number of "good judges" and "bad judges" in each condition at pre- and post-test.

Table 12: # of Good Judges and Bad Judges Across the Four Conditions at Pre- & Post-test

		Gesture + Diagram	Diagram only	No-gesture/No Diagram	Reading
Pre-test	# of good judges	6	9	9	9
	# of bad judges	17	15	14	14
Post-test	# of good judges	14	12	8	11
	# of bad judges	9	12	15	11

A binary logistic regression was conducted on post-test data with binary judging category as dependent variable; condition, cardinality category at pre-test, school, and gender as main factors; age and HTKS scores as covariates. Children who were already good judges at pre-test were removed from the analysis. All non-significant interactions were also removed from the model. Condition, X^2 (3, 60)= 8.136, p= 0.043, and whether or not children knew cardinality at pre-test, X^2 (1, 60)= 7.665, p= 0.006, and age F(1, 59)= 5.180, p= 0.023 had significant effect on children's likelihood of being good judges. Children who were in the gesture+diagram condition had a significantly better chance to become better judges comparing to the other conditions.

Use of Counting to Compare Numerical Relation Between Two Sets: Eight researchers were trained to code children's counting behavior using videotapes from pilot work. All researchers scored greater than 90% agreement when coding pilot children's counting behavior. Students' counting behavior on each trial was classified into one of four categories: (1) No Counting: the student responded without any apparent strategy; (2) Count 1-set: the student counted only one of the two sets; (3) Count 2 sets separately: the student counted each of the two sets separately; (4) Count 2 sets together: the student counted the two sets together as one whole set; for example, if one set had 3, and the other had 4 items, they would count 1, 2, 3, 4, 5, 6, 7. Since at each trial, the two sets differed only by one item, it was expected that if a child counted the two sets separately, he/she would have a higher chance to successfully compare the two sets. To examine this conjecture, a linear multinomial logistic regression mixed model analysis with counting behavior category (counting one or two sets; counting two sets together, no counting) at each trial as the dependent variable; condition, gender, and school as between subject variables; time of testing (pre vs. post-test) and trial number as within subject variables; and age and HTKS score as covariant variables was conducted. All non-significant interactions were removed from the model. Results confirmed that, in fact, counting behavior had a significant effect on children's likelihood of being correct in a trial, $F(3, 1097) = 15.823$, $p < 0.001$. In trials where children counted the two sets separately, on average they had a 74.5% (SE= 0.019) chance of being correct in comparing the numerical relation of the two sets, which was significantly higher than when they counted the two sets together (M= 0.481, SE=0.027), $t(1097) = 6.714$, $p < 0.001$, or when they did not engage in counting at all (M= 0.540, SE= 0.034), $t(1097) = 3.924$, $p < 0.001$. Further, when children counted only one of the sets, on average they had a higher chance of being correct in comparing the numerical relation of the two sets (M= 0.552, SE= 0.061) than when they counted the two sets together, $t(1097) = 2.202$, $p = 0.028$.

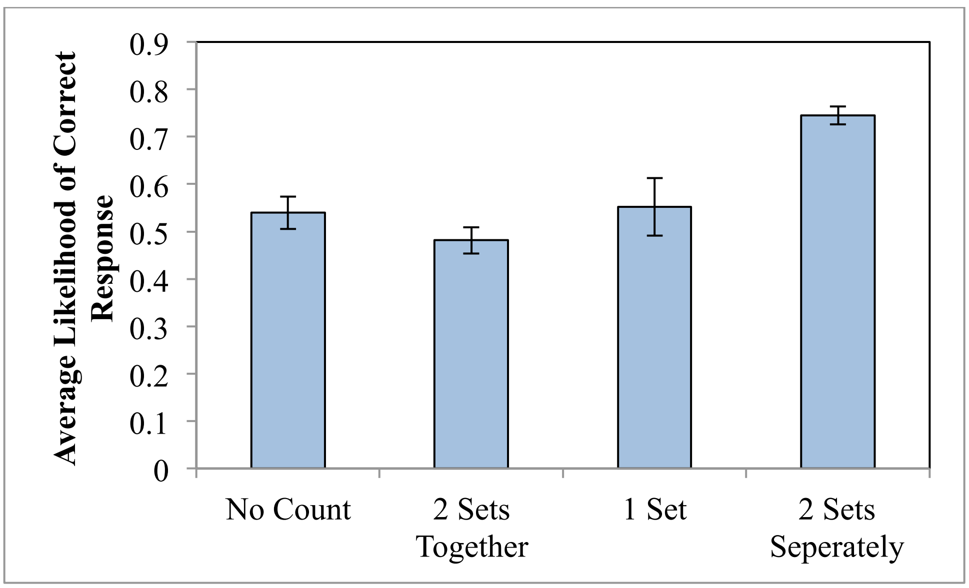

Figure 14: Average Likelihood of Correct Response for Each Counting Behavior

Based on the above results, strategies were ranked from the most to the least advanced: Count 2 sets separately, Count 1-set, Count 2 sets together, No Counting. If a child engaged in two different counting behavior in one trial, the more advanced strategy was coded for that trial. For example, if she counted two sets together and then counted the two sets separately, the coded strategy was "count 2 sets separately". Next, percentage of questions each child engaged in each of the above counting behaviors was calculated, and were averaged over condition at pre- and post-test. Table 13 summarizes the average frequency of each counting behavior in each condition at pre- and post-test.

Table 13: Counting Behavior Across the Four Conditions at Pre- and Post-test

Condition	Time of testing	No counting	Count 1 set	Count 2 sets Separately	Count 2 sets together
Gesture + Diagram	Pre-test	M= 0.174, SE= 0.068	M= 0.138, SE= 0.057	M= 0.4861, SE= 0.090	M= 0.203, SE= 0.072
	Post-test	M= 0.094, SE= 0.051	M= 0.094, SE= 0.040	M= 0.573, SE= 0.089	M= 0.283, SE= 0.079
Diagram Only	Pre-test	M= 0.153, SE= 0.070	M= 0.035, SE= 0.017	M= 0.451, SE= 0.086	M= 0.361, SE= 0.083
	Post-test	M= 0.132, SE= 0.060	M= 0.007, SE= 0.007	M= 0.521, SE= 0.089	M= 0.340, SE= 0.084
No Gesture/No Diagram	Pre-test	M= 0.160, SE= 0.064	M= 0.109, SE= 0.031	M= 0.420, SE= 0.089	M= 0.311, SE= 0.078
	Post-test	M= 0.065, SE= 0.041	M= 0.094, SE= 0.034	M= 0.464, SE= 0.095	M= 0.377, SE= 0.090
Reading	Pre-test	M= 0.138, SE= 0.062	M= 0.167, SE= 0.057	M= 0.5935, SE= 0.087	M= 0.102, SE= 0.042
	Post-test	M= 0.138, SE= 0.064	M= 0.044, SE= 0.019	M= 0.645, SE= 0.087	M= 0.174 SE= 0.06

It is important to note that many children engaged in variety of counting behavior from one trial to the other, and sometimes even in the same trial. At pre-test, forty five children (48%) always engaged in the same behavior, twenty nine of them always counted the two sets separately, eight of them always counted the two sets together, one of them always counted only one of the sets, and seven of them never counted in any of the trials; the rest of the children (52%) engaged in variety of counting behaviors from one trial to the next. Similarly, at post-test, fifty one children (54%) always engaged in the same behavior, thirty three of them always counted the two sets separately, thirteen of them always counted the two sets together, and five of them never counted in any of the trials; the rest of the children (46%) engaged in variety of counting behaviors from one trial to the next. Such behavior was in fact expected. From Siegler's (1987) research on children's strategy use, we know that children often engage in variety of strategies until they become proficient in using the more advanced strategies resulting in higher accuracy.

A multivariate repeated analysis of variance with average frequency of counting behavior (no-count, count one set, count two sets separately, count two sets together) as dependent variables; condition, gender, school, and cardinality category at pretest (cardinality knower vs. sub-knower) as dependent variables; age and HTKS score as covariates was conducted. Condition did not have a significant effect on children's counting behavior. On the other hand, on average children who knew cardinality at pre-test counted the two sets separately (correct strategy) significantly more frequently, both at pre-test and at post-test, than children who did not know cardinality at pre-test, $F(1, 93)= 14.749$, $p< 0.001$.

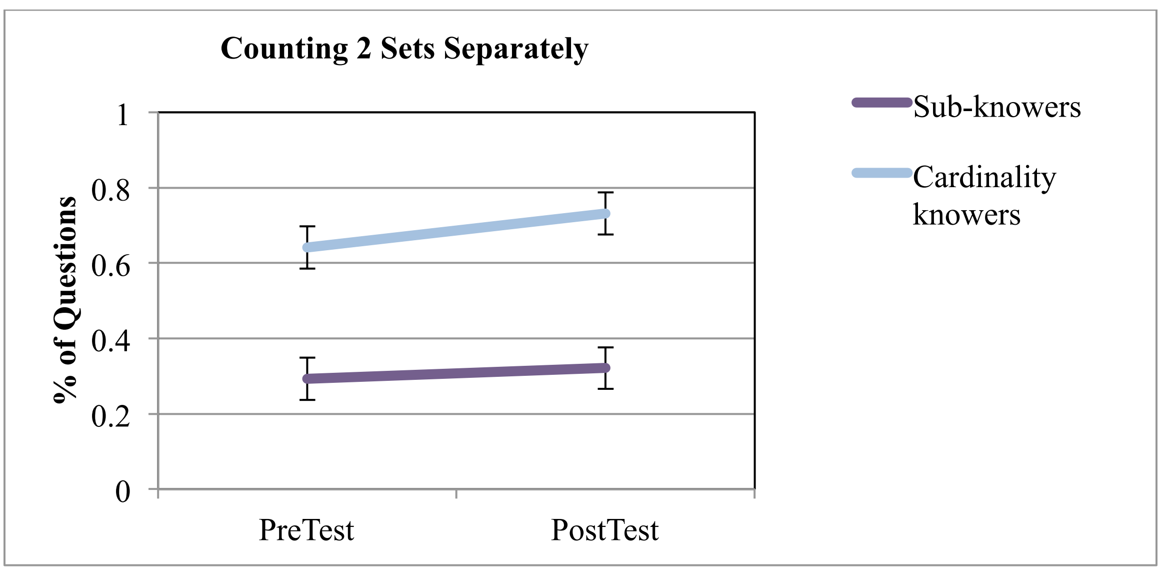

Figure 15: Frequency of Counting 2 Sets Separately at Pre- and Post-test for Sub-knowers and Cardinality Knowers

In addition, on average children with higher executive functioning skills counted the two sets separately significantly more often than children with lower executive functioning skills, $F(1, 93)= 4.016$, $p < 0.048$.

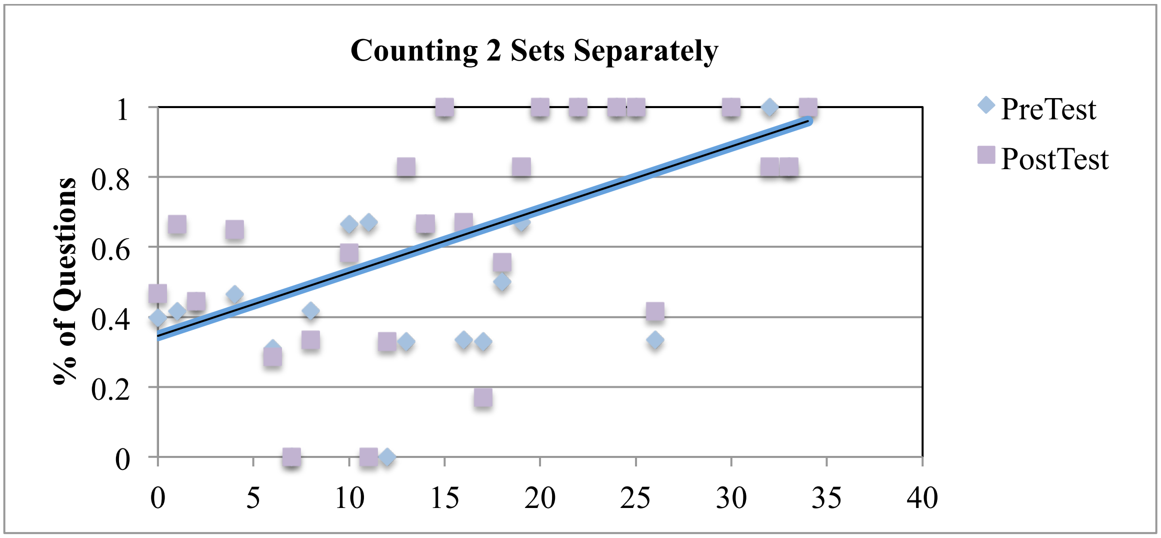

Figure 16: Frequency of Counting 2 Sets Separately with Respect to HTKS scores

Further, interaction between cardinality category and time of testing had a significant effect on children's counting two sets together behavior, $F(1, 93)= 5.278$, $p= 0.024$. Children who did not know cardinality at pre-test counted two sets together significantly more often at post-test (M= 0.415, SE= 0.061) than at pre-test (M= 0.265, SE= 0.057), but children who knew cardinality at pre-test contend the two sets together significantly less frequently at post-test (M= 0.199, SE= 0.049) than they did at pre-test (M= 0.231, SE= 0.047).

