THE UNIVERSITY OF CHICAGO

LEARNING MATHEMATICS THROUGH ACTION AND GESTURE:
CHILDREN’S PRIOR KNOWLEDGE MATTERS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY

ELIZA LINCOLN CONGDON

CHICAGO, ILLINOIS

AUGUST 2016

~ For Casey ~

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES v

ACKNOWLEDGEMENTS vi

ABSTRACT viii

CHAPTER 1 - INTRODUCTION 1

GENERAL OVERVIEW 1
PART I: LEARNING THROUGH ACTION 1
PART II: LEARNING THROUGH GESTURE 15
PART III: INDIVIDUAL DIFFERENCES, MATHEMATICS, AND MEASUREMENT 26

CHAPTER 2 – ESTABLISHING THE ROLE OF PRIOR KNOWLEDGE 32

INTRODUCTION 32
METHOD 36
RESULTS 41
DISCUSSION 49

CHAPTER 3 – TEACHING MEASUREMENT FROM ACTION TO ABSTRACTION 55

GENERAL OVERVIEW 55
INTRODUCTION 56
METHOD 58
RESULTS 62
DISCUSSION 69

CHAPTER 4 – ACTION, GESTURE, AND THE ROLE OF WORKING MEMORY 74

INTRODUCTION 74
METHODS 77
RESULTS 80
DISCUSSION 88

CHAPTER 5 – GENERAL DISCUSSION 94

PART I: PRACTICAL IMPLICATIONS 94
PART II: THEORETICAL IMPLICATIONS 97
PART III: BEYOND MEASUREMENT 99
PART IV: OPEN QUESTIONS 102
PART V: BROADER IMPACT AND CONCLUSIONS 103

REFERENCES 105

LIST OF TABLES

TABLE 1. PRE-TEST ONLY TASKS – EXPERIMENT 1……….………………………….………………….....46

TABLE 2. TRANSFER TASK IMPROVEMENT FROM PRE-TEST TO FOLLOW-UP – EXPERIMENT 1……47

TABLE 3. PERFORMANCE ON FOLLOW-UP TASKS – EXPERIMENT 1……………………………………..47

TABLE 4. TRANSFER TASK IMPROVEMENT FROM PRE-TEST TO FOLLOW-UP – EXPERIMENT 2……66

TABLE 5. PERFORMANCE ON FOLLOW-UP TASKS – EXPERIMENT 2……………………………………..67

TABLE 6. TRANSFER TASK IMPROVEMENT FROM PRE-TEST TO FOLLOW-UP EXPERIMENT 3……...84

LIST OF FIGURES

FIGURE 1. A SAMPLE "SHIFTED ITEM" 37

FIGURE 2. A PHOTOGRAPH OF EACH OF THE FOUR TRAINING CONDITIONS 40

FIGURE 3. DISTRIBUTION OF SCORES ON SHIFTED-PROBLEM TEST TRIALS 42

FIGURE 4. AVERAGE PERFORMANCE ON MAIN OUTCOME, EXPERIMENT 1 43

FIGURE 5. SHIFT IN STRATEGY USE, EXPERIMENT 1 45

FIGURE 6. SPATIAL LANGUAGE OR GESTURE USE AFTER TRAINING. 49

FIGURE 7. SAMPLE PERIMETER TEST ITEMS, EXPERIMENT 2.. 62

FIGURE 8. AVERAGE PERFORMANCE ON MAIN OUTCOME, EXPERIMENT 2 63

FIGURE 9. SHIFT IN STRATEGY USE, EXPERIMENT 2. 65

FIGURE 10. PERFORMANCE DURING TRAINING - HATCH MARK GROUP 68

FIGURE 11. PERFORMANCE DURING TRAINING - READ-OFF GROUP 68

FIGURE 12. CORRECT MOVEMENT DURING TRAINING - READ-OFF GROUP 69

FIGURE 13. TRAINING CONDITIONS, EXPERIMENT 3 79

FIGURE 14. EXPERIMENTER VIEW OF THE CORSI BLOCK TASK 80

FIGURE 15. AVERAGE PERFORMANCE ON MAIN OUTCOME, EXPERIMENT 3. 81

FIGURE 16. SHIFT IN STRATEGY USE, EXPERIMENT 3. 83

FIGURE 17. MEAN PERFORMANCE ON THE TWO WORKING MEMORY TASKS 85

FIGURE 18. DISTRIBUTION OF SCORES ON SPATIAL AND VERBAL WORKING MEMORY MEASURES... 86

FIGURE 19. AVERAGE WORKING MEMORY SCORES FOR LEARNERS AND NON-LEARNERS 87

FIGURE 20. AVERAGE WORKING MEMORY SCORES BY CONDITION 88

ACKNOWLEDGEMENTS

There are many people I would like to thank who have helped me at various stages of this project and throughout my graduate career. First and foremost, my advisors, Susan Levine and Susan Goldin-Meadow, who have provided me with invaluable professional guidance while serving as role models on both an academic and a personal level. At our very first meeting when I was a prospective student, the three of us shared research ideas over cappuccinos in the Meadow home and I immediately felt a sense of acceptance and belonging that has not wavered over the past six years. For that, and for so much more than I could possibly fit into an acknowledgments section, I am eternally grateful.

To my other committee members, Lindsey Richland, for providing constructive feedback and academic support throughout this process, and Amanda Woodward, for being an excellent mentor, teacher, and role model at every stage of my development as a scholar, I offer my sincere gratitude.

This work literally would not have been possible without the teachers, parents, and principals in the greater Chicagoland area who agreed to participate in each of these research projects. Thank you for allowing me an invaluable glimpse into your world. And to the research assistants and undergraduate scholars who dedicated their time and expertise to help me collect, analyze, and code data, I give you my deepest thanks. In particular, I want to thank University of Chicago undergraduates, Maria Bavaro, Marlene Lorenz, Frankie Carter, and Anna Rimlinger for expanding my thinking, for being willing learners, and for bearing with me through the moments of chaos and having the chutzpah to keep coming back for more. Thank you also to Joanna Schiffman, Claire Bradley, Hannah Degner, Rebecca Blackwell, Melissa Herrett, Anna

MacKinnon, Caroline Frazia, and Amanda Netburn for helping me to successfully run over 700 one-on-one instructional session over the years. It truly takes a village.

I also want to thank my developmental cohort 'siblings', Laura Garvin, Lauren Howard, and Miriam Novack for keeping my company on this journey. I am especially indebted to Miriam, who at various points has been my collaborator, editor, marriage counselor, bridesmaid, psychiatrist, stylist, life coach, gym buddy and dearest friend. I do not want to imagine what graduate school would have been like without you. I also want to extend a huge 'thank you' to my officemate and dear friend, Dominic Gibson, for keeping me grounded, for offering sage advice, for injecting his sense of style into our office décor, and for always being willing to talk shop over a bowl of ramen. And to E(Liz)abeth Wakefield, an invaluable collaborator, friend, and connoisseur of fine wines who has taught me so much about life, research, and tannins over these past few years. And to all of the other wonderful members of the Levine and Goldin-Meadow labs, thank you for countless hours of lab meetings, statistics workshops, and for always being willing to offer input and advice.

Last, but certainly not least, I want to thank my family. My husband, Casey, who has been my rock, my inspiration, and my biggest fan, and who makes the world seem five shades brighter. My kitties, Charlie and Kendall, for keeping my lap warm on long nights of writing, and for loving me no matter what p-value I get. To my siblings, Chal, Jodi, Dan, and Meryl, for teaching me that success comes in many forms and that true happiness comes from time with family. To my niece and nephew, Kate and Sam, whose light, life, energy, and antics keep me going even on the toughest days. And to my parents, Amy and Chal, for unconditional love and support, and for letting me go but always wanting me back. Thank you.

ABSTRACT

Teachers commonly use actions-on-objects (or actions with manipulatives) to help students understand mathematical concepts. This practice is based on the assumption that performing actions on *external* objects facilitates learning by changing or creating new *internal* ideas. Gestures are abstract, representational hand movements that can also help children to learn new ideas, but they differ from actions-on-objects in a key way -- they do not require learners to directly interact with the physical environment. I explore how this key difference between actions-on-objects and gestures affects learning outcomes in a particularly challenging area of elementary school mathematics, linear measurement. In Chapter 1, I find that for children with lower prior knowledge of measurement at pre-test, gesture-based instruction is largely ineffective. By contrast, actions-on-objects are effective for both higher and lower prior knowledge participants. In Chapters 2 and 3, I replicate this interaction between prior knowledge and movement type and further probe the boundaries of this effect. I end by situating the findings within the broader literature on the efficacy of gesture in instructional contexts. From a theoretical standpoint, the results suggest that the very features that make gesture so powerful and flexible in some instructional contexts (i.e. the fact that it does not necessitate physical interaction with specific objects) might make it inaccessible to some learners.

CHAPTER 1 - INTRODUCTION

General Overview

The overarching goal of this dissertation is to understand how individual differences in children's conceptual knowledge about linear measurement interacts with their propensity to learn through actions and gestures. To begin, I review the existing literature on *learning-by-doing,* a broadly defined term that captures a fascinating phenomenon: humans can acquire new conceptual knowledge through physical interactions with their environments. The first section focuses on when and how children can learn new ideas through their own actions-on-objects and through the actions of others. I then transition to the second section in which I review the existing literature on how children learn through hand gestures, which share some properties of transitive actions but also differ in some crucial ways. In the final section, I explore how individual differences in the learner might predict learning outcomes, and I review the existing literature on children's misconceptions in a particularly difficult area of mathematics, linear measurement.

Part I: Learning Through Action

1.1 Defining Action. Actions are a broad category that, when used colloquially, can refer to any time something is done or performed, or something is in a process or state of being active (e.g., "It's back in action"). In language, actions can be either transitive verbs, in which an action is directed at something named in the sentence (e.g., "I <u>kicked</u> the ball"; "I <u>pointed</u> at the sign") or intransitive verbs, in which the action is not directed at something named in the sentence (e.g., "I <u>walk</u> to school"). In this dissertation, the word "action" will be used to refer to a very specific subset of transitive actions—those that affect some type of change on an external object or representation. For example, drawing a diagram, moving a block from one place to another, in any way changing the spatial arrangement of objects, or erasing something would all be classified as actions. Tapping or pointing to an external representation would *not* be classified as "actions", as they do not affect any lasting change in the world. This classification will be important later, as the characteristics of gesture are defined to clearly include bodily actions or movements that do *not* change external representations.

1.2 Historical Perspective. John Dewey was a philosopher, psychologist and educational reformer and is perhaps one of the most famous early proponents of hand-on learning. In his book, *Democracy and Education*, first published in 1916, he wrote, "If knowledge comes from the impressions made upon us by natural objects, it is impossible to procure knowledge without the use of objects which impress the mind" (Dewey, 1916). He believed that education was a path to social reform, and that a true education and transmission of knowledge could not be achieved through the use of words alone, but instead that those words must be presented alongside the objects with which they are associated. He also believed strongly in the power of experimentation with physical materials as a path to acquiring new knowledge.

One of Dewey's contemporaries, Dr. Maria Montessori, founder of the modern day Montessori school was another vocal advocate for the idea of learning through action. Her educational philosophy was based on her experience that young children did not need to be verbally taught new ideas, but that they appeared to simply acquire new ideas through active exploration of their environments (Montessori, 1909; 1995). Such an observation was revolutionary at the time, and led Dr. Montessori to a long and illustrious career of real-world experimentation with classrooms of children in countries around the globe who learned not through traditional classroom lectures, but through touching, manipulating and exploring. These children could acquire new ideas simply by exploring enriched environments with thoughtfully designed toys, puzzles, tools and games. Modern day Montessori classrooms even have material where learners can discover ideas like the Pythagorean theorem.