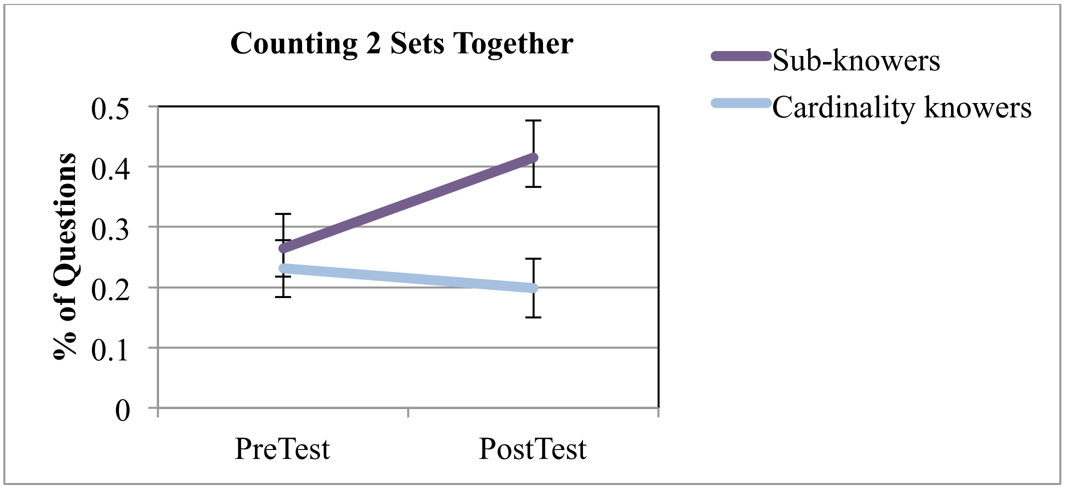

Figure 17: Frequency of Counting 2 Sets Together at Pre- and Post-Test for Sub-knowers and Cardinality Knowers

This result may seem to be baffling— why did children who did not know cardinality at pre-test make the counting mistake more often at post-test after the intervention? One plausible explanation is that when children start to use counting as a way to compare the numerical relation of two sets, they are not yet proficient enough in using their counting skills flexibly, so they may count the two sets together by mistake. Of course, counting two sets together may be a plausible strategy when they try to add the two sets together, but it is an ineffective strategy when they try to compare numerical relation of two sets. In other words, sub-knowers who were not counting at pre-test have started counting at post-test but they sometimes make the mistake of counting the two sets together. Figure 18 shows the frequency of each counting behavior at pre- and post-test across sub-knowers and cardinality knowers.

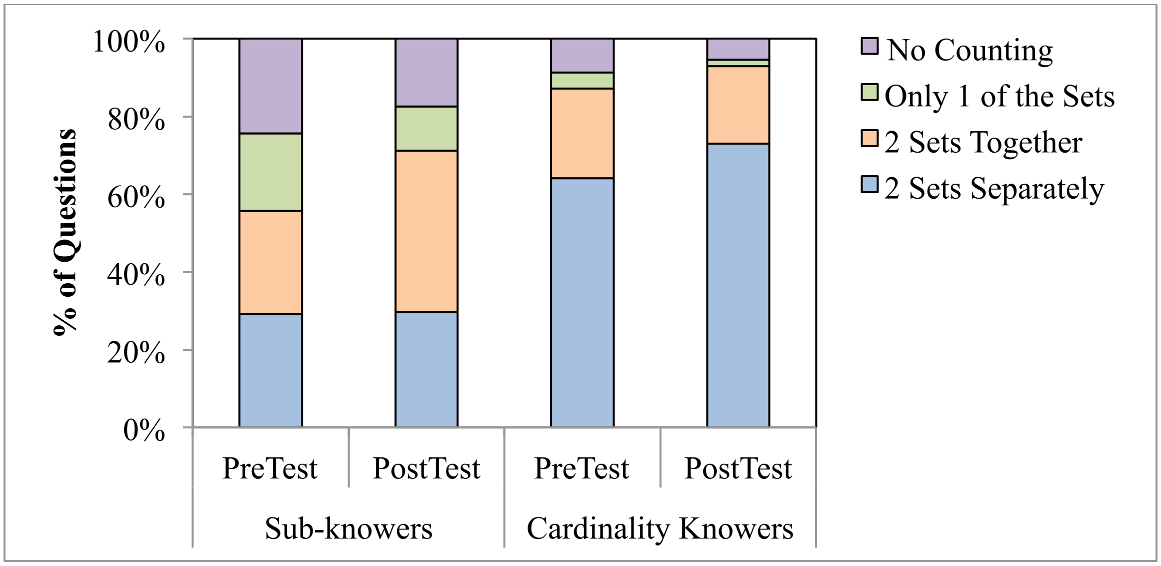

Figure 18: Frequency of the Four Counting Behaviors in Equivalence Judgment Task at Pre- and Post-test for Sub-knowers and Cardinality Knowers

As it is evident in this figure, sub-knowers engaged in counting more often at post-test and it seems that children who were not counters at pre-test, and have started counting at post-test, counted the two sets more often together than separately. In fact, data agree with that speculation. I categorized sub-knowers who did not count on 50% or more of the questions at pre-test, as "non-counters". Twenty-seven percent (11 out of 41 children) of sub-knowers were non-counters at pre-test. These non-counters counted two sets together more often (M= 0.334, SD= 0.387) than separately (M=0.227, SD= 0.327) at post-test. Therefore, we can conclude that counting two sets together in an intermediate behavior in the course of the path to use counting effectively when comparing the numerical relation of two sets.

Make the same (MS) task- Following, children's success rate and their counting behavior when producing two equal sets will be analyzed.

Success in Producing Two Equal Sets- A binary score for every trial at pre- and post-tests was given: children received a score of 1 for every trial in which they successfully made the two sets equal, and a score of 0 for the trials in which they failed to make the two sets equal. Next, an overall score was computed for each child at pre-test and at post-test. Two children from gesture+diagram condition, six from diagram-only condition, seven from no-gesture/no-diagram condition, and four from the reading condition scored perfectly at pretest and were removed from the analysis. A 2x4x2x3x2 repeated analysis of variance (repeated ANOVA) with time of testing (pre- or post-test) as the within subject variable; condition, gender, school, and cardinality category at pre-test (cardinality-knower or subset-knower) as fixed factors; age and HTKS scores as covariates; and average accuracy as the dependent variable was conducted. On average, children who knew cardinality at pre-test became significantly better in making the two sets equal at post-test (M= 3.56, SE= 0.191) comparing to pre-test (M= 2.00, SE= 0.221), but sub-knowers did not perform significantly different at pre-test (M= 1.09, SE= 0.315) and than they did at post-test (M= 1.41, SE= 0.32), F (1, 73)= 4.713, p= 0.034.

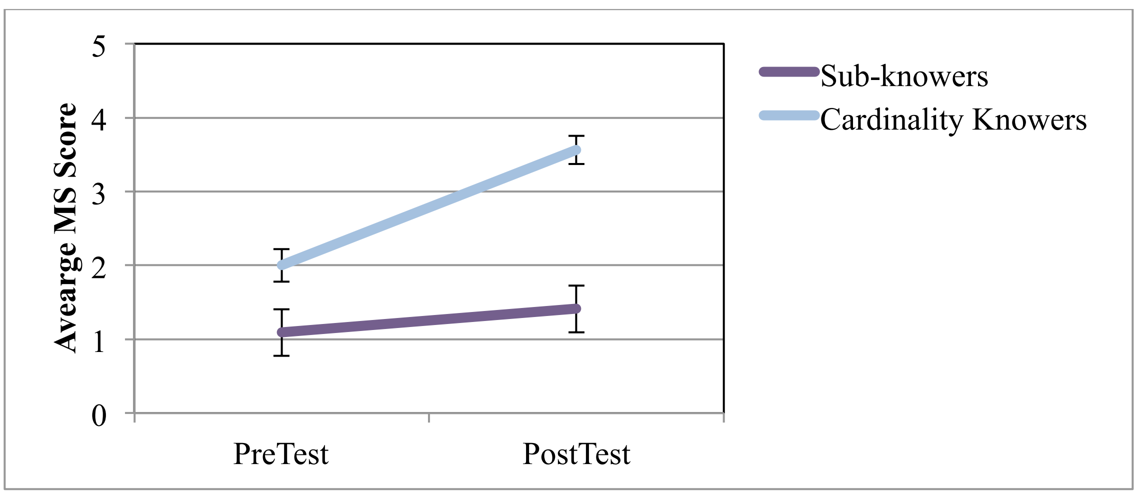

Figure 19: Average Scores in Make the Same Task at Pre- and Post-test for Sub-knowers and Cardinality Knowers

Further, HTKS scores explained significant variance in children's MS scores, F (1, 73)= 8.166, p= 0.006. Neither condition, nor its interaction with the time of testing, had a significant effect on Make the Same (MS) task scores.

Use of Counting When Making Two Sets Equal: Children engaged in variety of strategies when they tried to make the two sets equal. Some relied on visual matching, arranging blocks in the same visual pattern in the two sets, for example, they placed all blocks in a row, in a column, or based on a dice pattern; some would take away all blocks from both sets and then add the same number to both; and some would count the two sets and added or removed blocks from the sets. Here we focus on children's counting behavior to examine whether they engaged in the same counting behavior as they did when solving Equivalence Judgment (EJ) problems. As mentioned previously, when comparing numerical relation of two sets of candies in the Equivalence Judgment (EJ) task, children sometimes counted the two sets together as one whole set. In the EJ task, candies were presented as printed red squares on two separate cards— would children make the same mistake if they were comparing actual blocks that they could manipulate? This analysis will help to understand whether counting-two-sets-together is a task-specific mistake, or a mistake that children make in different contexts when they are comparing numerical relation of two sets.

Eight researchers were trained to code children's counting behavior using videotapes from pilot work. All researchers scored greater than 90% agreement when coding pilot children's counting behavior. Children's counting behavior on each trial were classified as one of the three categories: (1) No Counting: the student responded without any apparent strategy; (2) Count 1-set: the student counted only one of the two sets; (3) Count 2 sets separately: the student counted each of the two sets separately; (4) Count 2 sets together. In cases where two counting behaviors were observed, the most advanced behavior was coded for that trial. Next, percentage of questions each child engaged in each of the above counting behaviors was calculated, and was averaged over condition at pre- and post-test.

A multivariate repeated analysis of variance with average frequency of counting behavior (no-count, count one set, count two sets separately, count two sets together) as dependent variables, condition, gender, school, and cardinality category at pretest (cardinality knower vs. sub-knower) as dependent variables, age and HTKS score as covariates were conducted. Condition did not have a significant effect on children's counting behavior. On the other hand, on average children who knew cardinality at pre-test counted the two sets separately (correct strategy) significantly more frequently, both at pre-test and at post-test, than children who did not know cardinality at pre-test, $F(1, 92)= 12.983$, $p< 0.001$, a result similar to what we found in the Equivalence Judgment (EJ) task. Also, executive functioning skills measured by HTKS scores explained a significant portion of the variance of children's performance in the MS task, $F(1, 92)= 10.025$, $p= 0.002$, again a result similar to what we found in the Equivalence Judgment (EJ) task. Figure 20 shows the frequency of each counting behavior at pre- and post-test.

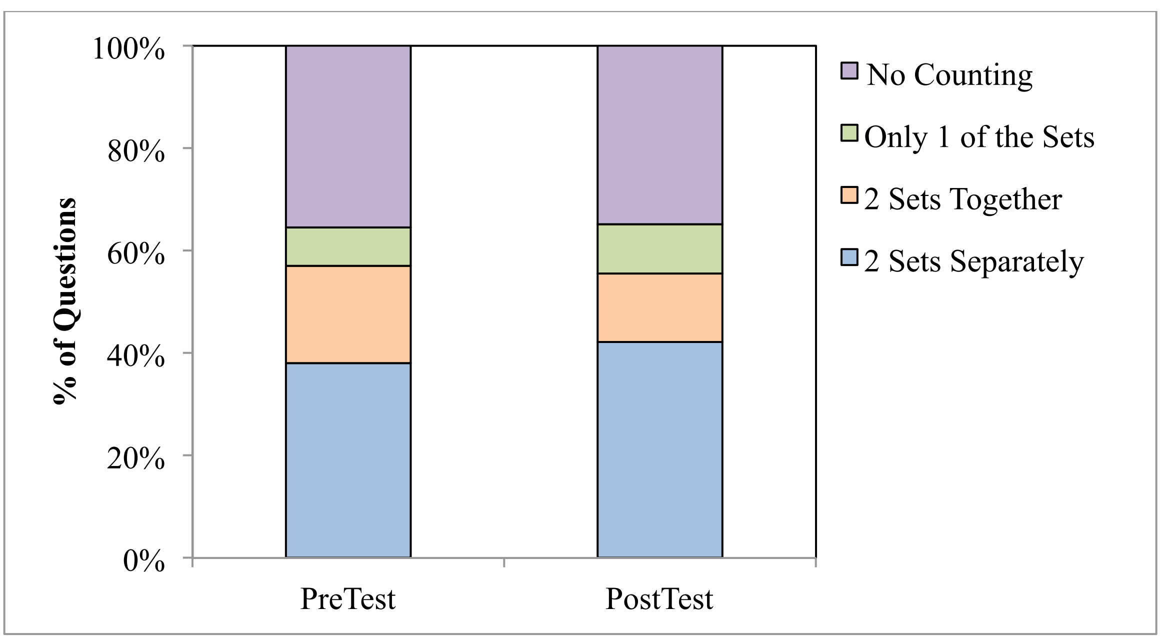

Figure 20: Frequency of the Four Counting Behaviors in Equivalence Judgment Task at Pre- and Post-test for Sub-knowers and Cardinality Knowers

Changes from Pre-test to Post-test- Summary

Key findings related to changes from pre- to post-test results are summarized in this section. These findings will be discussed in detail in the next chapter.

- Playing with Pop the Bubble activity helped preschoolers significantly improve their math skills. On average children in all three math conditions improved in their TEMA scores from pre- to post-test, while children in the reading condition did not significantly improve.
- Playing with Pop the Bubble activity helped preschoolers to improve their skills in recognizing written numerals, reciting the counting numbers, enumerating sets of objects, and comparing quantities. On average children in all three math conditions improved performance in specific TEMA items that were directly related to the above mentioned skills, while children in the reading condition did not significantly improve in these items from pre- to post-test.
- Grouping sets in a circular diagram, with or without the grouping gesture, helped children to improve their understanding of cardinality measured by Wynn's Give a Number task.
- Grouping sets in a circular diagram with the grouping gesture helped children who were not good judges of numerical relation of two sets at pre-test to become better judges at post-test.
- Executive functioning skills explained significant portions of the variance in all math measures.
- Children sometimes count two sets together as one whole set when they are comparing numerical relation of two sets. It seems that this counting behavior is an

intermediate stage when children are becoming proficient in using their counting skills to compare their two sets' numerosities.

- Children who know cardinality count the two sets together significantly less often than those who do not know cardinality.
- Children with higher executive functioning skills engage in the correct strategy of counting sets separately significantly more often then those with lower executive functioning skills.

Learning during the Math Intervention

In this section, I provide a precise depiction of the course of learning and strategy use of students in the three math conditions during the intervention. Students in the reading condition are excluded from the analysis.

Accuracy of judgment- Instead of computing an overall score for each session, trials in each session were treated as repeated observations; thus, children were given a binary score for every trial in the four intervention sessions. A Score of 1 was given if the child compared the numerical size of two sets correctly and a score of 0 was given if the child provided an incorrect answer. To examine whether condition had an effect on children's likelihood of being correct in comparing numerical relation of two sets and whether children in a certain condition became more accurate in their judgments at a faster pace than students in the other conditions, a linear binary logistic regression mixed model analysis with score at each trial as the dependent variable; condition, gender, and school as between subject variables; session number (first, second, third, or fourth); and question number as within subject variables; age and HTKS score as covariates were conducted. All non-significant interactions were removed from the model. The hypothesis was that children in all three math conditions improve in their judgment accuracy over time, but children in the diagram+gesture and diagram only conditions improve significantly more than do students in the no-diagram/no-gesture conditions. Further, it was expected that children with lower HTKS scores benefit from performing the grouping gesture in addition to seeing the enclosed diagrams, whereas for children with higher HTKS simply seeing the enclosed diagrams would be sufficient.

Results of the linear binary logistic regression mixed model analysis are summarized in Table 14.

Table 14: Summary of the Linear Binary Logistic Regression Mixed Model Analysis on the Math Intervention Data

Source	F	df1	df2	Sig
Condition	12.002	2	1378	0.000
School	1.736	2	1378	0.117
Session	5.255	3	1378	0.001
Question	3.712	4	1378	0.005
Age	36.122	1	1378	0.000
HTKS	6.724	1	1378	0.010
Condition x Session	0.283	6	1378	0.945
Condition x HSKT	8.916	2	1378	0.000

Condition had a significant main effect on children's accuracy, $F(2, 1378)= 12.002$, $p< 0.001$. Post-hoc sequential Bonferroni pairwise analysis confirmed that on average, children who played the version of the "pop the bubble" activity with enclosed diagrams around the sets, with ($M= 0.741$, $SE= 0.020$), $t(1378)= 2.489$, $p= 0.026$, or without the grouping gesture ($M= 0.7305$, $SE= 0.02$), $t(1378)= 3.146$, $p= 0.005$, were more likely to correctly compare numerical relation of the sets than children who played the version of the activity with no-gesture and no-diagram ($M= 0.6129$, $p= 0.02$).

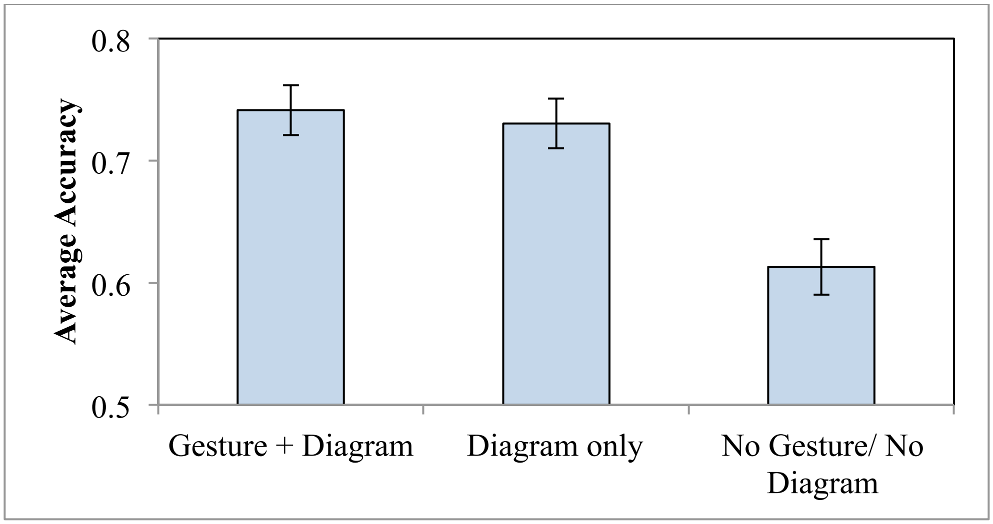

Figure 21: Average Accuracy in the Math Software Across the Three Conditions

Session of the intervention had also a significant main effect on children's accuracy, $F(2, 1378) = 5.255$, $p = 0.001$. Recall that level of the game was easiest in session one (number of fish in each set was between 2 to 5 and children were only asked to judge the equivalence of the two sets), harder in session two (each set had 5 to 7 fish, and children were only asked to judge the equivalence of the two sets), and hardest in sessions three and four (each set had 5 to 8 fish, and children were asked to judge the equivalence of the two sets and in cases the two sets were not equal they were asked to indicate which set has more fish). Sessions three and four had the same level of difficulty. As expected, post-hoc least significant difference pairwise analysis showed that on average, children were more likely to correctly compare the numerical relation of two sets in session one ($M = 0.7057$, $SE = 0.024$), $t(1378) = 2.346$, $p = 0.019$, and in session two ($M = 0.7571$, $SE = 0.02$), $t(1378) = 3.923$, $p < 0.001$, than in session three ($M = 0.6143$, $SE = 0.03$). Interestingly, there was no significant difference between sessions one and two even though session two was more advanced than session one. Further, on average, children were more likely to correctly compare the numerical relation of two sets in session four, the last session of the intervention ($M = 0.7029$, $SE = 0.024$) comparing to session three, which had the same level of difficulty as did session four, t(1378)= 2.436, p= 0.015, indicating that on average children in all three conditions were benefitting from the intervention.

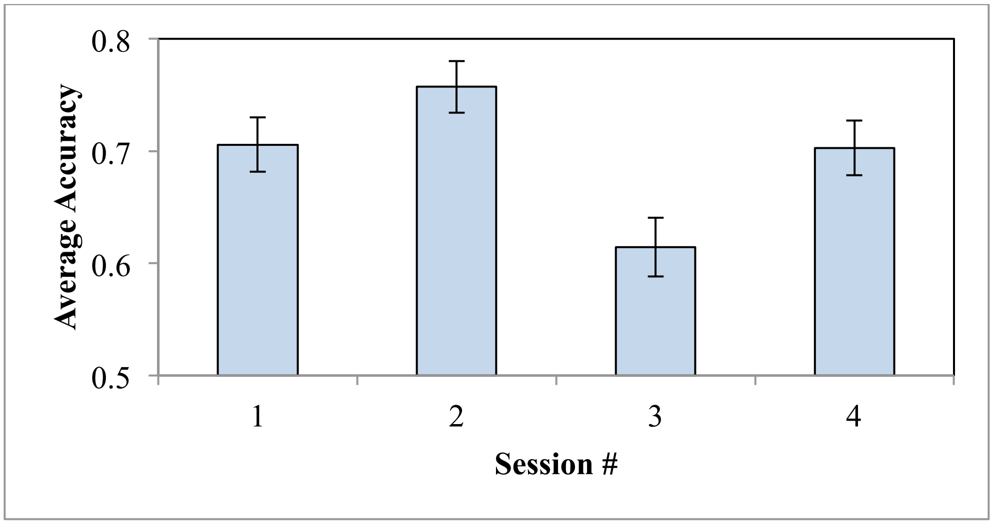

Figure 22: Average Accuracy in the Math Software Across the Four Sessions of Intervention

Further, children improved within each session from question 1 to question 5, F(4, 1378)= 3.712, p= 0.005. Post-hoc sequential Bonferroni pairwise analysis showed that on average, children were more likely to provide a correct answer in question five (M= 0.7607, SE= 0.023) than in question one (M= 0.6107, SE= 0.030), t(1378)= 3.738, p=0.002.

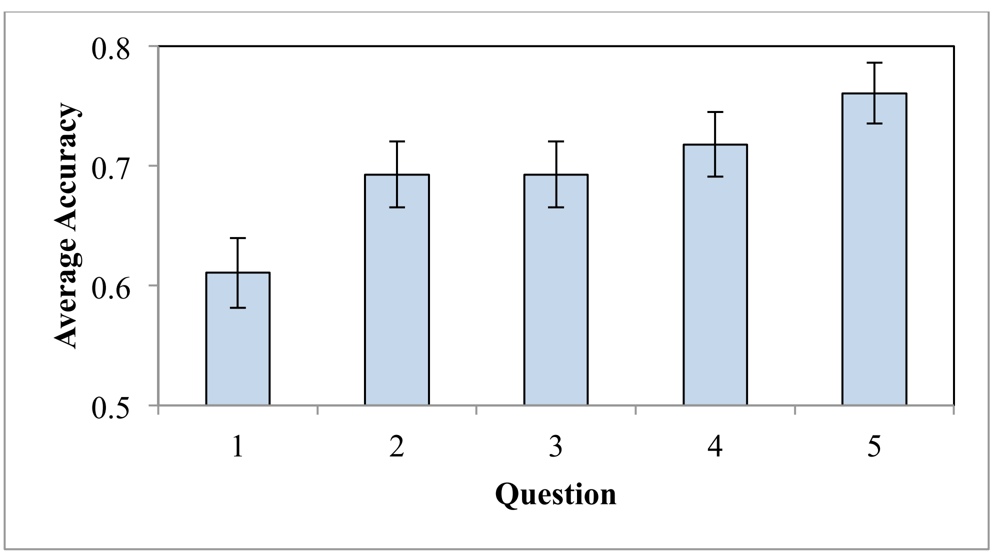

Figure 23: Average Accuracy in the Math Software in each Trail

Finally, analysis showed a significant interaction effect between condition and HTKS scores, $F(2, 1378)= 8.916$, $p< 0.001$. Figure 24 shows average accuracy versus HTKS scores for each of the three conditions. Post-hoc analysis showed that the slope of line for gesture + diagram condition is not significantly different than zero $t(460)= 0.259$, $p= 0.796$, whereas the slope of the lines for diagram-only condition, $t(475)= 2.715$, $p= 0.007$, and for no-gesture/ no-diagram conditions are significantly more than zero $t(465)= 4.058$, $p< 0.001$. Therefore, we can speculate that children with lower HTKS scores benefited from performing the grouping gesture in addition to seeing the enclosed diagrams, whereas for children with higher HTKS, simply seeing the enclosed diagrams was sufficient. Future studies with bigger sample sizes are needed to specifically investigate possible benefits of gesturing for children with lower executive functioning skills.

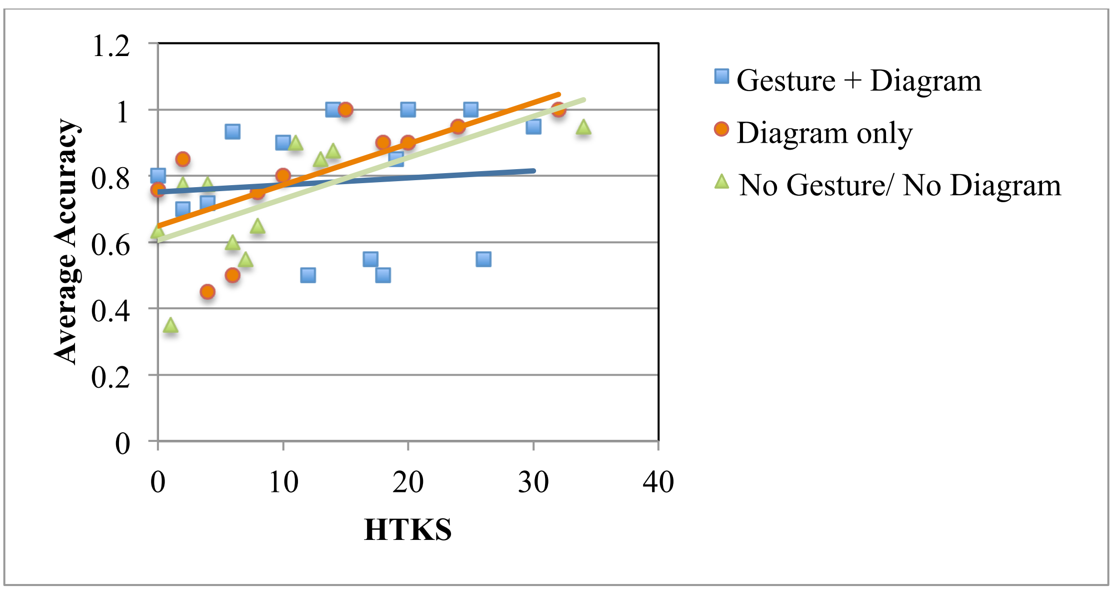

Figure 24: Average Accuracy in the Math Software versus HTKS Scores

Counting behavior- In a previous study (Jamalian, 2013), and in the current study during pre- and post-test sessions, we observed that preschoolers sometimes count two sets together as one set when comparing the numerical relation of the two. For example, to compare a set of 5 candies with another set of 4 candies, they count all candies together (1, 2, 3, 4, 5, 6, 7, 8, 9) and then make a judgment about the relation between the two sets. Such counting behavior is obviously not a correct strategy to judge numerical equivalence of two sets. Here, I asked whether seeing enclosed diagram around each set and/or performing grouping gestures around them can help to eliminate such counting mistakes. To investigate the above question, I counted number of students in each condition who counted the two sets together at least once during each session of the intervention. Interestingly, throughout the intervention, none of the children in the gesture + diagram condition counted the two sets together. On the other hand, 35% of children (8 out of 23) in the no-gesture/no-diagram condition and 17% of children (4 out of 24) in the diagram-only condition counted the two sets together at least once during session one. A 3x4 likelihood ratio chi square with condition (diagram only, diagram + gesture, no diagram/no gesture) and session number as independent variables, and number of children who made the mistake as the dependent variable, showed that in fact, condition had a significant main effect on children's counting behavior, X^2 (2, 92)= 31.722, $p< 0.001$. Further, as the intervention proceeded, fewer children made the above mistake, X^2(3, 92)= 28.442, $p< 0.001$, and by session four none of the children in the diagram-only condition counted the two sets together.