In 1966, Jerome Bruner, a psychologist and education researcher wrote an influential book called, *Toward a Theory of Instruction*. In it, he suggests ways to alter the manner in which academic instruction is designed. For example, Bruner outlines a learning trajectory in which a child begins with concrete objects and multiple exemplars of a concept, then moves towards an abstract or symbolic understanding, but keeps hold of those original concrete exemplars as a way to continually anchor their abstract symbolic representations (Bruner, 1966). Bruner goes on to suggest that while not *all* discoveries need to be made by the child on their own, encouraging a process of self-discovery in the learner will help them to acquire the necessary concrete exemplars needed to serve as a basis for subsequent manipulation and generalization of more abstract ideas.

Dewey's, Montessori's, and Bruner's conclusions were echoed by Jean Piaget, the father of modern developmental psychology. In his work entitled, *Development and Learning*, Piaget explains his belief that children learn best in an environment of interaction. He writes, "Knowledge is not a copy of reality. To know an object, to know an event, is not simply to look at it and make a mental copy, or image, of it. To know an object is to act on it." (Piaget, 1974). He goes on to argue that these operations on objects or representations, when added together, "constitute the basis of knowledge". In this influential work, Piaget unabashedly extols not just the virtues, but the necessity of learning through acting before being capable of succeeding on problems with symbolic representation. As an example of the importance of action on objects, Piaget provides an anecdote of a young boy discovering the theory of conservation by continuously reordering pebbles. The boy forms many arrangements of pebbles and counts them each time, always coming up with the same answer. This, Piaget argues, is an example of why learning through acting is crucial. The boy did not discover a property of pebbles, per se, but through his actions on pebbles, discovered a more broadly applicable theory about conservation of number.

As these ideas grew in popularity, they began to influence policy and practice in the United States. The United States Institute of Education was first established in 1972 and one of the first large scale reports it funded was a series of investigations known as the Mathematics Education Information Reports. The relevant section for the current dissertation is from the 1977 version of the report and is entitled, "Activity-Based learning in Elementary School Mathematics: Recommendations from Research" (Suydam & Higgins, 1977). The report focused primarily on the use of manipulative materials and diagrams in elementary mathematics education. It argues that while activity-based curricula had become a popular technique for teaching mathematics, there was little to no official reports of which parts of these increasingly popular programs were working and which were not. It summarizes the existing literature up to that point, citing studies with conflicting results as to benefit of physical manipulatives, and generates other factors to consider in evaluating the effectiveness of teaching tools. Some of the same issues raised in this 1977 report are still unanswered questions today. How does prior knowledge of the learner interact with manipulatives? Does the type of representation matter? Is it better to start with something concrete and then transition to symbolic representations or vice versa? Does the content of the problem matter? Is it important that the learner manipulates the object, or can the instructor do so with as great an effect? While the report was unable to speak definitively to many of these questions, its official recommendations were ultimately based on the finding that children, overall, tended to learn more from lessons taught with accompanying concrete materials than from lessons with “symbolic treatments alone”. This official report, in many ways ahead of its time, marks the beginning of the transition from earlier theories of action-based learning to formal policy recommendations and implementation plans.

As evidenced by this rich history, the idea that one can learn through action is not anything new. In fact, this idea has continued to gain momentum and influence as time has passed, affecting both classroom practice and cognitive science and education research. One clearly related theory from the field of cognitive science is embodied cognition theory. In its purest form, embodied cognition is the idea that people’s active experiences and interactions with the world form the basis for all of human cognition and language (e.g., Neidenthal, 2007; Raymond & Gibbs, 2006; Wilson, 2002; Clark, 1997; Glenberg, 2008; Smith, 2005). Those who endorse the tenants of embodied cognition believe that thinking and reasoning are not purely internal processes of the brain, but rather rely on a dynamic system that includes brain, body and environment. This framework for understanding human cognition clearly has its roots in the learning-through-action phenomena popularized by the likes of Montessori, Piaget and Bruner. Moreover, it provides one of the first comprehensive theories to posit a mechanism by which external information can affect the internal cognitive system.

In sum, the early psychologists, educators, and philosophers that first popularized the idea of action-based learning have spurred decades of experimental work and advances in curriculum design and development. Paradoxically, this research has probably raised more questions that it has answered. Recent research has started to focus not only on whether or not action-based learning is beneficial, but *when, how* and *for whom*. Answers to these mechanistic questions have started to come from closely controlled behavioral work, neuroscience methodologies, and continuing advances in cognitive theories of learning more broadly. I will review many of these findings here, in an attempt to synthesize the current state of the field before moving on to explore how and why learning through gesture differs in crucial ways from learning via actions-on-object.

1.3 Learning Through Actions. Acting on external objects can permanently change our internal representations. This phenomenon has been documented even very early in life and is not limited to what we think of as traditionally "academic" domains. For example, in a study by Dr. Jessica Sommerville and colleagues, three-month old infants were randomly assigned to have active experience reaching for an object or not (Sommerville, Woodward, & Needham, 2005). Because grasping an object is a prohibitively difficult task for most three-month olds, the children assigned to the active experience condition were outfitted with 'sticky mittens', which would stick to an object if the infant's hand grazed the object at any point during the session. Infants who were given this experience were better able to subsequently interpret another person's reaching action than infants who did not. The author's suggest a tight link between an infant's own ability to successfully perform a reaching action and their ability to interpret a goal-directed reaching action in another actor. A second study by this group showed 10 and 12-month old infants a scene where an actor pulled at a cloth in order to get a toy (Sommerville & Woodward, 2005). While 12-month olds, in general, understood the goal-directed nature of this scene, 10-month olds were mixed; those who could perform the action on their own understood it, while those who were unable to complete the action successfully did not. This provides another example of how an infant's own motor experience allows them to learn about the actions and intentions of others.

Though motor experience is clearly an important part of learning during infancy (see Campos et al., 2000 for further examples of how locomotion in infancy affects infant's social and cognitive development), there is evidence that this relationship between action and learning persists across the lifespan. For example, in a study by Karin James, pre-literate children were either given experience actively writing the letters of the alphabet, or experience visually recognizing letters (James, 2010). Children in the writing condition showed *more* activation in visual association cortex during an fMRI scan than children who were given lots of visual identification practice. This finding suggests that practice physically writing out letters may be particularly crucial for children who are learning the alphabet and learning to read. Similarly, adults learning about angular momentum learned more when they were given the opportunity to physically experience the consequences of torque in an active learning environment than when they learned about angular momentum in a more traditional verbal format (Kontra, Lyons, Fischer and Beilock, 2012).

Perhaps the most commonly studied domain of learning through action is research on mathematical manipulatives. Manipulatives are objects that are designed to represent an abstract concept in a tangible, physical way. For example, young children may learn to count or add using blocks or other sets of small objects before they are expected to count Arabic numerals (e.g., Huttenlocher, Jordan, & Levine, 1994). Similarly, older children may initially learn about geometric principles, fractions, place value, currency, or balancing equations by using manipulatives such as tangrams, pizza-shaped disks, Cuisenaire rods, play money, or a balancing scale. Plenty of qualitative and experimental research supports the use of such tools in mathematics education. Manipulatives allow children to “offload cognition” onto the environment, encourage the formation useful conceptual metaphors (Manches & O’Malley, 2012), direct attention to the relevant components of a complex problem (Mix, 2010), and engage young learners with limited attention spans (Peterson and McNeil, 2008). “Offloading cognition” is like making a list on a piece of paper; it describes the process of using the physical environment as a sort of temporary placeholder where you can store information or ideas until you need to retrieve them again (Mix, 2010). For example, a young child may have trouble adding four addends in his head (e.g.: 7 + 2 + 6 + 4 = __). Moving around groups of blocks that represent the quantities of each addend may help the child keep track of what has been added already and what is still left to add without requiring him to keep everything stored in working memory. Manipulatives also support conceptual metaphor formation when the properties of the physical object (such as the length of a Cuisenaire rod) clearly correspond to a property of the target math concept (in this case, place value). These relationships can encourage analogical reasoning that leads to further insight and understanding.

1.4 Mechanisms of Learning Through Action. This section reviews proposed mechanistic explanations of *how* action-based instruction or experience might lead to learning over and above verbal instruction. The first part explores the role of the motor system in forming long-lasting neural representations of learned concepts. The second part investigates how actions-on-objects can direct (and hold) a learner’s attention to relevant components of complex problems. Finally, we turn to the literature on embodied cognition to understand how physical interactions with objects might help learners to create useful conceptual metaphors for more abstract problems.

It has long been established that motor encoding improves recall and memory for words and sentences over and above verbal repetition alone (e.g., Engelkamp, 1991; Cohen, 1983; Nilsson & Cohen, 1988; Saltz, 1988), and that this effect is stronger for *doing* actions than for *seeing* actions. According to the theory of Parallel Distributed Processing (PDP), a widely accepted framework for thinking about the process of encoding and recall, forming a mental association between a motor action and a new concept increases the size and strength of the neural network that represents that concept, making it more likely that it will be appropriately reconstructed when cued (e.g., McClelland, 1994). To demonstrate this phenomenon mechanistically, researchers typically begin by establishing a behavioral advantage for those participants who learn a new idea or concept via action. Then, to reveal the 'neural signature' of such a process, they use neuroimaging methods to show that learners who *have learned* something (a novel word, object, idea) paired with an action will subsequently show motor activity even when they are no longer physically acting. For example, participants who are trying to learn paired associations between novel objects and novel sounds show faster and more accurate learning from self-generated motor actions on objects than from passive manipulations of the objects (Butler and James, 2013). Importantly, in this same paradigm, active learners show greater motor reactivation and greater connectivity between motor and visual cortices at later test than participants who learn through observing an experimenter perform actions (Butler, James & James, 2011). In addition, this paradigm promotes stronger neural associations between multi-sensory regions of the cortex (eg: superior temporal sulcus, which is implicated in audiovisual integration). To date, this kind of effect seems to be especially strong for self-generated actions, a finding that may be particularly true in younger populations (James and Swain, 2011). In one of the first studies investigating action-based learning from a neurological perspective in children, 5- and 6-year-olds were taught about novel object and word pairings. Only children with active experience showed motor reactivation during an auditory perception task in the scanners (James and Swain, 2011). Taken together, these findings imply that active learning may be a particularly beneficial way to enrich the representations of a new idea or object in one's memory to make it easier for subsequent recall.

Actions can also help to direct a learner's attention to important components of a complex problem. In the domain of mathematics, an expert can see mathematical relationships everywhere in the external environment, but for a novice, it may take a well designed manipulative and a proscribed set of actions to focus that learner's attention on the relevant properties of the environment (Mix, 2010). For example, a child who is learning about fractions may have been exposed to sliced pizza and half-full juice containers their entire lives without extracting a concept like the part-whole relationship of fractions. Exposure to a perceptually simple manipulative with accompanying actions that highlight the components of the formal definition of fractions (e.g., adding or removing parts of a whole circle while labeling each newly created fraction) can direct a learner's attention to important relationships in the environment more effectively than passive exposure to natural stimuli or manipulatives. Performing an action with an object may also make that object more interesting and engaging to a learner more generally, something that is particularly useful when instructing young children (Peterson & McNeil, 2008).

Beyond the creation of enriched neural representations and the ability to engage and direct a learner's attention, the theory of embodied cognition offers a stronger hypothesis about the power of action-based learning. As a reminder, embodied cognition posits that human thought is grounded in action because the brain is, by definition, situated within a moving, acting body. As such, it follows that actions are uniquely suited to help new conceptual information ‘enter’ the cognitive system. Art Glenburg argues that the body is specifically and specially evolved to perceive, act and emote, and that many of these processes contribute directly to “higher” cognitive processes (Glenburg, 2008). He claims that the body and higher order thinking, such as mathematics, are not simply related, but inextricably linked throughout development (see Andres, Seron & Olivier, 2007 for a TMS study relating adult counting to the hand long after finger counting has ceased). Other researchers have echoed these sentiments, arguing that all of mathematics, for example, is actually grounded in physical metaphors (Lakoff & Nunez, 2000). While somewhat controversial in it’s strongest form, this theoretical position is consistent with many of the findings on action-based learning. For example, thinking of the brain and the body as a bidirectional cognitive system could explain why learners can acquire new information through actions *and* use their actions to ‘offload cognition’ onto the external environment and decrease demands on working memory (Mix, 2010). It can also help to explain why the physical properties of mathematical manipulatives (e.g., weight, length, size, location), can give learners unique insight into mathematical ideas that they can use to create conceptual metaphors, which are then translated into more abstract symbols like Arabic numerals, diagrams, or mathematical equations (Manches & O’Malley, 2012).