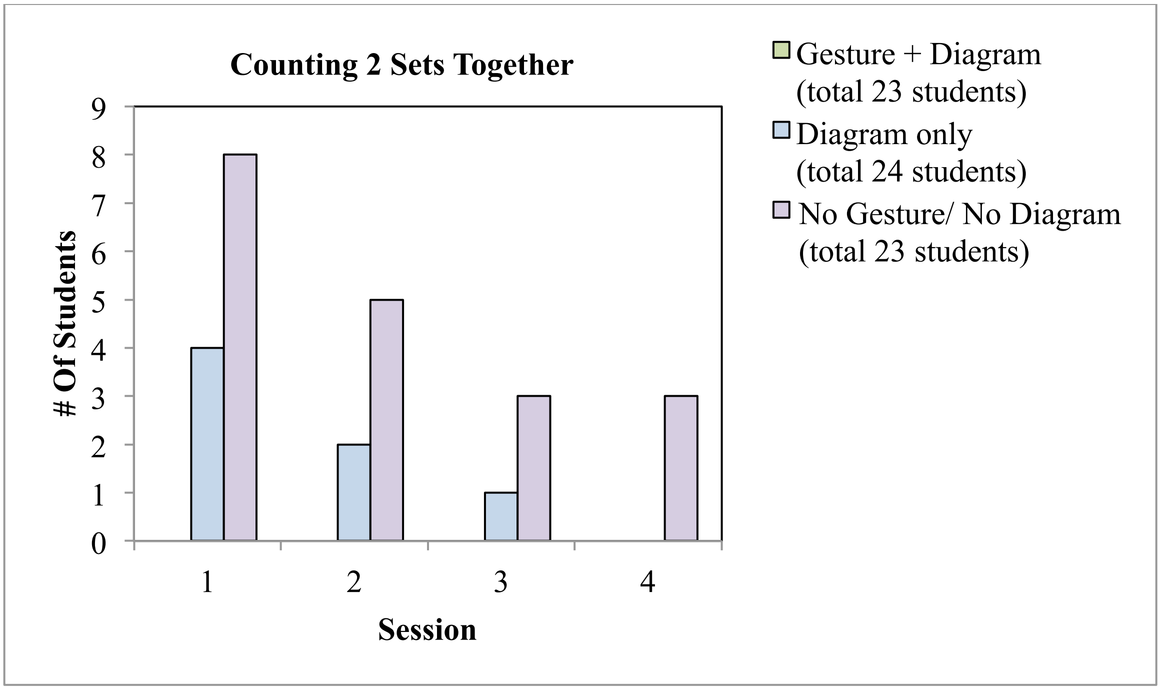

Figure 25: Number of Children who engaged in Counting 2 Sets Together in each of the Math Intervention Sessions Across the Three Conditions

It is important to note that children did not make the counting-together mistake as frequently as they did during the pre- and post-test Equivalence Judgment (EJ) task. Such behavior was in fact expected, as during the game play, children were explicitly asked to indicate how many fish are in each set, one after the other, and then were asked to make a judgment about the numerical relation between the two sets. By contrast, in the pre- and post-test Equivalence Judgment (EJ) task, children were only prompted to “check and tell the [researcher] if the boy and the girl have the same number of candies”. In fact, it is surprising that children engaged in such behavior during the intervention at all, where fish in each set differed in color from fish in the other set and in each trial the farmer explicitly asked, for example, “How many green fish are there?” and then “How many blue fish are there?”

A 3x4x2 Poisson loglinear regression with condition (diagram only, diagram + gesture, no diagram/no gesture), session number, and set (set of fish on the left side of the screen versus set of fish on the right side of the screen) as independent variables, and number of times counting-together mistake was encountered as the dependent variable, showed that condition had a significant main effect on the frequency of the counting-together mistake, X^2 (2, 92)= 73.696, p< 0.001. As mentioned previously, children in the gesture+diagram condition never made the above mistake, whereas, overall children in the no-gesture/no-diagram condition made that mistake in 19% of the trials (22 out of 115 [23 students x 5 trails per session]) in session one when they were asked to indicate how many fish are in the left set, and in 2% of the trials (2 out of 115) when they were asked to indicate how many fish are in the right set. Table 15 summarizes the frequency of count-together mistake in each condition for each set in each session. Children counted two sets together more frequently when they were asked to indicate how many fish are in the left set than when they were asked to indicate how many fish are in the right set, X^2 (1, 92)= 106.904, p< 0.001. Further, they counted the two sets together less frequently as the intervention proceeded, (3, 92)= 98.490, p< 0.001.

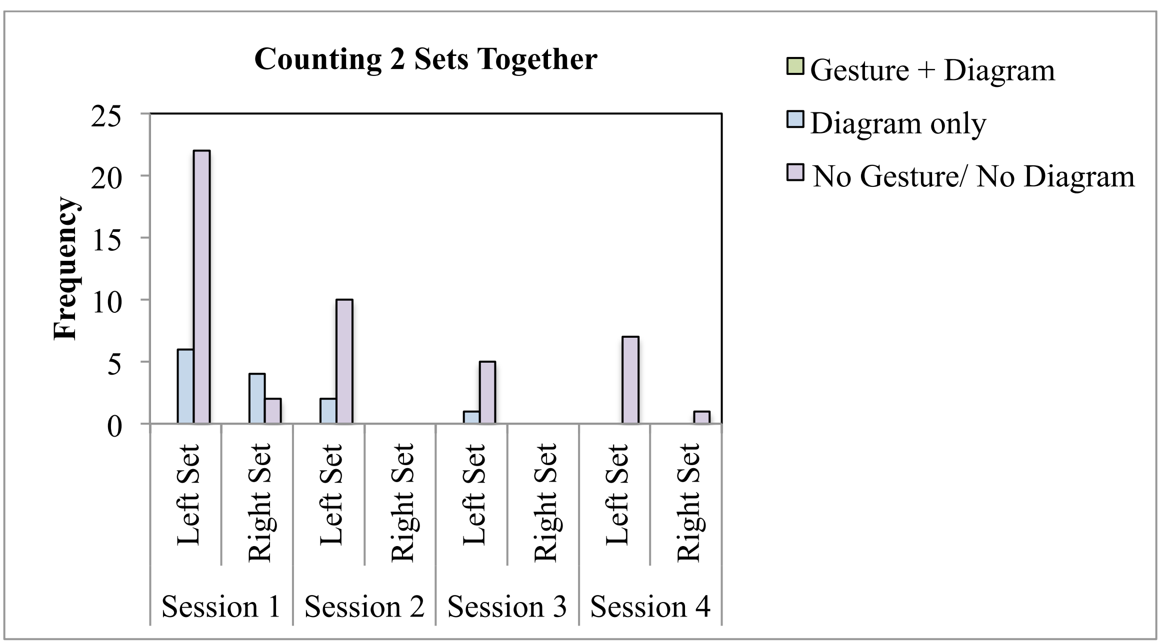

Figure 26: Frequency of Counting 2 Sets Together in each of the Four Math Intervention Sessions Across the Three Conditions

Table 15: Frequency of Counting 2 Sets Together in each of the Four Math Intervention Sessions Across the Three Conditions

Condition	Session	Set	Frequency of counting two sets together mistake
Gesture + Diagram	1	Left	0
		Right	0
	2	Left	0
		Right	0
	3	Left	0
		Right	0
	4	Left	0
		Right	0
Diagram-only	1	Left	6
		Right	4
	2	Left	2
		Right	0
	3	Left	1
		Right	0
	4	Left	0
		Right	0
No gesture/ No diagram	1	Left	22
		Right	2
	2	Left	10
		Right	0
	3	Left	5
		Right	0
	4	Left	7
		Right	1

From Intervention to Post-Test

To determine whether children's performance during the intervention was related to their performance in TEMA test at post-test, a univariate analysis of variance (ANOVA) with post-test TEMA scores as dependent variable; condition, school, and gender as fixed factors; age, HTKS score, and their average score in session 4 of the game as covariates, was conducted. Results showed that in fact performance in the last session of the intervention explained a significant portion of the variance in post-test TEMA scores, $F(1, 70)= 9.679$, $p= 0.003$.

Learning during the Math Intervention- Summary

Following, key findings related to children's learning path during the math intervention are summarized. These findings will be discussed in detail in the next chapter.

- All children significantly improved in comparing numerical relation between two sets over the course of the intervention.
- All children improved within each session, as on average they were significantly more likely to correctly compare the two sets in the final trial as they were in the initial trial of the session.
- Children who saw sets in a circular diagram, with or without the grouping gesture, performed significantly better than children who did not see the diagram.
- Grouping gesture, in addition to the circular diagram, was beneficial for children with lower executive functioning skills, but redundant for children with higher executive functioning skills.
- Grouping sets with gesture prevented children from counting the two sets together.

CHAPTER 4: DISCUSSION

Diagrams and Gestures

Are Diagrams and Gestures synonyms?

Preschoolers either saw an enclosed (circular) diagram around sets or they performed a grouping gesture around sets as well as seeing the diagrams around them. In a third condition, they neither saw a diagram nor performed a grouping gesture. Pre- and post-test results show that the circular diagram around sets with or without the accompanying grouping gesture helped children to learn cardinality measured by their performance on the Wynn's "Give a Number" task, and helped them to improve their overall math competence measured by their performance on the TEMA-3 standardized test of early math. Children who neither saw the circular diagram nor performed a grouping gesture around the sets did not benefit from the intervention as much, although they outperformed children who did not receive math instruction at all (i.e. children in the reading condition).

In addition, over the course of the intervention children who saw the circular diagram around sets, with or without the accompanying grouping gesture, improved significantly in their accuracy of numerical comparison on the "pop the bubble" math activity, more than did children who neither saw an enclosed diagram around sets nor performed a grouping gesture around them.

The grouping gesture is similar in form to an enclosed circular diagram— they both put a boundary around the items that belong to a set, represent, and highlight sets (rather than individual items) by *containing* items of the set within them. Lakoff and Nunez (2000) argue that we often conceptualize sets as containers, as if they contain their items *in* them. Tversky (2011) explains that bars in a bar graph are containers that contain similar instances, separating their content from dissimilar instances *in* other bars. Talmy (1999) uses the term “surround” to label the literal and metaphoric meaning of the linguistic proposition *in*; and agrees that that these fine linguistic structures, such as *in*, *on*, *out*, *over*, *across*, provide a systematic framework that schematizes space and uses spatial relations to convey meaning. Like language, gestures and diagrams schematize and abstract to convey meaning; unlike language, gestures and diagrams represent and organize information directly in space using relations in space. The grouping gesture and the circular diagram both represent interior, exterior, and boundary aspects of a container and disregard its size or volume to conceptualize a group of items as a whole set, an object in itself that may have distinct properties such as a cardinal value.

Unlike most words, diagrams and gestures, specifically iconic gestures, may have physical similarities (literal or metaphorical) to the things they refer to, for example, illustrating corner of a building by making an L shape gesture or by showing an L shape diagram on the paper; or representing sets by making a grouping gesture or by showing a circular diagram around them. Tversky (2011) explains that diagrams are the consequence of gestures, the drawing actions. Enfield (2003) argues that gestures establish “virtual diagrams” in space. Gestures as virtual diagrams or diagrams as crystalized gestures share representative features like position and form in space to convey meaning— and when they both have the same form they can be regarded as “synonyms”. In fact, we tend to replace gestures for diagrams when we cannot draw. In one study, participants gestured to solve problems when they did not have access to paper and pencil but drew diagrams when they could (Kessell & Tvesrky, 2006). Architects gesture to come up with new ideas and to communicate about their ideas when they are refrained from sketching, but when they have access to a piece of paper and pencil, they start sketching and gesturing over the sketches to design and to communicate about their designs (Bilda and Gero, 2006).

Thus, gestures and diagrams may convey similar meanings through similar representations, visualize and map thought directly in space using relations in space in a similar manner, and one could replace the other as synonyms. Grouping gesture and circular diagrams are synonyms in representing sets and either one may be enough to help children improve their understanding of cardinality. In the current study, we learned that the circular gesture alone is effective in helping children to improve mathematical thinking; future studies are necessary to test whether a grouping gesture alone is as effective.

Gestures accompanying speech, too, sometimes convey the same information as it is conveyed in the speech (McNeil, 1992). However, these redundant gestures help both the speakers in finding words (Krauss & Hadar, 2001), and the listener in comprehending the speech (Goldin-Meadow & Sandhofer, 1999). Children instructed in mathematical equivalence problems (e.g., 3 + 4 + 5 = __ + 5) are more likely to learn when the instruction includes speech and gesture than when it includes only speech (Cook & Goldin-Meadow, 2006, Perry, Berch, & Singleton, 1995; Singer & Goldin-Meadow, 2005). This has also been found with preschoolers instructed in symmetry (Valenzeno, Alibali, & Klatzky, 2003) and mental rotation tasks (Enrlich, Levine, & Goldin-Meadow, 2006), and with non-conservers instructed in Piaget’s liquid conservation task (Ping & Goldin-Meadow, 2008). Language conveys meanings indirectly through symbols, but gestures (as well as diagrams) represent information directly by mapping them in space. Hence they may be easier to interpret and easier to remember (Jamalian, Giardino, & Tversky, 2013; Tversky et al., 2013). Whereas redundant speech accompanying gestures helps thinking, learning, and communicating, redundant diagrams accompanying gestures may not add much— at least in the case of representing sets through grouping gestures and circular diagrams.