1.5 Drawbacks and Complications of Learning Through Action. In the domain of mathematics, the area of focus in this dissertation, action-based learning almost always refers to the use of mathematical manipulatives. As mentioned previously, manipulatives are objects whose physical properties are meant to instantiate some kind of math concept or idea, or whose properties afford actions that represent a specific mathematical concept (e.g. counting or comparison of length). They are used ubiquitously in classroom settings, particularly during elementary school. Until this point, we have primarily focused on the benefits of action-based learning, and the mechanisms that may underlie these benefits. However, there is a growing literature that has found drawbacks of action-based learning, and cautions against blind implementation of manipulative-based lessons (eg: Mix, 2010; Kamii, Lewis & Kirkland, 2001; Uttal, Scudder and DeLoache, 1997; Ball, 1992; Friedman, 1978; Fuson & Briars, 1990; Goldstone & Sakamoto, 2003; Goldstone & Son, 2005; Kaminski, Sloutsky & Heckler, 2005; 2006a, 2006b; Moyer, 2001; Peterson, Mercer & O'Shea, 1988; Resnick & Omanson, 1987; Sowell, 1989; Suydam & Higgins, 1977; Wearne & Hiebert, 1988). In the final section of Part I, I will discuss some of these potential drawbacks of action-based instruction. I focus on mathematical manipulatives, though the issues identified here are likely more broadly applicable to any object-based instructional technique.

In some cases, mathematical manipulatives are not only ineffective, but are harmful to students' ability to learn and generalize new mathematical concepts (e.g.: Mix, 2010; Kamii, Lewis & Kirkland, 2001; Uttal, Scudder and DeLoache, 1997; Ball, 1992; McNeil, Uttal, Jarvin & Sternberg, 2009; McNeil & Uttal, 2009). In the initial learning process, children do not always spontaneously create the appropriate conceptual metaphor that links the physical objects to the abstract concept they are meant to represent. For example, when learning to balance a mathematical equation, children might be presented with a physical scale that allows them to physically balance sets of weights. Though the link between the physical scale and the symbolic equation is clear to the instructor or teacher, the child may not spontaneously make this connection. Without this connection, it might also be hard for children to understand which physical properties of the stimulus are actually relevant to the underlying concept (Uttal, O'Doherty, Newland, Hand & DeLoache, 2009). The smooth texture or color of a block is completely irrelevant to its function as a counting object, but this may not be immediately apparent to a young learner.

One way to explain these difficulties for learners is to consider that even ‘concrete’ manipulatives have symbolic properties (Uttal, Scudder and DeLoache, 1997). And to successfully use these tools in an action-based lesson, children must keep track of the relevant properties of the physical object they are manipulating *and* the properties of the broader, abstract concept symbolized by those movements. This task for the learner, coined the “dual representation” problem, offers one explanation as to why object-based instruction may lead learners to see any learned actions as relevant only to the materials at hand, rather than to some broader, generalizable idea (DeLoache, 2000).

This dual representation hypothesis also has implications for the ideal perceptual properties of manipulatives. Perceptually rich objects may be most likely to obscure the crucial link between the object and it’s abstract conceptual counterpart. For example, we know that flashy or colorful materials, while they may get children’s attention, tend to cause learners to focus on the properties of the object rather than of its symbolic counterpart (Uttal et al., 2009). This may be especially true for manipulatives that serve an alternative, familiar function to the learning, like using toy cars to learn how to count (Peterson and McNeil, 2008). While some schools in the United States have adopted the use of perceptually sparse materials (e.g., Montessori schools), it is common in classrooms to use many exemplars of exciting, engaging materials on the assumption that these will be more beneficial to children than using the same simple materials over and over. In a cross-cultural comparison between Japanese and American schools, one study noted that American teachers tended to seek variety in their materials during a math lesson, while the Japanese teachers purposefully repeatedly used the same standard kit of materials (Stevenson and Stigler, 1992). While a number of other cultural factors are also at play, it is notable that Japanese students consistently outperform American students on international mathematical assessments across elementary, middle and high school (TIMSS, 2011; Foy, Arora & Stanco, 2013; PISA, 2012), suggesting that in some cases, perceptually simple materials might be best.

In addition to some of the issues with learning outlined above, it is also important to consider the degree to which students can *transfer* learned knowledge to novel situations and contexts. Transfer is one of the hallmarks of “deep learning” (the other being retention across time). A 2009 study by Kaminski and colleagues showed that while perceptually rich concrete symbols initially helped children learn new mathematical symbols, it was actually children in a “language only” training condition who were more readily able to transfer their understanding to new symbols (Kaminksi, Sloutsky, and Heckler, 2009). A similar pattern of results was reported in some of my own recent work whereby children in an action-based condition were less able to generalize to novel problem types than children in a more abstract gesture training condition (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014).

1.6 Interim Conclusions. Overall, it appears that manipulatives can be very useful for engaging learners and helping them to understand abstract concepts by grounding ideas in the physical properties of an object. Yet these very same strengths are also identified as some of the biggest limitations of manipulative-based learning. What can explain this apparent contradiction? A newer field of research on gesture-based instruction points towards possible answers. Gesture, as a type of action, involves movement of the body. Unlike action, however, gesture does not involve interacting directly with the physical environment of affecting permanent change on any external representations. What are the implications of this crucial difference? When does gesture help learners? When does it harm learners? What are some of the mechanisms of gesture-based instruction and learning and how is it similar or different from object-based action instruction? These are the questions that will be explored in the next section.

Part II: Learning Through Gesture

2.1 Defining Gesture. Gestures are movements performed with the hands that do not interact directly with objects or cause changes in the physical environment. They often accompany speech and have been described and categorized by scholars into several broad, descriptive classes (see Ekman & Friesen, 1969; Kendon, 2004; McNeill, 1992 for classification systems). In my dissertation I primarily focus on one specific type of gesture, representational gesture, which is very common in instructional contexts. Representational gestures include *iconic* gestures, which convey information through the similarity between their form and their referent (e.g., using ones hands as the edges of a triangle to talk about the angles of the triangle), *deictic* gestures, which identify objects or locations in the world (e.g., pointing to the edge of a triangle drawn on the chalkboard), and *metaphoric* gestures, which represent ideas through a metaphoric relationship between their form and their meaning (e.g., making one's hands into a triangle to talk about the components of a mediation analysis).

Gestures differ from actions-on-objects in a few key ways. The first way, as mentioned above, is that gesture does not cause a lasting change in the external world. When teachers use their hands to act on directly on representations in the physical world (e.g., moving around mathematical manipulatives, conducting science demonstrations, or drawing diagrams on the chalkboard), they are producing object-directed actions*,* not gestures. However, if a teacher moves her hands towards objects without touching them, or moves her hands in a way that represents or mimic a previously performed action, these movements are construed by both the teacher and the children as gestures, and are processed categorically differently than object-directed actions (Novack, Wakefield, & Goldin-Meadow, 2016).

The second important difference between gesture and other forms of movement is that gesture can represent information through its form or movement trajectory (e.g., making a twisting hand motion with curved fingers could represent taking the lid off of a jar). This makes gestures different from object-directed actions, which typically represent themselves (e.g., actually taking a lid off of a jar), and from movements that are performed for their own sake, such as dancing or exercising (Schachner & Carey, 2013). Because the purpose of a gesture is not the movement itself, but rather the idea that that movement represents, gestures are considered *representational* actions, and this special feature of gesture has crucial implications for thinking and learning. For example, gestures are produced prolifically in educational contexts while teachers are explaining new ideas, particularly in STEM (Science, Technology, Engineering, Mathematics) classes. Imagine, for example, you are a science teacher trying to explain what makes a molecule a stereoisomer. You need to explain to your students how mentally rotate the molecule comparing the locations of the various atoms. Describing this complicated mental transformation in words would likely be confusing and cumbersome, however, using your hands to *show* the rotation might be more natural, and make more sense to the students. Teachers will also spontaneously gesture while linking new concepts in math instruction in a classroom setting (Alibali et al., 2014), or when instructing children individually on math problems (Goldin-Meadow, Kim, & Singer, 1999) as well as when explaining new ideas about geoscience (Roth, 2007; Roth & Lawless, 2002). In fact, it is rare for teachers *not* to gesture while explaining new concepts (Goldin-Meadow, Kim & Singer, 1999). Students themselves are also likely to gesture while trying to learn new ideas. They spontaneously gesture while explaining their reasoning to Piagetian conservation problems (Church & Goldin-Meadow, 1986), mathematical equivalence problems (Perry, Church, & Goldin-Meadow, 1988), geological phenomena (Atit et al., 2013; Kastens, Agrawal, & Liben, 2007), or chemistry structures (Stieff, 2011).

2.2 Spontaneous Gestures and Learning. Teachers of all grade levels, including college professors, use gestures spontaneously, prolifically, and often unconsciously to help draw their student's attention to important components of a problem, demonstrate spatial relations between objects, or pantomime dynamic processes (Goldin-Meadow, Kim & Singer, 1999). But crucially for the current dissertation, spontaneous gesture is not just useful for clarifying a message to a listener; it also serves a function for the speaker or learner. For example, a blind individual talking to another blind individual will naturally produce gesture, as will a person who is talking on the phone (Goldin-Meadow, 2005). In neither case is the speaker gesturing for the benefit of clarifying a message to her audience. One favored hypothesis to explain this phenomenon is that producing gesture can reduce working memory demand on a speaker (Baddeley, 1986), particularly when the gesture is *meaningful* to the speaker rather than a series of meaningless movements (Cook, Yip & Goldin-Meadow, 2012). In an educational setting, this reduction in working memory demand could free up limited neural resources to help a learner reason through a difficult problem or situation.

Moreover, we know that the gestures that children produce when thinking through difficult problems can convey important conceptual information that is often absent from their verbal explanations. For example, a toddler still in the process of learning her numbers may say the wrong number if asked how many buttons are shown (e.g., say "two" when the correct answer is 3). Yet that same child may be able to use her hands to *show* the correct number of buttons (i.e., holding up three fingers) (Gunderson, Spaepen, Gibson, Goldin-Meadow, & Levine, 2015). Similarly, a 3rd grader given a missing addend equivalence problem (e.g., 4+6+9 =__+9) may incorrectly say that she can solve the problem by adding up all of the numbers (e.g., "I added the 4, 6, 9, and 9, and got 28"). However, while saying that, the student could sweep one hand under the left side of the problem (as she talks about adding up the numbers) and then sweep her other hand under the other side of the problem (while talking about arriving at the incorrect answer, 19), a subtle demonstration that she is beginning to notice something meaningful about the fact that the equation has two separate sides (Perry, Church & Goldin-Meadow, 1988).

These fascinating speech-gesture *mismatches*, in which learners provide more information in their gesture than in their speech, occur across age ranges and different academic domains. Importantly, learners who produce speech-gesture mismatches are actually more likely to learn from instruction than learners who do not (Alibali & Goldin-Meadow, 1993). In addition to the domain of mathematics, the phenomenon of speech-gesture mismatches predicting learning has been demonstrated among toddlers on the cusp of producing two word utterances (Iverson & Goldin-Meadow, 2005), 5-7 year olds learning about Piagetian conservation (Church & Goldin-Meadow, 1986), 9-year-olds discussing moral reasoning dilemmas (Beaudoin-Ryan & Goldin-Meadow, 2014), and even college students learning about organic chemistry (Ping, Decatur, Larson, Zinchenko, & Goldin-Meadow, 2016). Across all of these instances, the learner's spontaneous gesture and how the meaning expressed in its form or motion trajectory relates to the meaning expressed in speech can serve as a marker that a student is 'ready to learn', more so than focusing on the spoken explanations alone.