An important conclusion from the study is that representing sets through enclosed diagrams with or without accompanying grouping gesture helps preschoolers learn about cardinality and helps them to better judge numerical relation of two sets that differ by only one item. Our findings agree with Feigenson (2011) that *set representation* plays a critical role in representing the concept of cardinality.

Gestures benefit learning by adding a layer of meaning through action

Diagrams are static, but gestures are dynamic; they have an added channel to represent meaning through action (Tversky et al., 2013). In fact, there are many situations in which we rely on *both* diagrams and gestures to think, learn, and communicate. Collaborators often gesture over maps when they try to find routes (Heiser & Tvesrky, 2004). People gesture over diagrams when they try to assemble a piece of furniture (Tversky, Heiser, Lee, & Daniel 2009). Participants animate diagrams using gestures when explaining how a lock works (Engle, 1998).

Diagrams are effective in conveying structures, but they are not the best tools in representing action, change, and process (e. g., Heiser and Tversky, 2006; Tversky, Morrison, and Betrancourt, 2002). Gestures, however, are actions in space and can directly represent action and process. When participants see action gestures in addition to a diagram of the structure of complex engine along with verbal explanation, they gain a deeper understanding of the behavior (i.e. action and process) of the engine than participants who saw the same structure diagram and heard the same exact verbal explanation but with structure gesture (Kang, Tversky, Black, 2012). Thus, gestures provide a rich source of information about structure and action (e. g., Beattie &

Shovelton, 1999; Becvar, Hollan, & Hutchins, 2008), and due to their dynamic nature they can transmit information about behavior and process better than static diagrams.

That was exactly what we found in this study–– that gestures have more influence on actions than do diagrams. Results from the game sessions reveal that children who were asked to make a grouping gesture to make bubbles around each set never counted the two sets together to compare their numerical relation; on the other hand children who only saw bubbles around the sets (without the need to make the bubbles with grouping gestures) and children who never saw the diagram nor made a grouping gesture occasionally made the mistake of counting the two sets together. Counting two sets together as one set is a counting mistake that preschoolers make when they try to judge the numerical equivalence of the two sets (Jamalian, 2013). In the current study, we observed the same trend–– about one-third of our participant preschoolers counted the two sets together as one set when asked to compare the numerical relation of the two; some of them counted the two sets together at all times they engaged in counting; and some other counted the two sets together in one trial, and counted them separately (the correct behavior) in another. Performing the grouping gesture around sets helped them to completely eliminate this mistake while playing on the software activity. Counting is strongly linked to pointing actions, and grouping gesture helped to alter the counting action but the diagram with the same form did not have the same effect.

Individual Differences

Gestures can have the same form as diagrams, and hence, they may carry information that is redundant with diagrams. Such redundancy reinforces the message by presenting information in two modalities— a redundancy that may not be necessary for some, but beneficial to others. In the current study, we observed that, on average, preschoolers benefited the same from the intervention when they saw a circular enclosed diagram around the sets with or without a grouping gesture. That is children who saw the circular diagram with or without the grouping gesture improved the same in their overall math competence and their understanding of cardinality from pre to post-test and they improved in their accuracy of comparing numerical relation of two sets over the course of intervention. Further, both groups improved more than children who never saw the circular diagram nor were asked to make a grouping gesture. However, we realized that over the course of intervention, the grouping gesture helped children with lower executive functioning skills perform similarly to children with higher executive functioning skills.

Executive functioning skills have been defined as cognitive control functions involved in planning and organizing thought, remembering and attending to rules, focusing on one's behavior and inhibiting initial false impulses when performing a certain task (Diamond & Lee, 2011). Core to executive functioning skills are working memory, flexibility in thinking, and inhibition— involving self-control and self-regulation (Miyake, et al, 2000). In a previous study, researchers investigated the role of gesturing in young children's executive function by analyzing children's gestures during a sorting task of executive functioning skills (O'Neil & Miller, 2013). Results of the study show that preschoolers (ages 2.5 to 6 year old) tend to spontaneously gesture more when the task becomes more difficult and that children who gesture more perform better in the sorting task. Based on these results, they concluded that gestures might play a role in the development of executive functioning skills by helping working memory and/or by providing additional representation of information. In addition, it has been suggested that gestures lighten working memory load when explaining math solutions (Cook, Yip, & Goldin-Meadow, 2012; Goldin-Meadow & Beilock, 2010), and when solving problems that exceed working memory limit (Kessell & Tversky, 2006).

In the current study, we hypothesized that grouping gesture may help children to focus on items in only one set at a time and interrupts their counting routine by performing an action that has a form different than the repeated pointing gesture often used in the counting process; hence, grouping gesture may be beneficial for children with lower executive functioning skills. If that is the case, then performing grouping gesture in addition to seeing an enclosed diagram should have benefited children with lower executive functioning skills but for children with higher executive functioning skills such redundancy would not be necessary. We tested this hypothesis by analyzing children's accuracy in the game as a function of their executive function skills and the format of the game they played (gesture-plus-diagram, diagram only, no-gesture/no-diagram). Interestingly, we found that the slope between accuracy and executive functioning skills was not significantly different than zero for children in the gesture-plus-diagram condition; however, the slope was significantly positive for children in the diagram condition and in the no diagram/no gesture condition. In other words, in the diagram only and no-diagram/no-gesture conditions, children with higher executive functioning skills performed significantly better than children with lower functioning skills; whereas, in the gesture-plus-diagram condition children's performance was not a function of their executive functioning skills. Thus, the gesture-diagram redundancy benefited children with lower functioning skills but did not make a difference for children with higher executive functioning skills.

Sequential Representation of Diagrams

Diagrams on a piece of paper are static; diagrams shown in the "pop the bubble" activity were sequential, and not truly static. Figure 27 shows the sequence of which the circular enclosed diagrams appeared around the sets over the course of the gameplay:

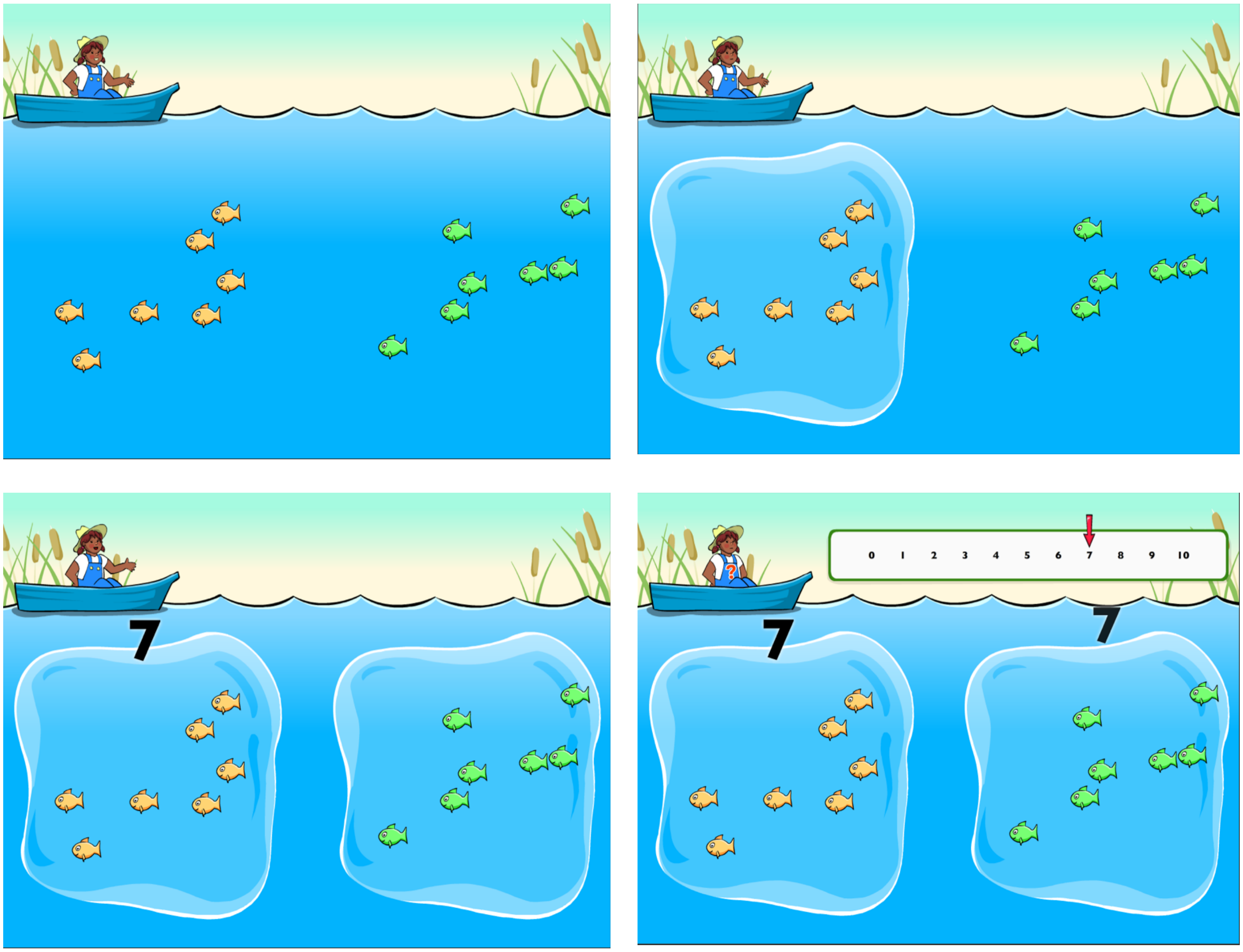

Figure 27: Pop the Bubble Math Activity Sequence

As it is evident from the figures, bubbles appeared sequentially around each set of fish one after the other. The child first saw two sets of fish, then a bubble appeared around the first set of fish, and the child was asked to indicate how many fish were in that bubble. Then another bubble appeared around the second set, the child was asked to indicate how many fish were in the new bubble, and finally the child was asked to judge whether there were the same number of fish in each bubble. Children in the diagram + gesture condition made each bubble with a grouping gesture; whereas, in the diagram only condition bubbles appeared automatically around the sets, but in the exact similar sequence. Such sequencing is not inherent to diagrams— on a piece of paper diagrams are static. In fact, the sequenced representation of bubbles around the sets made the two conditions (diagram-only and diagram + gesture) too similar to each other and perhaps such representation resulted in somewhat underestimating the effects of gesturing. Future studies are needed to compare truly static diagrams to the diagram + gesture condition (and to the sequential diagram condition) to further investigate the importance of the dynamic nature of gestures.

In textbooks, circular diagrams are usually presented around every set at the same time in cases where they are used to represent sets. Compared to the sequential representation, the all-show-at-the-time representation may not be as effective in highlighting sets. Obviously, it is almost impossible to represent diagrams sequentially on a piece of paper, as one can do in a computer activity. Thus, the conclusion that seeing circular diagrams around sets is as effective as making those diagrams with a grouping gesture may not be applicable to the diagrams in textbooks. We recommend that teachers and parents encourage children to draw circles around groups of things and indicate the number of items in the circle. Such intervention has been shown to improve children's understanding of cardinality and their overall math competence; showing two circles around two sets of items may not be as effective.

General Discussions on Diagrams and Gestures

Unlike most words, diagrams and gestures, specifically representational gestures, may have physical similarities (literal or metaphorical) to the things they refer to, for example, illustrating corner of a building by making an L shape gesture or highlighting a set by making a grouping gesture around it. Gestures and diagrams tend to convey information more directly than the symbolic words and they may be easier to interpret and easier to remember (Jamalian, Giardino, & Tversky, 2013; Tversky et al., 2013). Effective diagrams facilitate thinking and learning, in both those who make them and those who use them (Jamalian, 2013; Mayer, 2001; Tversky, 2011a). Effective gestures also assist thinking and learning, in both those who perform them and those who see them (Goldin-Meadow, 2003; McNeil, 1992).

When solving complex problems, people draw diagrams and the diagrams help them to find solutions (Kessell & Tversky, 2006). When they do not have access to paper and pencil, they simply gesture (Kessell & Tversky, 2006; Jamalian, et al., 2013). In one study, a majority of participants gestured when they were reading descriptions of new environments on their own alone in room (without an access to paper and pencil), and those who gestured were more accurate in making inferences about the environments (Jamalian et al., 2013). Gestures create “virtual diagrams” in the air, and with the same form, they may be the exact synonyms to diagrams; hence, one may replace the other. However, gestures have an added layer of action, and can use the action to better represent actions, behaviors, and processes. In the current study, we investigated whether asking children to make a grouping gesture in addition to seeing circular diagrams around sets helps them to improve their understanding of cardinality and their overall math competence. Grouping gesture and the circular diagram have the same form, so in some respect they are synonyms; but the grouping gesture has the additional layer of action and that layer of action may be more beneficial in representing sets and altering children’s counting routine when comparing sets. And that was exactly what we found: seeing the circular gesture with or without the grouping gesture was effective in helping children to learn cardinality and to improve their math competence. The circular diagram highlighted sets as a whole with certain properties and therefore, helping children to conceive of cardinal values as an emergent property of a set rather than property of the last counted item in the set. As mentioned previously, the above finding is only applicable to the cases where diagrams are presented sequentially around the sets and may not be applicable to static diagrams in a textbook.

We also found that performing the grouping gesture in addition to seeing the circular diagrams is beneficial to children with lower executive functioning skills, perhaps due to the added representational power through actions.

Finally, in a previous study and in the current study we observed that preschoolers often count two sets together as one set when they are comparing the numerical relation of the two sets. Interestingly, performing the grouping gesture around sets helped them to completely eliminate this mistake while engaging in the software activity. Counting is strongly linked to pointing actions, and grouping gesture helped to alter the counting action but the diagram with the same form did not have the same effect.

Implications for teachers and parents

Diagrams and gestures could be synonyms–- one could replace the other. But sometimes gestures could help more by “adding a layer of meaning through action” (Tversky et al 2013). We recommend teachers and parents to encourage children to gesture in forms that reflect the meaning of the concept to be learned, even in the presence of diagrams. In this study, we demonstrated that set-representation through enclosed diagrams and grouping gestures helps children to improve their understanding of cardinality and their overall math skills. Presenting sets in enclosed diagrams and/or encouraging children to draw circular diagrams around sets is a good practice to help preschoolers improve their early mathematical thinking.