2.3 Manipulating Gesture to Promote Learning. The prevalence of learners' spontaneous gestures in learning and instruction settings led researchers to ask whether manipulating learners' gesture production might lead to insight. As it turns out, just like asking a child to perform a specific set of actions with a mathematical manipulative, you can ask a child to perform *specific* representational gestures to improve learning outcomes. This effect has been replicated across a variety of academic domains including algebra, chemistry, geometry and word learning (e.g., Wakefield & James, 2015; Macedonia, Müller, & Friederici, 2011; Ping & Goldin-Meadow, 2008; Valenzeno, Alibali, & Klatzky, 2003; Singer & Goldin-Meadow, 2005). There is even some promising evidence that producing gesture helps children to transfer their knowledge to new contexts (Cook, Duffy & Fenn, 2013; Novack, Congdon, Hemani-Lopez & Goldin-Meadow, 2014) and better retain newly learned information across time (e.g., Cook, Mitchell & Goldin-Meadow, 2008; Levine, Goldin-Meadow, Carlson & Hemani-Lopez, 2016).

For example, in one study, children were taught to produce a grouping gesture (a *v-point* to the first two addends on the left hand side of the equation, followed by a *one-finger point* to the blank) while saying an equalizer strategy aloud ("I want to make one side equal to the other side") (Goldin-Meadow, Cook & Mitchell, 2009). In the equation, 4+2+7=__+7, the children would make a *v-point* gesture to the 4 and the 2, then a single point to the blank. A second group of children was taught the same mismatching speech and gesture, but they were taught to produce the grouping gesture toward the incorrect grouping addends (2 and 7 in the example above). A third group of children was only taught to reproduce the spoken equalizer strategy.

Importantly, children in the two gesture conditions were never told the meaning of the gesture. Results showed that children who were taught to produce the correct gesture and mismatching speech learned significantly more from training than their peers who did not produce any gesture. This suggests that mismatching speech and gesture, even when it is artificially manipulated, still promotes conceptual change. Somewhat surprisingly, children in the *incorrect gesture* training condition learned more from training than children in the speech alone condition. This could mean that the idea conveyed in the gesture (i.e., add two numbers together), and not just the attention to specific numbers by the gesture, may have aided in conceptual change. Finally, across both gesture groups, children actually incorporated the grouping strategy into their spoken explanations after training, meaning that they learned and internalized a new problem-solving strategy that they had only ever produced in gesture.

2.4 Mechanisms of Learning Through Gesture. In thinking about the differences between actions-on-objects and gestures and how they may differentially promote learning and insight, it is useful to consider the known mechanisms underlying each type of movement experience. In Part I, we explored how self-produced actions engage the motor system, direct a learner's attention, and help learners to develop conceptual metaphors through guided physical interactions with real-world objects. Here, we explore some of the mechanisms that have been proposed to explain gesture's effect on learning outcomes. Similar to actions-on-objects, gestures engage the motor system and direct a learner's attention to relevant components of a problem. In contrast to actions-on-objects, gestures are produced in perfect synchrony with spoken language during communicative and instructional settings, and they are representational and are physically removed from the actions or ideas they represent. All four of these mechanisms has implications for information encoding, learning, and memory.

Like other kinds of actions, gesture engages the motor system and can change the way information is processed. For example, in one study, participants were either allowed or prohibited the use of gesture when solving a spatial gear-task. Those who were allowed to gesture persisted in using a perceptual-motor based strategy whereas those who could not, used an abstract reasoning strategy (Alibali, Spencer, Knox & Kita, 2011). In a different study, when individuals were asked to explain their solution to a Tower of Hanoi puzzle problem, the gestures they produced during their explanations influenced their internal representations of the problem, which affected how they subsequently solved it (Beilock & Goldin-Meadow, 2010; Goldin-Meadow & Beilock, 2010). There is even some behavioral evidence using a motor interference task that interpreting someone else's gesture engages one's own motor system (Ping, Goldin-Meadow, & Beilock, 2014). That is, if a listener is asked to move their arms or hands while a speaker is gesturing, the listener's comprehension of the speaker's gestures is impaired.

Neuroimaging evidence further supports the role of the motor system in processing information learned through gesture. When watching co-speech gesture, adults show activation in motor planning regions (Wakefield, James, & James, 2013), directly supporting the behavioral finding of Ping et al. (2014). After producing symbolic or iconic gestures while learning new information, such as musical melodies or new vocabulary words, motor areas are also activated when these stimuli are subsequently encountered (Macedonia, Muller, & Friederici, 2011; Wakefield & James, 2011), suggesting that the learner formed a link between the concept and corresponding motor movements. Recent work has also found that children who learned the concept of mathematical equivalence through a speech and gesture strategy activated motor regions when later solving math problems, compared to children who learned through a speech alone strategy (Wakefield, Congdon, Novack, Goldin-Meadow, & James, 2016). These neuroimaging results, together with the behavioral finding that gesture instruction enhances learning, suggest that the motor engagement during learning with gesture may cause subsequent motor reactivation, which provides learners with richer and more robust representations of newly acquired ideas.

Gesture, like some kinds of action-on-objects, can also help to direct a learner's visual attention to important locations in the spatial environment. Even children as young as 4.5 months will shift their visual attention following a dynamic, deictic gesture (Rohlfing, Longo, & Bertenthal, 2012). Adult learners also pay attention to gestures, particularly those that pause in space, as those gestures are often indicating the relevance of a particular spatial location (Gullberg & Holmqvist, 2006). In a chaotic visual world with many competing sources of potential information, this ability to direct or capture visual attention has clear consequences for learning. For example, gesturing towards the referent for a new word can facilitate learning a label for that object (Rader & Zukow-Goldring, 2012); tracing an outline of two symmetrical objects highlights the relation between the two objects and facilitates learning of the concept of symmetry (Valenzeno, Alibali, & Klatzky, 2003); and gesturing towards two sides of a mathematical equivalence problem can clarify the role of the equals sign (e.g., Cook, Duffy, & Fenn, 2013).

We know that gesture itself can highlight particular regions of space and represent information, but it does not do this in a vacuum. Gestures are predominately produced with spoken language, and some have even argued that these two streams of communication, gesture and speech, emerge from a single integrated system (McNeill, 1992, 2005). As such, it may be the case that gesture's privileged relationship with speech is one of the key mechanisms underlying its power to promote learning. For example, when a learner's own gesture reveals semantic information that is different but complementary to the information found in speech (a *speech-gesture mismatch*), that learner is very likely to profit from instruction. Singer and Goldin-Meadow (2005) investigated whether the same was true when children were watching gesture instruction. Children were given instruction on mathematical equivalence by a teacher who explained either one or two strategies, and varied whether these strategies were presented through speech, speech with 'matching' gesture (expressing the same information as speech), or speech and 'mismatching' gesture (expressing two correct, but different strategies in speech and gesture). Children performed best on a posttest if they learned through gesture and speech that expressed different information, suggesting that the integration of the two complementary ideas across two modalities provided the most comprehensive instruction.

In my own recent work, I expanded upon these findings by showing that the temporal simultaneity of the information in the mismatching speech and gesture instruction is crucial for learning (Congdon et al, 2016). In a comparison of instruction through simultaneous speech-and-gesture, sequential speech-then-gesture instruction, and sequential speech-then-speech instruction, children learned most from the simultaneous speech-and-gesture training condition. We argued that part of the facilitative effects of gesture on learning stem from the integration of information from speech and gesture that occur when these modes of communication are presented simultaneously. Because speech and gesture can both communicate information, but do so through different modalities, speech and gesture instruction is uniquely situated to allow for this integration. Neuroimaging data supports this claim, finding that processing speech and gesture activates overlapping neural regions (e.g., Holle, Obleser, Rueschemeyer, & Gunter, 2010; Willems, Özyürek, & Hagoort, 2009). Kelly and colleagues (2014) found that information present in gesture influenced how simultaneous speech was processed and vice versa. Crucially for the purposes of this dissertation, gesture seems to share a privileged relation to speech that is not shared by other forms of action: in the same study, the authors found a significantly smaller bidirectional influence of speech and object-directed action processing (Kelly, Healy, Ozyurek & Holler, 2014).

Finally, gestures are highly representational. That is, the form or motion of the gesture can represent ideas, actions, spatial information or spatial relationships (Kita & Ozyurek, 2003; McNeill, 1992). For example, when explaining mathematical equivalence, an instructor might make a *v-point* gesture to the first two addends of the equation to represent the idea that those two addends should be combined into a single quantity. Or someone who is reasoning through a difficult mental rotation task may use their hands to represent the features of the to-be-rotated object (Chu & Kita, 2008). Eye tracking studies with adults suggest that when we look directly towards these kinds of representational gestures (Beattie, Webster, & Ross, 2010), it is typically because the form of this type of gesture has previously provided us with useful semantic information. Adults are especially likely to attend to gestures during the stroke phase of a gesture when that semantic information is most likely to be present (Beattie et al., 2010).

The fact that gestures, unlike other forms of actions-on-objects can convey semantic information or ideas without directly manipulating the physical world may be a particularly important component of gesture. As mentioned in Part I, actions-on-objects can occasionally cause learners to focus on irrelevant physical features of the objects. Gesture, by contrast, are not beholden to the affordances of any single object or set of objects, but instead can provide an abstract representation that highlights only the most crucial features of a concept or idea. Such a possibility is supported by my work showing that children who learned a novel concept through gesturing towards objects were able to generalize that concept to novel problem types, more so than children who learned through actions-on-objects (Novack, Congdon, Hemani-Lopez, & Goldin-Meadow, 2014). This work suggests that the features that make gesture *different* from actions-on-objects, while still engaging the motor system and directing a learner's visual attention, may also be important to its effects on learning, retention, and generalization.

2.5 Potential Drawbacks and Complications. Just like object-based instruction, there is reason to think that gesture-based instruction may not work well under all circumstances. Very young children can understand another person's action, like demonstrating how to twist off the top of a container, long before they can interpret a gesture that represents that action, like miming a twisting motion near the top of a container (Novack et al, 2014). This evidence suggests that iconic gesture interpretation follows a later and more protracted developmental time span than action interpretation. Consequently, the meaning of iconic gestures may be opaque to some children, particularly if they are unfamiliar with the specific problem or context being demonstrated. There is also a possibility that gestures may be better suited for teaching abstract problems, or problems that require flexible conceptual understanding, while actions-on-objects may be better suited for teaching more concrete problems. In general though, the drawbacks and limitations of gesture-based instruction are poorly characterized. As such, this will be a major focus of the current dissertation.

To date, research suggests that *because* gestures are flexible and dynamic in form and do not require interaction with objects, they are free from the influence of any potentially misleading physical properties of manipulatives. As such, they may be better suited than actions-on-objects for promoting learning, transfer, and generalization. In this dissertation, I explore whether gesture's physical and metaphorical "distance" from the object-based action may make the meaning of a gesture unclear to some learners, particularly learners who have lower levels of prior conceptual knowledge. In other words, what are the potential boundaries of gesture's positive effects on learning, and how can that inform our understanding of the conceptual representations formed by young learners when they are trying to learn a new idea through actions or gestures?

Part III: Individual Differences, Mathematics, and Measurement

3.1 Individual Differences in the Learner. The goal of this dissertation is to understand how action- and gesture-based instruction on a linear measurement task might differentially promote learning for children who are at different stages of conceptual development. In doing so, it is important to understand the existing literature on individual differences in learners and how that can predict the efficacy of different types of instruction. For example, in some recent work by Siler and Willows, the concreteness of the lesson in a modular arithmetic task interacted with both age and ability (Siler and Willows, 2014) such that younger, 6^{th} grade students or students who were low performers at pre-test benefitted most from a more concrete representation and older, 8^{th} grade students or students who were high performers at pre-test seemed to benefit most (and transfer best) after learning from a more abstract representation of the concept. In this work, the authors demonstrate that the benefits of concreteness during instruction depend on the age of the child, the deductive reasoning skills of the child, and the characteristics of the task itself.