Implication for designers

The results also have important implications for software designers: by designing gestural interactions that are congruent with the concept to be learned, designers can improve the learning effects of their software. These gestural interactions may emphasize a visual diagram in the activity and can directly effect the users' actions. In particular, results of this study demonstrate the positive effect of set-representation in a sequential manner through enclosed diagrams, with or without a grouping gesture. We encourage software designers to include such set representations in their activities when applicable.

Improvement in the Overall Math Competence

On average students who played MathemAntics' "Pop the Bubble" math activity improved in their TEMA scores at post-test comparing to pretest. Here, we highlight two principles of early mathematics instruction that were shared among all three conditions and discuss how each of these principles may have contributed to the preschoolers' improvement of math competency. It is important to highlight that these principles were not directly tested in the current study, and therefore, future studies are needed to specifically investigate possible effectiveness of these principles. Following, we discuss how each of these instruction principles may have benefitted children's overall math achievement based on findings from previous studies.

Principle #1: Focus on number comparison

Many scholars in early mathematics education have advised to focus instructions on numerical relations rather than a mere focus on vocabulary and rote memorization of mathematical elements (Cross, Woods, & Schweingruber, 2009). Some have even proposed that, "the age-old question, "Who has more?" is foundational to all of mathematics" and have suggested that "human fixation on comparison was a fundamental impetus for [the] invention of numbers" (Kepner, 2010).

Some recent experimental studies talk to the same point. In one study, Hyde, Khanum, and Spelke (2013) showed that children who briefly practiced tasks that engaged primitive approximate numerical quantities performed better on subsequent exact, symbolic arithmetic problems than did children given other tasks involving comparison and manipulation of non-numerical magnitudes (brightness and length). Children were presented with two arrays of dots, one after the other, and were asked whether the second array was more or less numerous than the first. In another condition, they were asked to estimate the numerical sum of two sequentially presented array of dots and judge whether a third array was more or less than the actual sum. Children in both of these conditions outperformed children in a third condition that simply compared brightness of two objects. Therefore, the mere practice of numerical comparison can help children to improve their symbolic mathematical thinking; a practice that all children in the math comparison groups in the current study had.

The ‘‘RightStart’’ curriculum designed by Griffin et al. (1994; now known as ‘‘NumberWorlds,’’ Griffin, 2004) also included number comparison, and put an emphasis on the relationship between counting and quantity to compensate for a lack of environmental experience in early mathematical skills for the children from low SES families. Following the same principal, ‘‘The Number Race’’ game designed by Wilson & Dehaene (2007) included number comparison as an essential part of the game. Both ‘‘RightStart’’ and ‘‘The Number Race’’ has been shown to be effective in improving children’s early mathematical understanding.

Comparing the numerical relation of two sets of items involves many fundamental early mathematics thinking processes: engaging primitive approximate numerical quantities, attending to sets of items (as opposed to merely focusing on individual items that make up the set), relationship between counting and quantity, concept of exact quantity, and order of numbers—all of which important in early math competence.

Principle #2: Connecting different representations of numbers

It is important for children to understand the meaning of mathematical symbols, connect them to non-symbolic mathematical ideas, and more generally to appreciate that mathematics is meaningful (Ginsburg, Jamalian, & Creighan, 2013). Vygotsky (1978) proposed that early mathematics education should help children learn to synthesize their "spontaneous concepts" of everyday mathematics with "scientific concepts" of the symbolic formal mathematics. "The strength of scientific concepts lies in their conscious and deliberate character. Spontaneous concepts, on the contrary, are strong in what concerns the situational, empirical, and practical" (Vygotsky 1986, p. 194). In other words, early interventions in mathematics should help children to mathematize their meaningful everyday mathematics (Ginsburg, Lee, Boyd; 2008).

Recent studies also emphasize the importance of connecting symbolic and non-symbolic representation of numbers (Ginsburg et al, 2013; Mazzocco, 2005; Wilson & Dehaene, 2007). It has been even suggested that dyscalculia, or mathematical learning disability presumed to be related to impaired brain function, may involve deficiency in the ability to connect non-symbolic magnitudes to symbolic representation of numbers (Mazzocco, 2005; Wilson & Dehaene, 2007). As noted in chapter 1, much evidence suggests that children start reciting number words without an understanding of the actual meaning of these words (Fuson, 1988; Spelke, 2011; Wynn 1990; 1992) and only after they learn cardinality can they link verbal and symbolic number words to non-symbolic representation of magnitudes (Le Corre, 2013; Le Corre & Carey, 2007). Yet, whereas the connection between symbolic and non-symbolic representation of numbers is automatic in adults (Dehaene & Akhavein, 1995; Naccache & Dehaene, 2001; Pavese & Umilta, 1998; Rusconi, Priftis, Rusconi, & Umilta, 2006), the process is not spontaneous for young children (Girelli, Lucangeli, & Butterworth, 2000; Moyer & Landauer, 1967; Rousselle & Noel, 2007). These findings lead to an important instructional principle in early mathematical learning, which emphasizes the importance of connecting the links between various representations of numbers; as opposed to practices that are merely focused of rote memorization of formal mathematics.

Two important features in the pop of the bubble activity are designed with this instructional principle in mind: (1) asking children to respond to the "how many" question by choosing a number from a linear array of numbers, and (2) receiving audio and visual feedback highlighting individual items while being counted, repeating last counted number word, and showing the number word representing cardinality on top of the set. Next we discuss these two features in more detail.

Design feature 1: linear number representation as a response mechanism to the "how many" question

Playing "Pop the Bubble" activity, children were asked to indicate how many fish are in each set by tapping on a number from an ordered list of numbers on a line, albeit not connected with a line. Previous studies have shown that linear representation of numbers help preschoolers to improve their performance on numerical comparison, counting, numerical identification, and numberline estimation tasks (Wilson, Dehaene, Pinel, Revkin, Cohen, & Cohen, 2006; Griffin, 2004; Siegler & Ramani, 2008; 2009). In one study, Siegler and Ramani (2008) had preschoolers play on either a linear number board game with numbers shown on 10 equal-sized squares of different colors arranged in a horizontal array increasing in magnitude from left to right, or on a circular board with numbers increasing in magnitude in a clock-wise or counter clock-wise direction. Children who played on the linear number line improved more significantly on number comparison and numberline estimation tasks than children who played on the circular board game. Representing numbers on a horizontal line makes an association between symbols (Arabic number digits) and non-symbolic linear space which in turn may help children to improve their "number sense access"— defined as the spontaneous conversion of symbols to non-symbolic quantities (Rousselle & Noel, 2007; Wilson, Dehaene, Dubois, & Fayol, 2009). Previous research suggests that adults and children access the non-symbolic representation of quantities when solving symbolic math problems (Gilmore, McCarthy, & Spelke, 2007). Some evidence suggests that the conversion between symbolic and non-symbolic representation of numbers has become automatic for adults (Dehaene & Akhavein, 1995; Naccache & Dehaene, 2001; Pavese & Umilta, 1998; Rusconi, Priftis, Rusconi, & Umilta, 2006), but yet to become automatic for young children (Girelli, Lucangeli, & Butterworth, 2000; Moyer & Landauer, 1967; Rousselle & Noel, 2007). Thus, exposing children to games that represent numbers on a line may help children improve their "number sense access".

Another explanation for the positive effect of linear number representation may be the literal ordering of numbers one after the other, from the smallest to the biggest. The visual representation of numbers in a linear order provides an additional channel of information for children when learning order of numbers in their count list— not only they say number 6 after number 5 and before number 7, but also they *see* the symbol for number 5 before 6 and after 4. Much evidence has shown that visualization of information has a positive effect on comprehension (Tversky, 2011b) and learning (Mayer, 2001).

<u>Design feature 2: Counting, highlighting, and repeating last counted number as feedback</u>

Children were given two chances to indicate how many fish were in each set. In case they made a mistake on their second trial an auditory and visual feedback was given to them, modeling the counting procedure and highlighting the cardinality of the set. For example, if there were 6 fish in a set, an audio was played saying, “there are 1, 2, 3, 4, 5, 6; 6 fish all together”. Along with the audio, fish were highlighted one by one while being counted, and then highlights would turn off and on again on all the fish at the same time, while the audio repeated the cardinality phrase, “x fish all together”, x being the cardinality of the set. In addition, the number representing the magnitude of the set would appear on top of each set after counting them (See Figure 28).

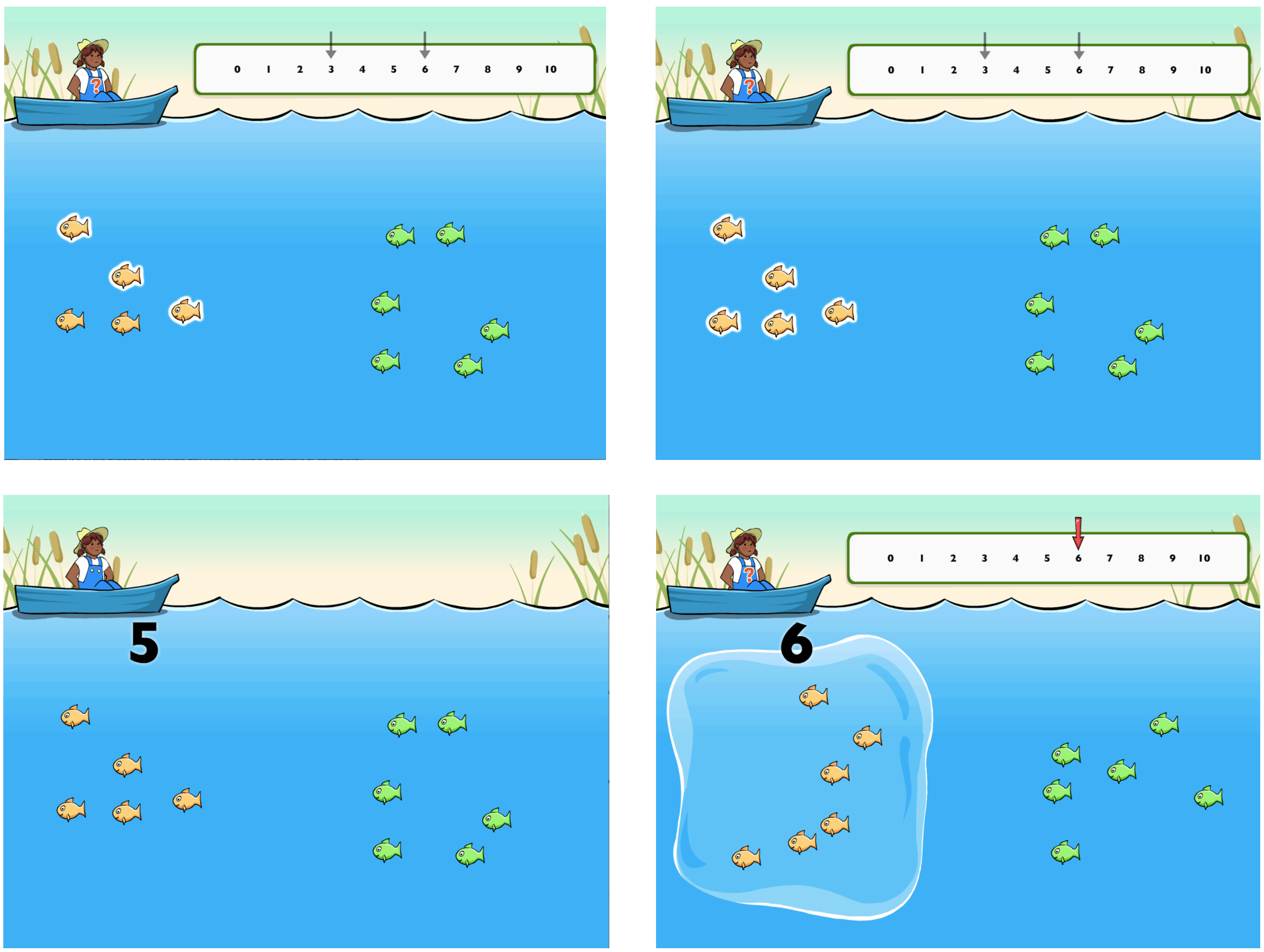

Figure 28: Counting and Cardinality Feedback in the Pop the Bubble Math Activity

Highlighting individual fish while being counted helps children to keep track of fish that were counted and separate them from the ones to be counted. In addition, highlighting the fish while being counted would help to coordinate saying the number words with tracking the individual items and emphasizing the one-to-one correspondence principle- an important counting principle noted by Gallistel and Gelman (1992). Therefore, the highlighting feature may have helped children to improve their counting skills. Of course, since we did not control for the presence of this feature, we cannot come to the conclusion that this feature was definitely effective, but we can speculate such conclusion based on prior research.

Turning off the individual highlights and having the whole set to be highlighted at once when repeating the cardinality number, emphasizes the set an entity and makes a connection between the cardinality magnitude and the set as a whole, instead of the last fish that was counted. Therefore, in a way all students who played on the software were presented with some form of set representation, which may have helped them in understanding the concept of cardinality. However, children who saw the bubble diagram around the set benefited more from the intervention perhaps because the bubbles were a good representation of an enclosed container and a better depiction of the abstract concept of sets. Finally, repeating the last counting word along with the phrase "all together" has been suggested to be an effective method to help children learn the cardinality principle (Cross et al., 2009).

Therefore, the audio and visual feedback presented to children in all conditions when they made a mistake in answering the "how many" question was probably helpful in improving their overall math competency. However, given the fact that children who saw the diagram of bubbles around set improved even more than children who did not see the diagram, it is fair to conclude that showing an enclosing diagram around sets is a very effective method to help preschoolers learn about cardinality and improve their overall math competency.