When one group of children learns best from one type of instruction and another group of children learns best from another type of instruction, this is referred to as an 'aptitude-treatment interaction' (Cronbach and Snow, 1977; Snow, 1989). For example, in a study teaching college students a new concept in chemistry, participants with lower working memory learned significantly more from audiovisual input than from visual-only input (Seufert, Schütze & Brünken, 2008). Participants with higher working memory learned equally well from both types of input. Crucially, higher working memory students showed much higher rates of *transfer* from visual-only input, while lower working memory students showed much higher rates of transfer after audiovisual input. This effect can also be modulated by a learner's prior knowledge of a given domain. For example, in multimedia learning, instructional design that may help a learner with low prior knowledge may not help, or may even hinder learning for someone with higher prior knowledge (Kalyuga, 2005). In the learning process, the learner must integrate the new information in the instruction with their pre-existing knowledge, a process that must, necessarily, be calibrated to that level of prior knowledge to avoid unhelpful or harmful interference.

There are also some cases where one group of children shows lower rates of learning overall, irrespective of instruction type. In previous work within the domain of linear measurement, children who began a training study with a more severe misconception learned less from instruction overall than their higher prior knowledge peers across four different types of interventions (Kwon, Ping, Congdon & Levine, 2016). In measurement, children who have this more severe misconception also tend to be the younger students, or students from lower socio-economic status (SES) backgrounds (Levine, Kwon, Huttenlocher, Ratliff, & Dietz, 2009). Being from a low SES background, in particular, is generally correlated with slower rates of learning and cognitive change throughout ontogenetic development and across multiple domains (e.g., Landry, Denson & Swank, 1997; Alexander, Entwisle & Olsen, 2001; Hoff, 2008). In the domain of word learning, researchers found that children who had low phonological competency learned less from instruction overall than children who had high phonological competency (Wakefield & James, 2015).

3.2 Mathematics as a Domain. Many of the examples in this text have centered on the field of mathematics. For several reasons, mathematics may be a particularly fruitful domain in which to consider this question of how children learn new conceptual ideas through actions and gestures. From a practical perspective, children in the United Stated consistently perform poorly on international assessments of mathematics achievements as compared to other developed countries (TIMSS, 2011; Foy, Arora & Stanco, 2013). Furthermore, early success in mathematics is highly predictive of a number of positive educational attainment outcomes including graduation rate, college attendance, and success in STEM (science, technology, engineering, mathematics) disciplines (The National Science Education Standards, National Research Council, 1996). With ever-widening racial and socio-economic achievement gaps in our nation (Lee, 2002 and Reardon, 2011, respectively), there has never been a more important time to support research that aims to understand and optimize mathematical learning outcomes for all students.

From a more theoretical perspective, mathematics offers a variety of subdomains that vary in the degree to which they are abstracted away from the physical world. For example, areas like Euclidean geometry and measurement have very clear links to the concrete, physical world. By contrast, areas like category theory and model theory are so far removed from their origins in the physical world that even mathematicians do not see them as having any relevant connection to concrete examples. Something like algebra, which has origins in the physical world but is typically taught as a set of more abstract rules, might fall someone in between the two extreme ends of the spectrum. Overall, the majority of mathematics, even geometry and measurement, might be considered abstract on some level, in that the purpose of the domain is to quantify relationships in the natural world through the use of symbols and agreed-upon conventional notation that can be generalized to any number of novel contexts.

Determining the efficacy of teaching interventions or understanding 'instructional complexity' while taking into consideration the content area, the individual characteristics of the learner, the learning goal, and the type of instruction is a computationally prohibitive problem for researchers (Koedinger, Booth & Klahr, 2014). For now, the recommended course of action is to choose a computationally tractable set of variables, understand the boundaries of the instructional effect, and then expand outwards to consider other factors and how they may interact with the documented phenomenon.

3.3 Focus on Measurement. In the current dissertation, I accomplish this goal by focusing on a particular content area within mathematics, linear measurement. Linear measurement is a relatively concrete area of elementary school mathematics that also happens to be a source of pervasive student misconceptions. In fact, measurement remains the only domain of mathematics in which elementary school children consistently perform lower than the international average (TIMSS, 2011; Foy, Arora & Stanco, 2013). A true understanding of units of measurement requires children to understand that equal partition of space is important. However, almost all mathematics curricula focus on teaching procedural strategies of measurement, and focus less on improving a necessary conceptual understanding of measurement (Sophian, 2003; Smith et al, 2008; Smith, Males, Dietiker, Lee, & Mosier, 2013).

When children are asked to measure an object that is not aligned with the "0" point of a ruler, they consistently make one of two kinds of well-documented errors, which reflect two distinct levels of conceptual understanding (Lehrer, Jenkins & Osana, 1998; Solomon, Vasilyeva, Huttenlocher, & Levine, 2015). Some children incorrectly count the hatch mark lines instead of the intervals of space that fall between an object's left-most and right-most edges (hatch-mark counting strategy). Other children simply read off the number on the ruler that aligns with the rightmost edge of the object (read-off strategy). Importantly, both strategies consistently provide the correct answer when the object to be measured is aligned at the "0" point of a ruler, a scenario typical of classroom instruction. These errors on shifted-object measurement tasks make it clear that children are largely dependent on a set of procedural rather than conceptual skills that help them arrive at the correct answer on traditional kinds of measurement test items (Kamii & Clark, 1997; Martin & Strutchens, 2000; Lehrer, Jenkins & Osana, 1998). Moreover, this work shows that children consistently struggle with conventional rules such as where to start counting, what to count, and the significance of the hatch marks and numbers on the ruler (Solomon et al., 2015).

To date, the majority of research on linear measurement has documented the nature of children's difficulties and misconceptions. Very little research has focused on designing effective ways to address though misconceptions. From the field of education research, one group did an in-depth case study with eight students who were given a number of different measurement instruction activities and continuously assessed across nearly a full year (Barrett et al., 2012). Based on their findings, the authors propose some useful instructional tasks that might help push children from one conceptual stage of measurement understanding to the next. However, this work used many instructional strategies at once, and did so over a long period of development, making it difficult to ascertain which specific features of the instruction might be driving improvement. In some of my own previous work, I tested two principles from the field of cognitive science, disconfirming evidence and structural alignment, to see which was most effective in improving children's performance after a very short training intervention (Kwon, Ping, Congdon & Levine, 2016). We gave children a lesson on either shifted-object ruler problems or more traditional unshifted-object ruler problems with either small plastic unit chips or no chips. We found that shifted-object practice was crucial, and that alignment of discrete plastic unit chips marginally increased the speed of learning during the training session.

In the current study, I aim to expand upon this literature by testing a new type of unit representation, gesture. More specifically, I explore the relative efficacy of action- and gesture-based instruction within two groups of students at different levels of conceptual understanding (read-off strategy vs. hatch-mark strategy). In doing so, I explore the boundaries and limitations of these types of instruction, and identify the features of gesture, in particular, which prove difficult for low prior knowledge learners. This effort represents the beginning of a broader discussion of learning through action and gesture, and the general discussion will situate my findings within a broader literature that is addressing different pieces of this question.

CHAPTER 2 – ESTABLISHING THE ROLE OF PRIOR KNOWLEDGE

Introduction

We know from decades of experimental psychology research that asking children to act directly on external representations can affect their internal ideas (e.g., Wilson, 2002; Sommerville & Woodward, 2010; James, 2010; Kontra, Goldin-Meadow & Beilock, 2012; Gerson, Beckering & Hunnius, 2014). In fact, children succeed in solving many problems grounded in the physical world well before they can succeed with abstract, symbolic forms of parallel problems (Bruner, Olver & Greenfield, 1966; Piaget, 1953). These findings suggest that acting on, or manipulating, objects is a powerful way for children to learn new ideas.

Gestures -- a special category of action – can represent information, engage the motor system, and reference external representations in an instructional context, but unlike actions-on-objects, gestures are representational and do not create lasting change in the external environment. Here, we directly compare gestures with actions-on-objects in a linear measurement lesson with first grade children to investigate whether these different kinds of actions might differentially affect children's understanding of spatial units of measure. This foundational math concept is one that many children struggle with throughout elementary school and even middle school (Lindquist & Kouba, 1989). While traditional classroom instruction activities are largely ineffective, there is some recent work showing that giving children instruction with actions on manipulatives and self-discovered feedback, or 'disconfirming evidence', can improve learning outcomes (Kwon, Ping, Congdon & Levine, 2016).

In general, previous research identifies both benefits and drawbacks of learning through action in math contexts. Using manipulatives, objects designed to represent abstract math concepts in a tangible, physical way is one of the most common ways that action-based learning is instantiated in elementary school math lessons. For example, young children may learn to add using blocks or other sets of small objects before they are able to add Arabic numerals (e.g., Huttenlocher, Jordan, & Levine, 1994). Acting with manipulatives allows children to offload cognition onto the environment and encourages the formation of useful conceptual metaphors (Manches & O'Malley, 2012). It also directs attention to the relevant components of a complex problem (Mix, 2010), and engages young learners with limited attention spans (Peterson and McNeil, 2008). Yet some recent research cautions against action-based learning, citing instances where children may become distracted by irrelevant components of the manipulatives such as color or texture (Peterson, 2008), or may see the learned actions as relevant only to a specific set of objects rather than to a broader mathematical principle (e.g., Uttal, Scudder & DeLoache, 1997; DeLoache, 2000; Kaminski, Sloutsky & Heckler, 2009).

Gestures differ from actions on manipulatives in that they do not require children to interact directly with physical objects and do not result in changes in the location or orientation of these objects. Importantly, research shows that asking learners to gesture can promote learning and insight across a variety of academic domains including algebra, chemistry, geometry and word learning (e.g. Wakefield & James, 2015; Macedonia, Müller, & Friederici, 2011; Ping & Goldin-Meadow, 2008; Valenzeno, Alibali, & Klatzky, 2003; Singer & Goldin-Meadow, 2005). Gesture may be a particularly effective way of helping children to focus on important relational structures or spatial features of a problem. Consistent with this possibility, children instructed in mathematical equivalence problems (e.g., $3 + 4 + 5 = \_ + 5$) learn more from a lesson with a gesture that emphasizes making the two sides of the equation equal than from verbal instruction alone (Singer & Goldin-Meadow, 2005; Cook, Mitchell & Goldin-Meadow, 2008).

Although both action and gesture can be used as powerful learning tools, there is an open question as to *who* can best take advantage of the properties each type of tool offers. In fact, the very features that differentiate gestures from actions (i.e. the fact that they are representational, do not interact with objects, and do not affect change on the external world) may make gestures difficult to understand for some learners. In other words, some children may have trouble either mapping the form of a gesture to its symbolic content, or perhaps keeping all the pieces of a problem active in their minds, which could render gesture ineffective as a teaching tool for that child. In support of this idea, we know that very young children can understand another person's actions, like demonstrating how to twist off the top of a container, before they can interpret a gesture that represents that action, like miming a twisting motion near the top of a container (Novack, Goldin-Meadow & Woodward, 2015). This evidence suggests that iconic gesture interpretation follows a later and more protracted developmental time span than action interpretation. Consequently, the meaning of iconic gestures may be unclear to some children, particularly if they are unfamiliar with the specific concept being represented by the gesture.

One study to date has directly compared action and gesture in a learning paradigm (Novack et al, 2014). In this study, the authors trained 3rd grade children to produce a problem-solving strategy with either an action, a concrete gesture or an abstract gesture in a mathematical equivalence task (i.e.: 3+7+2=__+2). While children in all groups performed similarly on a post-test, children in both of the gesture conditions performed better on a near-transfer task, and children in the abstract gesture condition performed best on a far-transfer task. The intriguing findings suggest that the features that differentiate gesture from action may be particularly helpful for giving children an abstract, flexible, and generalizable understanding of an idea. Yet this study raises the question of whether abstract gesture is more helpful than actions-on-objects for all students, even if they have a very rudimentary understanding of a concept.

To address this open question, we gave children a lesson with either action or gesture on a linear measurement task. Linear measurement provides a fitting case study for two reasons. First, measurement is a foundational mathematical concept that young children consistently struggle with, making it an important focus of research on math learning. In fact, measurement remains the only domain of mathematics in which elementary school children in the US consistently perform lower than the international average (TIMSS, 2011; Foy, Arora & Stanco, 2013). Second, children consistently and robustly make one of two interesting errors on shifted-object linear measurement problems (problems where the to-be-measured object is not aligned with the zero-point on the ruler). These errors, described in more detail below, allow us to directly ask how action and gesture training might interact with learning outcomes in two groups of children that are at distinctly different levels of conceptual understanding.

In a hatch-mark counting error, children count the hatch mark lines on the part of the ruler that is aligned with the object being measured instead of the intervals of space that fall between an object's left-most and right-most edges. In a read-off error, children simply read off the number on the ruler that aligns with the rightmost edge of the object no matter where the object's left most edge starts on the ruler. Notably, both errors provide the *correct* answer on typical measurement problems where the object-to-be-measured is aligned with the zero point of the ruler (e.g. Blume, Galindo, & Walcott, 2007; Kamii & Clark, 1997; Lehrer, Jenkins, & Osana, 1998; Solomon, Vasilyeva, Huttenlocher, & Levine, 2015). There is reason to believe that children who primarily use the read-off strategy on shifted-object problems are more behind in their understanding of linear measurement than those who use the hatch-mark strategy. For one, the hatch mark strategy at least reflects knowledge that measurement involves counting some kind of unit. Second, the read-off strategy generally negatively correlates with both age and socio-economic status (Solomon et al., 2015; Levine et al., 2009). Third, we know from previous work that some students switch their strategy from read-off to hatch-mark counting after instruction, but the reverse pattern is never observed (Kwon, Ping, Congdon & Levine, 2016). Taken together, these pieces of evidence suggest that children who use the read-off strategy at pre-test have lower conceptual knowledge of linear measurement than those who use the counting hatch mark strategy at pre-test.

In the current study, we begin by assessing first grade children's measurement pre-test strategies to determine whether they primarily use a hatch-mark strategy or a read-off strategy on shifted-object measurement problems. We then explore whether children at these two different levels of conceptual understanding benefit differentially from a short lesson that is accompanied by either an action (moving discrete plastic unit chips) or a gesture (counting units with a *thumb-and-forefinger* pinching gesture). Lastly, we look for any evidence that action or gesture instruction differentially promote generalization to other unit-based tasks.

Method

Subjects: 122, 1st grade students (60 females; 62 males; mean age at test: 7.13 years) were recruited and tested at a Chicago area private school. Children whose parents signed a consent form participated in three one-on-one sessions across two weeks (Session I; Session II; Session III).

Procedure: At each of the three visits, all children received a 14 question multiple-choice paper and pencil test (see Figure 1 for a sample question). Trials were pseudo-randomized into three versions of the task, counterbalanced across visit so each child received each version only once across the three sessions. In the first four test items, the crayon image was aligned with the "0" point on the ruler ("unshifted problems"). In the 10 subsequent test items the crayons were shifted to different points on the ruler ("shifted problems"). All crayons started and ended at a whole unit. The four answer choices reflected the correct answer, a read-off strategy answer, a hatch-mark strategy answer, and a fourth random choice that did not match any of the other three strategy-related options. This multiple-choice test was the main outcome of interest.

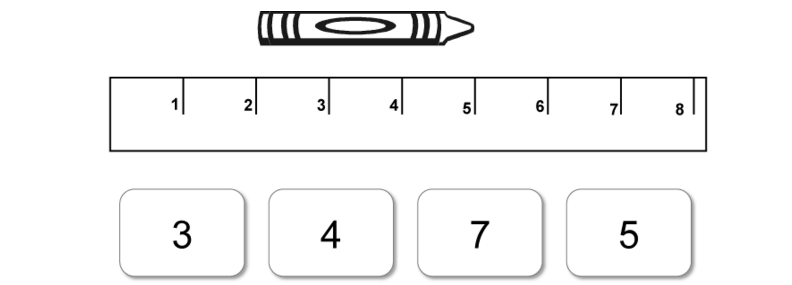

Figure 1. A sample "shifted item" from the multiple choice crayon measurement task.

Because we were interested in the effect of training, we excluded children who already understood how to correctly answer these questions. Children who answered 6 or more of the 10 "shifted object" multiple-choice problems correctly on this task in Session 1 were excluded from the study and did not participate in Session II or III. This criterion was based on probability values of the binomial distribution. Answering 6 or more multiple-choice questions correctly on a task with 4 options is significantly above chance ($p < .01$). Accordingly, 27 children were excluded at Session 1, leaving 95 children in the final sample.

Session I: In addition to the multiple-choice crayon task (Pre-Test), children in the first session also received a set of tasks that were intended to assess their understanding of linear units and measurement tools more broadly. The first task required children to "draw a picture of a ruler that is 8-units long" on a blank piece of paper. The second task required children to examine a set of four computer drawn images of rulers, which either had equal or unequal spaces and numbers or no numbers. The experimenter asked the child if each ruler, in turn, was a "useful ruler" and the child was asked to explain their answer. The purpose of this task was to assess whether or not children believe that numbers and/or evenly spaced units are crucial components of a useful measuring tool.

In the third task, participants were asked to "color a unit" on a picture of a blank ruler. The purpose of this task was to directly assess understanding of the word "unit". The fourth task was a perimeter measurement task in which children were presented with two hatch-marked shapes and asked, "How many units would it take to go all the way around the outside edge of the shape?" Finally, the children concluded the first session with a number line task in which they were asked to place 6 different numbers on a number line that ran from 0 to 100, and 6 numbers on a number line that ran from 0 to 1000 (e.g., Siegler & Opfer, 2003). Half of the participants received the 0-100 number line first, and half received the 0-1000 number line first. Children did not receive any experimenter feedback on any of the tasks in Session I.

Session II: In Session II, which took place at a convenient time between one and seven days after the first session (mean delay = 3.33 days, SD = 1.67 days), children were pseudo-randomly assigned to one of four training conditions, counterbalanced by both the gender of the child and their dominant answer strategy on pre-test trials (hatch-mark counting strategy or read-off strategy). The four training conditions were: unshifted unit, shifted unit, unshifted gesture and shifted gesture (Figure 2). In all four conditions, an experimenter showed children how to measure a colorful wooden stick with a 9-unit paper ruler and either discrete plastic unit chips or a thumb and forefinger gesture. After the experimenter placed the stick above the ruler, participants were asked to guess how long the stick was. Then they were told to check their answer with either the unit chips or the "gesture unit". The unit chip instruction conditions were modeled after a linear measurement training study with 2nd grade students (Kwon, Ping, Congdon & Levine, 2016). The gesture training conditions were developed to represent the same concepts as the unit chip instruction using a gesture that is spontaneously produced when people talk about the size or length of small objects. In addition, there is some previous work showing that children as young as 2.5 years old can map this gesture to the size of an object (Novack, Filippi, Goldin-Meadow & Woodward, 2016). The child was corrected if he or she performed any of the movements incorrectly. Finally, the experimenter performed the movements while counting aloud to ensure each child understood the correct answer for each trial. This procedure was repeated 8 times with different colorful sticks that varied in length. Following training, children received a second version of the multiple-choice crayon measurement task (Post Test).

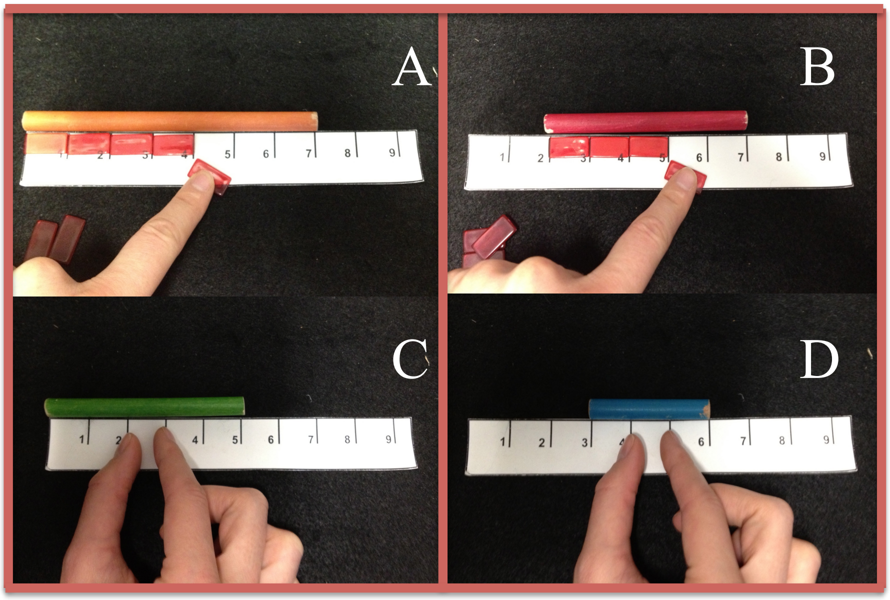

Figure 2. A photograph of each of the four training conditions. A – Unshifted unit chip training condition; B – Shifted unit chip training; C – Unshifted gesture training; D – Shifted gesture training

Session III: One week after the second session (mean delay = 7.05 days, SD = 1.77 days), each participant received the multiple-choice crayon task a third time followed by a series of generalization tasks aimed at characterizing each child's ability to transfer his or her understanding of the shifted ruler task to other tasks tapping the concept of spatial unit. In one generalization task, children were asked to measure three real-world objects with a "broken" ruler, which started with a jagged edge at the 2.5- or 3.5-unit mark. The purpose of this task was to see whether or not children would try to align the object with the broken edge of the ruler, or whether they would use the middle of the ruler to give an answer that reflected either a read-off strategy, a hatch-mark counting strategy, or the correct strategy. On a second generalization task, we asked children to use two paper clips to measure how many "paper clip units" it would take to measure a line. In addition, to assess growth across the training session, each child was again asked to color a unit on a picture of a blank ruler, complete the number line task, and find the perimeter of two novel but similar test items to those used on the pre-training perimeter task.

Results

As expected, performance on the four unshifted items on the multiple-choice crayon test was virtually perfect at all three time points for all participants (*M*=3.93, *SD*=0.39 at pre-test; *M*=3.92, *SD*=0.38 at immediate posttest; *M*=3.95, *SD*=0.37 at the 1-week follow-up). As such, we only carried out formal analyses on children's performance on the ten shifted-item questions.

Main Outcome: The first analysis examined whether children's starting strategy (read-off vs. hatch-mark) interacted with the efficacy of the two different training conditions. Because individual children tended to get most of the problems right or most of the problems wrong at each of the three sessions, the data were non-normally distributed (Figure 3). (Note that children who got most of the problems correct at pretest are not represented in this figure because they were excluded from the training). Accordingly, instead of performing an analysis on average scores, the data were fit with a mixed effects binomial logistic regression model that predicted correct performance on each shifted-object test item. All analyses were performed using R (R Development Core Team, 2008). In the first model, participant was a random effect, and session (pre-test, posttest, follow-up), training condition (unshifted unit, shifted unit, unshifted gesture, shifted gesture), starting strategy (read-off or hatch-mark), gender, and the two-way interaction between starting strategy and condition were used as fixed effects.

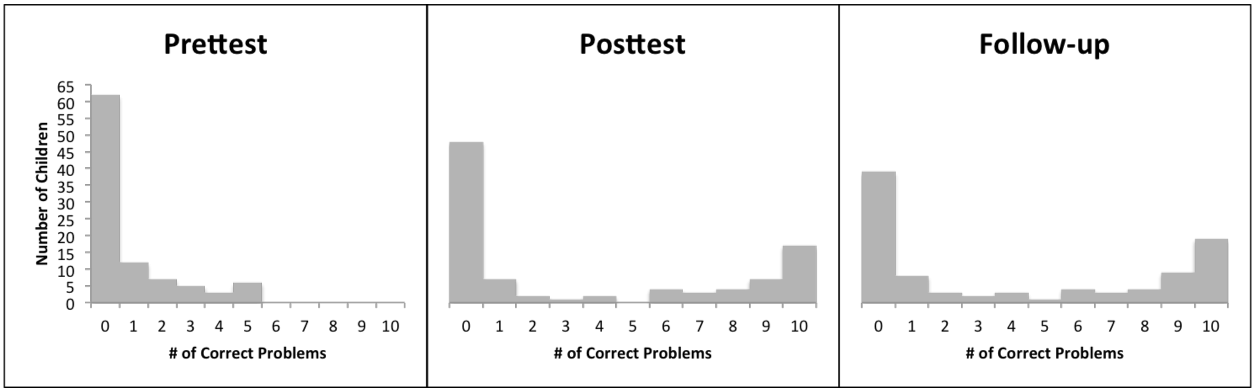

Figure 3. Each panel shows the non-normal distribution of scores on shifted-problem test trials for all children in the sample, regardless of training condition, at each of the three time points.