In summary, bubbles and grouping gestures were not the only valuable features of the “pop the bubble” activity; the mere practice of number comparison, and connecting different representations of numbers through well designed response and feedback mechanisms, presented in all three conditions, were helpful in improving children’s overall math competence.

Implication for teachers and parents

Early math education goes beyond memorization of counting words in a fixed order and using counting skills to answer the “how many” question. We recommend that parents and teachers engage children in activities that require them to compare numerical relation of sets, and encourage them to use their counting skills when solving math problems, like identifying the set with more items. We also highly recommend parents and teachers to connect different representation of numbers by asking children, for example, to represent quantities with number symbols and/or to indicate their answer to the “how many question” on a number line.

Implication for designers

When designing software activities for early mathematics, it is important to focus on children’s thinking processes and the substance of the mathematics to be learned rather than on mere presentation of mathematical symbols. Designers may rely on the instruction principles discussed above as guiding principles when designing early math activities, and may incorporate the two design features in their own designs when applicable.

Executive Functioning Skills and Early Mathematics

Executive functioning skills, measured by performance on the Head-Toes-Knees-Shoulders (HTKS) task, was positively correlated with preschooler’s overall mathematics competence, measured by Standard Test of Early Mathematics Ability (TEMA). Executive functioning skills was also positively correlated with children’s accuracy in comparing numerical relation of two sets, and their performance on the “Pop the Bubble” software math activity. In addition, when comparing the numerical equivalence of two sets, children with higher executive functioning skills were more likely to count the two sets separately to make their judgment; this shows a positive relation between executive functioning skills and children’s successful use of counting to solve relational problems.

Core to executive functioning skills are working memory, flexibility in thinking, and inhibition— involving self-control and self-regulation (Miyake, et al, 2000). Previously, it has been shown that working memory predicts school age children’s (between the ages 7 and 14 years) performance in mathematics (Gathercole, Pickering, Knight, & Stegmann, 2004). Epsy et al (2004) have shown that inhibitory skills are strong predictors of children’s early arithmetic and overall math skills. Further, in a longitudinal study, Bull, Espy, and Wiebe (2008) found that children with higher executive functioning skills have an immediate advantage in their math and reading performance, and that they maintain this advantage throughout the first three years of primary school.

Ponitz et al (2009) developed the Head-Toes-Knees-Shoulders (HTKS) task as a direct assessment of children’s behavioral regulation-- as an indicator of executive function skills in children’s gross motor actions. HTKS task is easy and quick to administer and is suitable for evaluating preschool and kindergarten children’s behavioral regulation. Previously, it has been shown that gains in behavioral regulation predict gains in mathematics but not in language and literacy over the kindergarten year (Ponitz et al, 2009). Current study provides further evidence for the importance of behavioral regulations in overall early mathematics competence and also specifically in accuracy of numerical equivalence judgment tasks, and in children’s correct use of counting strategy when comparing numerical relation of two sets.

Counting Sets Together- Task related or a Matter of Maturity?

An intriguing finding of the study was the insight into the maturity process of preschoolers' counting strategy when they compare two sets of items. In the current study, and also in a previous one (Jamalian, 2013), we observed that preschoolers sometimes count both sets together when they are asked to judge the numerical relation of two sets. For instance, to compare a set of 4 candies with a set of 5 candies, children sometimes counted both sets together as 1, 2, 3, 4, 5, 6, 7, 8, 9, and then provided their judgment about the numerical relation between the two sets. Interestingly, a majority of children who counted the two sets together made this mistake in some trials but not in others. These children counted two sets together in some trials, but counted each set separately or counted only one of the sets or none of the sets in other trails. Some children counted two sets together in one trial and corrected themselves by recounting the sets separately in the same trial. Only a small percentage of children counted the two sets together in all trials that they engaged in counting.

Counting two sets together was not related to problem representation. In our previous study, both sets of items were presented to children on one piece of paper, and we speculated that perhaps children would not make such mistake if the two sets were presented on two separate pieces of papers. However, in the current study, we presented the two sets of items on two separate pieces of paper and still we got the same result: quite often children counted the two sets together even when they were presented separately. We also observed the same trend in the "Make the same" task where children were dealing with real blocks. Therefore, it seems that counting-two-set-together mistake is not related to the specific task but it is perhaps part of the maturation of their counting skills.

We hypothesized that such error may be due to children's limited understanding of the link between counting and evaluating cardinality of sets, and/or it may be due to their limited working memory and inhibitory skills. In fact, children who were not cardinality knowers at the pre-test did not engage in counting frequently to compare numerical relation of two sets at pre-test, despite prompting them to count by encouraging them to "check" and then answer or "to show" how they know their answer. At post-test they engaged in counting more frequently, but they often counted the two sets together as one set. Children who knew cardinality at pretest, engaged in counting more often, and they counted the two sets separately (the correct strategy) significantly more frequent at post-test than pretest. Thus, as children improved their counting skills, they tended to count the two sets separately more often. From the above observations, we can conclude that when children who do not know cardinality engage in counting less often when they try to solve math problems, and when they start counting, they have difficulty in using their counting skills effectively, and therefore may count sets together when it is not applicable. Only after they are advanced in their counting skills, they can engage in counting effectively to compare numerical relation of two sets.

As mentioned above, we hypothesized that the counting-two-sets-together error may be due to children's limited understanding of the link between counting and evaluating cardinality of sets, *and/or* to their limited working memory and inhibitory skills. Children need to be able to count one set, stop, and start from number 1 to count the other set. Furthermore, children's failure to stop counting and start from 1 for the second set may also be due to their limited cognitive flexibility: they may find it difficult to think about the concept of sets while counting at the same time. In other words, they may be aware that counting is a good strategy to compare sets, but find it difficult to use their counting skills purposefully. Based on these lines of reasoning, we hypothesized that children with higher executive functioning skills may be more likely to count the two sets separately than children with lower executive functioning skills. In the current study, we found that, in fact, when asked to compare numerical relation of two sets children with higher executive functioning skills were more likely to count the two sets separately than children with lower executive functioning skills.

When presented with two sets of items, one can compare them, or perhaps add them together. To compare the two sets, it is feasible to count each set separately and then compare their cardinality values; to add them together, the feasible strategy is to count them together and evaluate their sum. In other words, whereas counting two sets together is the correct strategy when adding things together, it is not a logical strategy when comparing sets. It may be that children who are not proficient in counting are not flexible enough in using their counting skills effectively based on the context. As they improve their numerical knowledge (become cardinality knowers) and as they improve their executive functioning skills, they become better in using their counting skills effectively to solve various types of math problems.

Limitations

There are certain limitations to this study that must be addressed. First limitation is related to the usability difficulties some children had in the gesture + diagram condition. Occasionally, children in this condition had difficulty in performing the grouping gesture on iPad in order to make bubbles around the sets; in these cases, the researcher would help them to accomplish the task by holding on their finger and directing their finger movement on the screen. It is important to note that children did not have to draw perfect circles around the sets for the bubbles to appear, but they had to draw an enclosing diagram; yet, some would take their fingers off the screen in the middle of their drawing and the software would require them to restart. One way of eliminating this issue for future studies is to allow children to take their fingers off the screen and have the opportunity to continue their drawings from where they have left it off within a short pause.

The other limitation of the study was the setting in which the pre- and post-tests and the intervention sessions were conducted. In a typical setting, three to four researchers worked with children on a one-on-one arrangement, all in one hallway or in one classroom. In such setting, children could easily hear their classmates counting or making comments about different elements in the game. A few children in the no-gesture/no-diagram condition, for example, asked about the bubbles and wondered why they do not see any bubbles in their game. In a few occasions, some children in the diagram-only condition pretended that they are drawing bubbles by chasing their finger around the boundary of the bubbles. Hearing other children counting may also have altered other children's counting behavior. Despite our efforts to make the conditions as separate as possible, these issues were inevitable as they may be present in any study done in public preschool settings.

Another important limitation of the study is the relatively small sample size. Having twenty-three to twenty-four children in each condition does not allow making strong claims about individual differences and effects of gesturing, per se. The finding of this study suggests that gesturing might benefit children with lower executive functioning skills, at least in the context of set-representation for learning about cardinality. The relation between gesturing and executive functioning skills should be investigated further in future studies with larger sample sizes.

Finally, there were few limitations to the design of the pre- and post-test tasks. In Give the Same (GS) task, children were shown a card with a picture of a boy or a girl having few “candies” (printed on the card) and were asked to give the boy or the girl on a separate card the same number of candies (i.e. blocks). To solve the problem, children engaged in variety of strategies: some placed blocks on the empty card in the exact visual arrangement of the printed candies, which resulted in a correct response. However, it is not clear whether these children understood the meaning of “same number” or they simply relied on the visual information to make the two sets *look* similar. Some placed blocks on top of the printed “candies” and then moved all the blocks over to the empty card. These children, therefore, relied on one-to-one correspondence to make an equal set with the set they were shown. Thus, children could solve the task accurately without the need to rely on their counting skills; and perhaps that was the reason we did not find any group differences in children’s performance on this task. To specifically study children’s using counting skills in solving equivalence problems, the task should have been designed in a way to make counting the only feasible strategy for coming up with a correct response. For example, the researcher could have hidden the card with the printed candies and then ask the child to place the same number of candies on the empty card. In that case, it would have been harder for the child to rely on the visual information or to directly match one set with the other; instead, to succeed in the task, child would have had to evaluate cardinality of the first set and place the same number of blocks on the other set that would match the cardinality of the first set.

Similarly, in the Make the Same (MS) task, children engaged in variety of strategies to make the two sets equal. Some relied on visual matching, arranging blocks in the same visual pattern in the two sets. For example, they placed all blocks in a row, in a column, or based on a dice pattern; some would take away all blocks from both sets and then add the same number to each side; and some would count the two sets and added or removed blocks from the sets. To specifically study children's proficiency in using their counting skills when producing equal sets, the task should have been designed in a way to make counting the only feasible strategy to solve the problem. For example, blocks in one set could have been arranged in a circular tray, and in the other set, they could have been arranged in a rectangular tray. In that case, it would have been harder for children to rely on the visual information or to directly match one set with the other to make them equal.

Future Directions

This study demonstrated the positive effects of set-representation through enclosed diagrams and grouping gestures on children's understanding of cardinality and on their improvement of overall math skills. The study specifically compared effects of grouping gesture accompanying circular diagram with the effects of diagram alone, and results showed that the enclosed diagram with or without the grouping gesture is effective in helping preschoolers to learn about cardinality and to improve their math proficiency. However, it is not clear whether the grouping gesture alone would have similar effects as the enclosed diagram alone. If similar forms of diagrams and gestures are synonyms to each other, then perhaps the grouping gesture alone will be as effective as the enclosed diagram to represent sets. Future studies are needed to compare directly the grouping gesture with or without the enclosed diagram to further investigate the role of gesturing in early mathematical thinking.

Further, as discussed previously, over the course of the intervention, diagrams appeared around the sets in a sequential manner, one after the other. It is not clear whether the visual representation of the diagram, per se, was effective or whether the sequential representation of diagrams was particularly important in representing sets. Diagrams in textbooks and on paper are static, and hence, may essentially not have the same positive effects as of the sequential diagrams used in this study. Future studies are needed to compare effects of sequential representation of diagrams in software with that of the static diagrams in textbooks.

This dissertation also alluded to the possible role of executive functioning skills in children's counting behavior. We observed that children engage in variety of counting behaviors when they compare numerical relation of two sets: they may count the two sets together as one whole set (incorrect strategy), they may count the two sets separately (correct strategy), or may count only one of the sets or none of them at all. Results showed that children with higher executive functioning skills tended to engage in the correct counting strategy (i.e. they counted the two sets separately) more often than children with lower executive functioning skills to judge numerical equivalence of two sets. Future studies with larger sample sizes are needed to further investigate the link between executive functioning skills and use of counting skills to solve different types of mathematical problems.

Finally, results of the study suggest a link between executive functioning skills and gesturing. Finding of the study imply that children with lower executive functioning skills benefit from performing the grouping gesture in addition to seeing the enclosed diagrams around the sets, whereas, children with higher executive functioning skills do not need such redundancy. Future studies with larger sample sizes should further investigate the possible benefit of gesturing for children with lower executive functioning skills.

REFERENCES

Alibali, M.W. & DiRusso, A.A. (1999). The function of gesture in learning to count: more than keeping track, *Cognitive Development*, *14*(1), 37- 56.

Bates, E. (1976). *Language and context.* New York: Academic Press.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, *123,* 1-30.

Becvar, A., Hollan, J., & Hutchins, E. (2008). Representational gestures as cognitive artifacts for developing theories in a scientific laboratory. In M. S. Ackerman, C. A. Halverson, T. Erickson, & W. A. Kellogg (Eds.) *Resources, Co-Evolution and Artifacts: Theory in CSCW* (pp. 117-143). London, England: Springer-Verlag.

Bilda, Z and Gero, J.S. (2006). Reasoning with internal and external representations: A case study with expert architects. In R. Sun (ed), *Proceedings of the Annual Meeting of Cognitive Science Society* (pp. 1020- 1026). Mahaw, NJ: Lawrence Erlbaum Associates.

Bull, R., Epsy, K.A., & Wiebe, S.A., (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Developmental neuropsychology*, *33*(3), 205-28.

Carey, S. (2004). Bootstrapping and the origins of concepts. *Daedalus*, 59-68.

Carey, S., & Xu, F. (2001). Beyond object-files and object tracking: infant representations of objects. *Cognition, 80*, 179–213.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33* (4), 747-756.

Cross, C.T., Woods, T.A., & Schweingruber, H. (Eds.). (2009). *Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity*. Washington, DC: National Academy Press.

Cook, S.M., & Goldin-Meadow, S. (2006) The role of gesture in learning: Do children use their hands to change their minds? *Journal of Cognition and Development*, *7*, 211 - 232.

Cook, S. W., Yip, T. & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*, 594-610.

Dehaene, S. (1997). *The number sense*. Oxford University Press, Penguin press, New York, Cambridge (UK).

Dehaene, S. & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 314--326.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*, 959–964.