An analysis of variance of the factors in the regression model showed a main effect of training condition ($X^2 = 46.80$, $p<.001$) and a main effect of starting strategy ($X^2 = 35.78$, $p<.001$). Importantly, these main effects were qualified by a marginal, though theoretically significant, condition x starting strategy interaction ($X^2 = 7.62$, $p=.054$). There was also a strongly significant main effect of session ($X^2 = 335.31$, $p<.0001$) and a marginal effect of gender ($X^2 = 3.54$, $p=.060$). To better explore these results, particularly the interaction between starting strategy and training condition, we built two separate models; one for children who predominantly used the hatch-mark counting strategy at pre-test, and one for those who began by using the read-off strategy. Means and standard errors of the means for the two groups at each session are displayed in Figure 4.

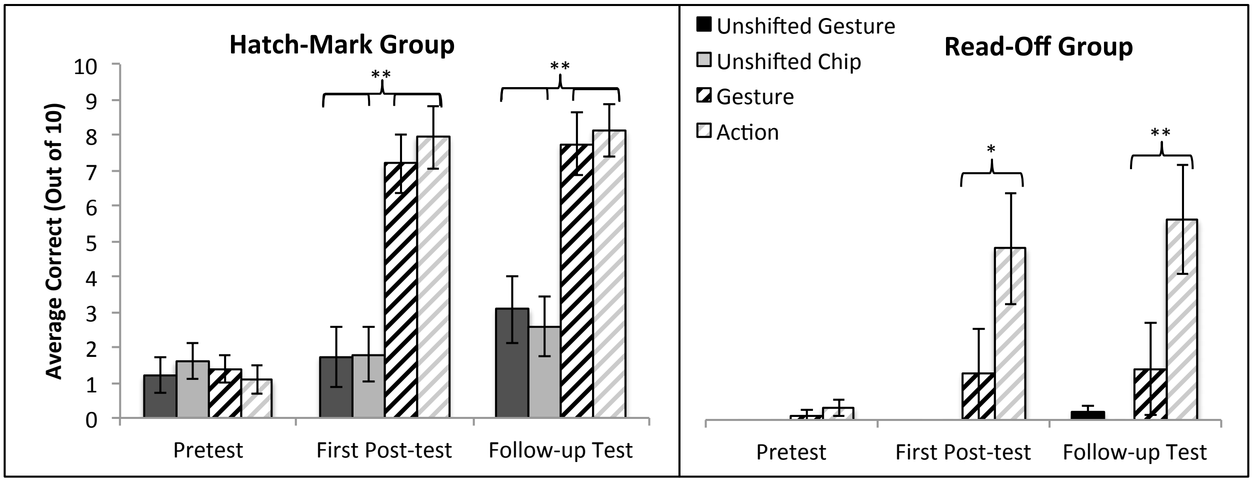

Figure 4. Average performance by starting strategy and training condition across the three sessions. Bars represent +/- 1 standard error of the mean when the data are aggregated by participant[1].

For the children who began the study by counting hatch-marks, an analysis of variance of a binomial linear regression model with subject as a random effect and training condition, session, gender and the interaction between condition and session as fixed effects revealed a main effect of condition ($X^2 = 29.74$, $p<.001$), whereby the shifted gesture and shifted unit training conditions were more effective than the two unshifted conditions. There was also a main effect of session ($X^2 = 170.51$, $p<.001$) whereby responses at posttest and follow-up were significantly more likely to be correct than responses at pre-test. These main effects were qualified by a significant condition by session interaction ($X^2 = 102.95$, $p<.001$), driven by the fact that there was no effect of training condition at pre-test but significant differences at posttest and follow-up. There was no effect of gender in this model ($X^2 = 0.46$, $p=.50$).

[1] Note that the graphs present mean performance of participants, but the data are analyzed using a binomial model with a random effect of participant. The model is sensitive to the largely bimodal nature of the data (i.e., some children failed to learn at all, and those who did learn got most or all of the problems correct on a given test) and thus more accurately reflects children's performance than the means presented in the graph.

For the read-off strategy participants, analyses revealed a main effect of condition (X^2 = 13.30, p<.01), whereby children in the shifted unit condition performed better overall than each of the other three conditions. Again, there was an expected main effect of session (X^2 = 41.74, p<.001), with a higher chance of correct responses at posttest and follow-up than at pre-test. Finally, in this model, there was a main effect of gender (X^2 = 4.61, p<.05), whereby girls outperformed boys. This gender effect was likely driving the marginal effect of gender reported in the original omnibus model, and future work should investigate why males in the read-off group were less receptive to instruction than females.

Strategy Analysis: Motivated by the low overall rates of learning in the read-off strategy group, we performed a descriptive analysis of the kinds of errors children in both groups were making before and after training to ask whether some children were showing qualitative improvements that were not captured by our main outcome (Figure 5). This analysis showed that training led some children in the read-off group to switch their strategy to the more sophisticated, yet still incorrect, hatch mark strategy. While the strongest effect was observed in the most successful training condition, shifted unit training, there were some children who switched to the hatch-mark strategy after training in each of the other three conditions. By contrast, none of the children in the hatch-mark group switch to a read-off strategy after training in any instructional condition. The overall pattern further supports the original distinction between the two groups as being at different levels of understanding. In other words, these data suggest a progression in learning from the most rudimentary strategy (read-off) to a more sophisticated but incorrect strategy (hatch-mark counting), though our data are clear that the intermediate hatch-mark counting stage is not a necessary precursor to correct performance (as a few read-off children did jump right to a correct strategy after our brief training, particularly the shifted unit training).

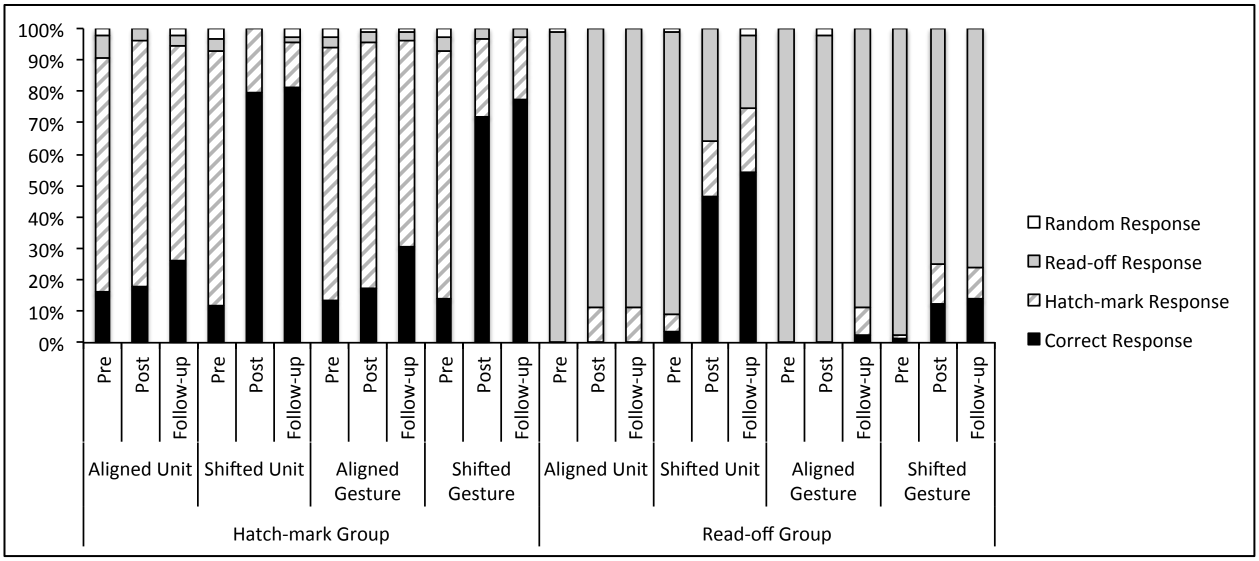

Figure 5. Each trial was coded based on the child's response. This chart, broken down by starting strategy (hatch-mark and read-off), training condition (aligned unit, shifted unit, aligned gesture, shifted gesture), and time point (pre-test, posttest, follow-up) shows the distribution of strategy use across the entire study. While some children in the read-off group switched to the hatch-mark strategy after training, children in the hatch-mark group never switched to the read-off strategy.

Pre-Training and Generalization Tasks: For the two tasks administered only at pre-test, a set of simple linear regression models used performance on the pre-test task to predict learning score on the main ruler and crayon outcome (post-test score minus pre-test score) after controlling for training condition. Children's ability to appropriately draw a ruler predicted their propensity to learn from subsequent training. This suggests that familiarity with the key features of a ruler may have been an important foundation for successful training (Table 1).

For the tasks administered at both Session I and Session III, a set of linear regression models used improvement on the main ruler and crayon outcome (post-test score minus pre-test score); training condition; and the interaction between condition and improvement to predict the change in performance on the generalization task from Session I to Session III. For the last set of transfer tasks, which were administered only at Session III, the same fixed effects were used to predict performance on each transfer task. For each of these models, an analysis of variance of the regression model gives an estimate of the main effects and interaction terms, reported here as X^2 values (Tables 2 and 3).

Somewhat surprisingly, these analyses of each generalizations task revealed very little evidence that training condition differentially affected transfer. The results show that while a few of the tasks (Draw a Ruler, Perimeter, Color a Unit, Broken Ruler) were at least marginally related to learning outcomes in general, there was no evidence of any significant effects of training condition or any interactions between learning and training condition.

A.	Task	Group Means (SD)	Group Comparison	All Participants Relation to Learning
	Draw A Ruler	RO = 0.51 (.66) HM = 0.78 (.86)	$p = 0.11^{\dagger}$	β= 0.95 (p = 0.031)*
	Useful Ruler	RO = 0.79 (.98) HM = 0.98 (.98)	p= 0.37	β= 0.17 (p = 0.42)

Table 1. This shows children's performance on the tasks that were administered only at pre-test.

B.	Task	Group Means of Δ Score (SD)	Group Comparison	All Participants Relation to Learning
	Perimeter	RO = 0.29 (.67) HM = 0.57 (.81)	$p = 0.086^{\dagger}$	$X^2_{Learning}= 1.93$ ($p = 0.06$)† $X^2_{Condition}= 2.83$ ($p = 0.17$) $X^2_{Interaction}= 1.58$ ($p = 0.41$)
	Number Line (1 – 100)	RO = 0.06 (.54) HM = 0.10 (.61)	$p= 0.72$	$X^2_{Learning}= 0.06$ ($p = 0.67$) $X^2_{Condition}= 0.96$ ($p = 0.42$) $X^2_{Interaction}= 0.33$ ($p = 0.81$)
	Number Line (1 – 1000)	RO = 0.03 (0.38) HM = -0.02 (0.47)	$p = 0.63$	$X^2_{Learning}= 0.06$ ($p = 0.59$) $X^2_{Condition}= 0.98$ ($p = 0.18$) $X^2_{Interaction}= 0.11$ ($p = 0.90$)
	Color a Unit	RO = 0.11 (0.47) HM = 0.27 (0.52)	$p = 0.16$	$X^2_{Learning}= 1.48$ ($p = 0.01$)** $X^2_{Condition}= 1.37$ ($p = 0.13$) $X^2_{Interaction}= 0.95$ ($p = 0.27$)

Table 2. This shows children's performance on the tasks that were administered at both pre-test and follow-up. Improvement on the perimeter task and on the color a unit task were correlated with learning, even after controlling for condition.