Ehrlich, S. B., Levine, S., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42*, 1259 - 1268.

Enfield, N. J. (2003). Producing and editing diagrams using co-speech gesture: Spatializing non-spatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology*, *13*(1), 7-50.

Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Espy KA, McDiarmid MM, Cwik MF, Stalets MM, Hamby A, Senn TE (2004). The contribution of executive functions to emergent mathematical skills in preschool children. *Developmental Neuropsychology*, *26*, 465– 486.

Feigenson, L. (2011). Objects, sets, and ensembles. In S. Dehaene & E. Brannon (Eds.), *Attention and Performance, Volume XIV*, Oxford University Press.

Feigenson, L., & Carey, S. (2003). Tracking individuals via object files: evidence from infants' manual search. *Developmental Science, 6*(5), 568–584.

Feigenson, L., & Carey, S. (2005). On the limits of infants' quantification of small object arrays. *Cognition, 97*(1), 13–23.

Feigenson, L., Carey, S., & Hauser, M., (2002). The representations underlying infants' choice of more: object files versus analog magnitudes, *Psychological Science, 13* (2), 150–156.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8* (7), 307-314.

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

Fuson, K., & Hall, J. (1983). The acquisition of early number word meanings: A conceptual analysis and review. In H. Ginsburg (Ed.), *The development of mathematical thinking* (pp.

49-107). New York: Academic Press.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43–74.

Gallistel, C. R. & Gelman, R. (2005). Mathematical Cognition. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning*. Cambridge university Press, Cambridge, New York.

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from National Curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18*, 1–16.

Gelman, R. (1993). A rational-constructivist account or early learning about numbers and objects. Advances in research theory. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation* (pp. 61-96). San Diego: Academic Press.

Gelman, R., and Gallistel, C.R. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.

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

Gelman, R., and Meck, E. (1992). Early principles aid initial but not later conceptions of number. In J. Bideaud, C. Meljac and J. P. Fischer (Eds.). *Pathways to number* (pp. 171- 189). Hillsdale, NJ.: Erlbaum Associates.

Ginsburg, H.S. (1977). *Children's Arithmetic*. New York: Van Nostrand.

Ginsburg, H. P. (1989). *Children's arithmetic* (2nd ed.). Austin: Pro-Ed.

Ginsburg, H. P. (1997). *Entering the child's mind: the clinical interview in psychological research and practice*. New York: Cambridge University Press.

Ginsburg, H., & Baroody, A. (2003). *Test of Early Mathematics Ability--Third Edition*. Austin, TX: Pro-Ed.

Ginsburg, H.P., Carpenter, K., Labrecque, R., & Pagar, D. (2013). Guidelines for Developing a System of High-Quality Software to Promote Young Children's Meaningful Mathematics Learning. In R. Cohen Kadosh & A. Dowker, (Eds.), *Oxford Handbook of Numerical Cognition.*

Ginsburg, H. P., Jamalian, A., & Creighan, S. (2013). Cognitive Guidelines for the Design and Evaluation of Early Mathematics Software: The Example of MathemAntics. In L.D. English & J.T. Mulligan (Eds), *Reconceptualizing Early Mathematics Learning* (pp. 83-120). Springer Netherlands.

Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: what it is and how to promote it. *Social Policy Report*, *22* (3),

Ginsburg, H. P., & Opper, S. (1987). *Piaget's Theory of Intellectual Development* (3^{rd} ed.). Englewood Cliffs, NJ: Prentice Hall.

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticy of accessing number magnitude. *Journal of Experimental Child Psychology*, *76*, 104–122.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*, Cambridge, Mass.: The Belknap Press.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1-19.

Goldin-Meadow, S., & Alibali, M.W. (2013). Gesture's role in learning and development. *Annual Review of Psychology*, *64.* 257-283.

Goldin-Meadow, S. & Beilock, S. L. (2010). Action's influence on thought: The case of gesture. *Persepctives on Psychological Science*, 5(6), 664-674.

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teachers' hands tell the students' minds about math. *Journal of Educational Psychology, 91*, 720–30.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*, 516-522.

Goldin-Meadow, S., & Sandhofer, C. M. (1999). Gesture conveys substantive information to ordinary listeners. *Developmental Science, 2* , 67–74.

Gopnik, A., Sobel, D. M., Schulz, L. E., & Glymour, C. (2001). Causal learning mechanisms in very young children: Two-, three-, and four-year-olds infer causal relations from patterns of variation and covariation. *Developmental Psychology, 37*, 620–629.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: using children's hands to read their minds. *Cognition and Instruction, 9*, 201–19.

Greeno, J. G., Riley, M. S., and Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology, 16*, 94-143.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447*(7144), 589–591.

Griffin, S. A. (2004). Building number sense with NumberWorlds: A mathematics program for young children. *Early Childhood Research Quarterly*, *19*(1), 173.

Griffin, S. A., 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: MIT Press.

Hauser, M.D. & Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: Examinations of content and format. *Cognitive Psychology*, *47*, 367-401.

Hauser, M. D., & Spelke, E. S. (2004). Evolutionary and developmental foundations of human knowledge: A case study of mathematics. In M. Gazzaniga (Ed.), *The Cognitive Neurosciences, Vol. 3.* Cambridge: MIT Press.

Heiser, J., & Tversky, B. (2004). Characterizing diagrams produced by individuals and dyads. In T. Barkowsky (Editor). *Spatial cognition: Reasoning, action, interaction* (pp 214-223). Berlin: Springer-Verlag.

Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science, 30*, 581-592.

Hostetter, Autumn B. & Martha W. Alibali (2004). On the tip of the mind: Gesture as a key to conceptualization. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 589–594). Mawah, NJ: Erlbaum.

Iverson, J.M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*, 367-371.

Izard, V., Sann, C., Spelke, E.S., & Streri, A. (2009). Newborn infants perceive abstract numbers, *Proceeding of National Academy of Sciences U.S.A. 106* (25), 10382–10385.

Izard, V. & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, 106(3), 1221-47.

Jamalian, A. (2013), Metaphoric gestures and mathematical education for children from low-income families, *In Proceedings of American Educational Research Association (AERA)*.

Jamalian, A., Giardino, V., and Tversky, B. (2013). Gestures for Thinking. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society.*

Jamalian, A. and Tversky, B. (2012). Gestures alter thinking about time. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, Pp. 551-557. Austin TX: Cognitive Science Society.

Kang, S. Tversky, B., & Black, J. B. (2012). From hands to minds: How gestures promote action understanding. . In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 551-557). Austin TX: Cognitive Science Society.

Kepner, H. (2010, Februrary). A Missed Opportunity: Mathematics in Early Childhood. Retrieved from http://www.nctm.org/about/content.aspx?id=25054).

Kessell, A. M. & Tversky, B. (2006). Using gestures and diagrams to think and talk about insight problems. In R. Sun and N. Miyake (Eds) *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2528).

King, A. S., McDonald, R., & Gallistel, C. R. (2001) Screening for mice that remember incorrectly, in *International Journal of Comparative Psychology*, *14*, 232-257.

Krauss R. M. & Hadar U. (2001). The role of speech-related arm/hand gestures in word retrieval. In R. Campbell & L. Messing, (Eds.) *Gesture, speech, and sign* (pp. 93-116). Oxford: Oxford University Press.

Lakoff, G. & Núñez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books, 2000.

Le Corre, M. (2013). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, 1- 15.

Le Corre, M., Brannon, E.M., Van de Walle, G., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. *Cognitive Psychology, 52(3)*, pp. 130-169.

Le Corre, M. & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105,* 395-438.

Lipton, J., & Spelke, E. S. (2003). Origins of number sense: Large number discrimination in human infants. *Psychological Science, 14*, 396-401.

Lipton, J., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy, 5*, 271 – 290.

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*(5), 978–988.

Mayer, R.E. (2001). *Multimedia learning*, Cambridge, England: Cambridge University Press.

Mazzocco M.M. (2005). Challenges in Identifying Target Skills for Math Disability Screening and Intervention. *Journal of Learning Disabilities, 38*, 318-323.

McCrink, K. & Wynn, K. (2004). Large-Number Addition and Subtraction by 9-Month-Old Infants. *Psychological Science*, *15* (11), 776-781.

McNeill, D. (1992). *Hand and mind: What gesture reveals about thought*. Chicago: University of Chicago Press.

Michie, S. (1984). Why preschooler's are reluctant to count spontaneously. *British Journal of Developmental Psychology, 2(4)*, 347-358.

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.

Mix, K. S. (1999). Similarity and numerical equivalence: appearances count, *Cognitive Development, 14*, 269-297.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.

Moyer, R. S., & Landauer, T. K. (1973). Determinants of reaction time for digit inequality judgments. *Bulletin of the Psychonomic Society*, *1*, 167-168.

Naccache, L., & Dehaene, S. (2001). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex*, *11*, 966–974.

O'Neil, G., & Miller, P. H. (2013). A Show of Hands: Relations Between Young Children's Gesturing and Executive Function. *Developmental Psychology, 49*(8), 1517-28.

Pavese, A., & Umilta, C. (1998). Symbolic distance between numerosity and identity modulates Stroop interference. *Journal of Experimental Psychology: Human Perception & Performance, 24*, 1535–1545.

Perry, M., Berch, D., & Singleton, J. L. (1995). Constructing shared understanding: The role of nonverbal input in learning contexts. *Journal of Contemporary Legal Issues, 6,* 213–236.

Ping, R. & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology*, *44*(5), p. 1277.

Piaget, J. (1952). *The child's conception of number.* New York: Norton.

Piaget, J., and Inhelder, B. (1962). *The Psychology of the Child*. New York: Basic Books.

Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology, 45*, 605-619.

Rowe, M.L., Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*, 951–53.

Rousselle, L., & Noel, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition*, *102*, 361–395.

Rowe, M.L., Özçalişkan, S., Goldin-Meadow S. (2008). Learning words by hand: gesture's role in predicting vocabulary development. *First Lang*, *28*(2), 185–203.

Rusconi, E., Priftis, K., Rusconi, M. L., & Umilta, C. (2006). Arithmetic priming from neglected numbers. *Cognitive Neuropsychology*, *23*, 227.

Sarnecka, B. W. (2007). From grammatical number to exact numbers: early meanings of 'one,' 'two,' and 'three' in English, Russian, and Japanese. *Cognitive Psychology, 55*(2), 136–168.

Sarnecka, B.W. & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition, 108*(3), 662-674.

Saxe, G. B., & Kaplan, R. G. (1981). Gesture in early counting: A developmental analysis. *Perceptual and Motor Skills, 53*, 85l- 854.

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology, 6,* 357-379.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250-264.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11, 655–661.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games -- but not circular ones -- improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, 101, 545-560.

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teachers gestures and speech differ. *Psychological Science, 16*, 85-89.

Sophian, C. (1987). Early developments in children's use of counting to solve quantitative problems. *Cognition and Instruction, 4*(2)*,* 61 - 90.

Sophian, C. (1995). Representation and reasoning in early numerical development: Counting, conservation, and comparisons between sets. *Child Development, 66*, 559 – 577.

Spelke, E. S. (2011). Natural number and natural geometry. In E. Brannon & S. Dehaene (Eds.), *Attention and Performance Vol. 24. Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought* (pp. 287-317). Oxford, United Kingdom: Oxford University Press.

Steffensen, M. S. (1978). Satisfying inquisitive adults: Some simple methods of answering yes-no questions. *Journal of Child Language*, *5*, 221–236.

Suriyakham, L. W. (2007). *Input effects on the development of the cardinality principle: does gesture count?* (Doctoral dissertation). University of Chicago.

Talmy L., (1983). How language structures space. In Herbert L. Pick, Jr. & Linda P. Acredolo (eds.), *Spatial orientation: Theory, research, and application* (pp. 225-282). New York: Plenum Press.

Tversky, B., Heiser, J., Lee, P. and Daniel, M.-P. (2009). Explanations in gesture, diagram, and word.. In K. R. Coventry, T. Tenbrink, & J. A. Bateman (Eds), *Spatial Language and dialogue* (pp. 119- 131). Oxford: Oxford University Press.

Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology, 23*, 515-557.

Tversky, B. (2011a), Visualizing Thought. *Topics in Cognitive Science, 3*, 499–535.

Tversky, B. (2011b). Tools for thought. In B. Benedetti and V. Cook (Eds.), *Language and bilingual cognition* (pp. 131-139). New York: Psychology Press.

Tversky, B., Jamalian, A.**,** Giardino, V., Kang, S., & Kessell, A. (2013). Gestures as Diagrams: Towards a Semantics for Gesture. In *Proceedings of the Tilburg Gesture Research Meeting (TiGeR 2013).*

Tversky, B., Morrison, J. B. & Betrancourt, M (2002). Animation: Can it facilitate? *International Journal of Human Computer Studies, 57*, 247-262.

Valenzeno, L., Alibali, M. A., & Klatzky, R. (2003). Teachers' gestures facilitate students' learning: A lesson in symmetry. *Contemporary Educational Psychology, 28,* 187–204.

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language.* Cambridge: MIT Press. A. Kozulin, trans.

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychological Science, 10*, 130–137.

Wilson, A.J., & Dehaene, S. (2007). Number sense and developmental dyscalculia. In D. Coch, G. Dawson, and K. Fischer (ed.) *Human behavior, learning and the developing brain: Atypical development*. Guilford Press, New York.

Wilson, A.J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an Adaptive Game Intervention on Accessing Number Sense in Low-Socioeconomic-Status Kindergarten Children. *Mind, Brain, and Education*, 3(4), 224-234.

Wilson, A.J., Dehaene, S., Pinel, P., Revkin, S., Cohen, L., & Cohen, D. (2006). Principles underlying the design of The Number Race, an adaptive computer game for remediation of dyscalculia. *Behavioral Brain Functions*, 2(1), 19, 2006.

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155-193.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749 – 750.

Xu, F., Spelke, E., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*, 88– 101.

Xu, F., & Arriaga, R. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology , 25* (1), 103-108.