C.	Task	Group Means (SD)	Group Comparison	All Participants Relation to Learning
	Broken Ruler	RO = 0.63 (.84) HM = 1.28 (.90)	$p = 0.00074$**	$X^2_{Learning}= 2.44$ ($p = 0.09$)† $X^2_{Condition}= 5.01$ ($p = 0.12$) $X^2_{Interaction}= 2.51$ ($p = 0.40$)
	Paperclip Task	RO = 0.33 (.48) HM = 0.50 (.50)	$p= 0.124$	$X^2_{Learning}= 0.13$ ($p = 0.48$) $X^2_{Condition}= 0.76$ ($p = 0.40$) $X^2_{Interaction}= 0.005$ ($p = 0.99$)

Table 3. This shows children's performance on the tasks that were administered only at follow-up. Children in the read-off group performed worse than children in the hatch-mark group on the broken ruler task, and performance on this task marginally correlated with learning outcomes overall.

Spatial Language and Gesture: In order to ensure that improvement on the main task was driven by an improved understanding of spatial units or spatial extent, rather than by unintentionally teaching children some other pattern, trick, or algorithm, we coded the speech and gesture of a subset of the participants (66% of read-off; 75% of hatch-mark)[2] as they explained their answers on the final two trials of the crayon task. One research assistant, blind to experimental condition, transcribed the child's spoken responses and transcribed the child's gestures by noting the hand shape, motion, and location of each hand movements. That same research assistant coded the transcriptions for evidence of spatial language and gesture. A second trained coder coded a randomly selected 30% of the trials for reliability purposes (agreement was very high = 95%, un weighted Kappa = .83). In speech, a child was coded as using spatial language if she used the word "unit(s)", or "space(s)" or appropriately described the spatial extent of the crayon, even in the absence of a unit-like word (e.g. "It started at the 3 and went to the 8 so I knew it was 5 long"). Simply counting aloud (e.g. "1, 2, 3"), a common verbal response, was not considered sufficient evidence of spatial language. Children's gestures were coded as correct, unit-based gestures if they pointed to the spaces (but not the lines), traced the exact extent of the crayon, or framed the extent of the crayon with either hands or fingers. For example, if a child counted aloud, "1, 2, 3", while pointing to the spaces on the ruler, she would be coded as having spatial gestures, but not spatial language.

Figure 6 displays the proportion of children who showed evidence of spatial thinking in their speech or gesture. As expected, we found evidence of a general increase in spatial language and spatial gesture after training. In a set of two binomial linear regression models, we used training condition to predict 1) spatial language and 2) spatial gesture use at the follow-up session after controlling for use at pre-test for children in the hatch mark group. Analyses showed that there were no significant differences by condition for spatial talk (X^2=1.28, p=0.74),

[2] Videos that are not included in this analysis were corrupted due to experimenter error and could not be transcribed or coded. The sample of videos lost is believed to be random, allowing us to draw conclusions for the remaining, representative videos.

but for spatial gesture, children in the shifted chip condition increased their spatial gestures significantly more than those in the aligned chip condition (B=2.84, p=0.047) and marginally more than those in the aligned gesture condition (B=1.91, p=0.10). All other comparisons were non-significant. The extreme nature of the data in the read-off group (in which no children produced spatial language or spatial gesture at either the pretest or the follow-up in three of the four training conditions) precluded us from performing any statistical analysis. Qualitatively, however, we observe an increase in both spatial language and spatial gesture in the shifted chip condition, the only effective training condition.

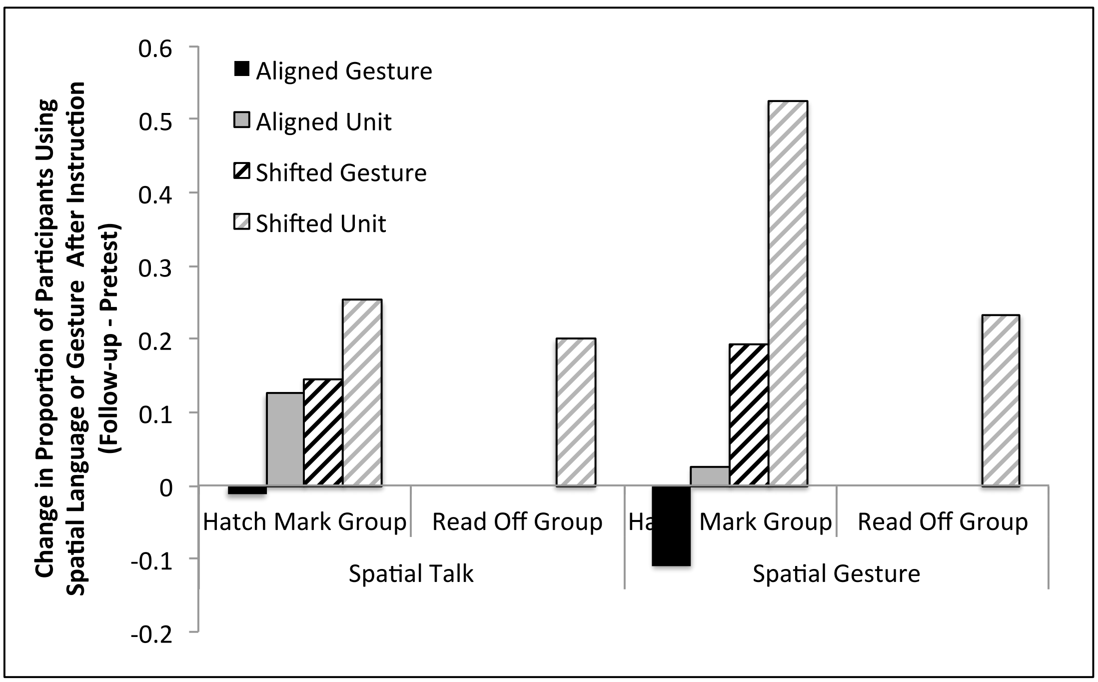

Figure 6. This figure shows the change in the proportion of subjects who reveal correct unit-based strategies in either speech or gesture after training.

Discussion

The results of this study add to a growing literature that explores how the qualitative differences between actions and gestures, two similar though not identical types of movement, contribute to learning and cognition. Specifically, this study is the first to show that it is critical to consider individual differences in children's conceptual understanding of a given problem before implementing gesture-based instruction. Overall, the results suggest that the abstract properties of gesture, the very properties that may make it so powerful for generalization in some scenarios, can also make it inaccessible to some learners in problem-solving contexts.

From a practical perspective, this study emphasizes the necessity of providing children with linear measurement instruction involving shifted-object problem types. As reported in previous work, these types of problems reveal children's misconceptions about measurement in a way that unshifted problems do not, and also support learning in a way that shifted problems do not (Solomon et al, 2015; Kwon et al., 2016). They discourage the use of simple procedural strategies and encourage the development of a more flexible and conceptually rich understanding of measurement. The kind of rapid and robust learning observed in the current study can best be explained by the idea of disconfirming evidence, or prediction error (e.g., Rescorla & Wagner, 1972; Ramscar, Dye, Poppick, O'Donnell-McCarthy, 2011). Encouraging a child to make a guess and then allowing them to discover that their answer is consistently wrong can powerfully drive conceptual change and adoption of new strategies by causing learners to question their current strategies and assumptions (Siegler & Svetina, 2006).

Although shifted-object training was necessary for improvement in the current study, it was not sufficient. Instead, we found an interaction between a child's starting level of conceptual knowledge and the effectiveness of gesture- and action-based instruction. First, the results suggest that representational gesture is more abstract than actions-on-objects, and that this distinction has context-dependent implications for cognition and learning. Second, the results add to existing literature claiming that the read-off strategy is a more rudimentary procedural strategy than counting hatch marks (Solomon, Vasilyeva, Huttenlocher & Levine, 2015; Kwon, Ping, Congdon & Levine, 2016). Not only did children in the hatch mark group learn more from training overall, but we found that after training, some children in the read-off group switched to the hatch-mark counting strategy, and we never observed the opposite change in error type. Taken together, these results demonstrate that in the context of linear measurement, children who have a more rudimentary understanding require more concrete, tangible tools, while those with a more advanced albeit still erroneous understanding can learn from either action or gesture.

So which property or properties of gesture are driving this dissociative effect? There are several intriguing possibilities that emerge from examining the differences between the two kinds of movement. The first possibility is that the meaning of the gesture itself was opaque to the children in the read-off strategy group, and the meaning of the plastic unit chip was more obvious, or perhaps more familiar to those same students. Gesture understanding and interpretation does follow a more protracted developmental time course than does action understanding (Novack, Goldin-Meadow & Woodward, 2015). Indeed, even adults require certain contextual cues to consider something gesture and to be able to interpret it appropriately (Novack, Wakefield & Goldin-Meadow, 2015). Therefore, the failure of the children in the read-off group to learn from gesture is potentially reflective of the fact that they did not have the appropriate preexisting conceptual basis upon which to map the iconic gesture. Recall that students who use the read-off strategy have not demonstrated *any* of the conceptual bases for measuring spatial extent with units. In lacking this conceptual foundation, they may have been unable to glean any novel insights from the iconic "pinching" size gesture. This may not be a problem of gesture, per se, but perhaps is a more general phenomenon that learners must have a conceptual basis upon which to map any symbolic or representational learning tool. Indeed, a prior study (Kwon, Ping, Congdon & Levine, 2016) showed that in a group of slightly older second-grade children who were persisting in using the read-off strategy on a linear measurement pretest, even training with unit chips was unsuccessful. This finding suggests that even the unit chips, which are, themselves representational of a unit, may not be immediately apparent to all learners, particularly for those children who have continued to use a rudimentary strategy as they progress through elementary school.

The second possibility, which is not mutually exclusive from the first, is that gesture is cognitively demanding in this particular context because it is iterative and does not leave a trace. In other words, it is possible that even if children in the read-off group understood that the gesture was meant to represent a small length or unit of measure, they were subsequently overwhelmed by the pragmatics of the problem; unable to keep in mind the gesture instructions, what they were supposed to be counting with the gesture, what the gesture represented, and what the final numerical answer mapped onto. In contrast, the plastic unit chips are manipulable, countable, objects that create a lasting trace in the form of a set that can be counted. Thus, it's possible that children in the read-off group, who had to make a larger conceptual leap than their peers who began with the hatch-mark strategy, found the gesture counting task unduly taxing for their working memories. Decreasing working memory load has been offered before as a potential benefit of using real-world manipulatives to offload some cognitive processes (Manches & O'Malley, 2012). And while similar mechanisms have been suggested for gesture-based instruction (e.g., Morsella & Krauss, 2004; Ping & Goldin-Meadow, 2008; Cook, Yip & Goldin-Meadow, 2012), it is possible that some familiarity with the target concept is necessary to capitalize on that feature of gesture.

While we did find some evidence of transfer in the current study, action and gesture did not *differentially* predict rates of transfer as has been reported in previous work (Novack et al., 2014). There are several ways to interpret this finding. First, there is existing research on how difficult it is for learners to apply newly acquired knowledge in novel contexts (e.g., Catrambone & Holyoak, 1989; Mix, 2010). The training we provided here was not only brief, but required children to switch between a real-world, 3D training scenario and a 2D posttest even before we assessed “transfer”. Such a dimensional shift between training and testing could push the limits of flexibility in children’s representational systems (Barr, 2010). Furthermore, the low rates of transfer on the farther generalization tasks would suggest that perhaps the tasks were not appropriately calibrated to capture meaningful differences by training condition.

The second possibility, however, is that for linear measurement, it is the learning and insight process itself that matters for success on transfer tasks and not the manner in which the task was learned. Though there are many features that differentiate the current study from that of Novack et al., one notable difference is the type of mathematics problem being trained (linear measurement vs. mathematical equivalence). Perhaps gesture, a more abstract tool, is better suited for learning and transfer in a more abstract mathematical domain like algebraic equivalence and equation balancing. By contrast, linear measurement is a spatial problem, and it may be the case that either action-based or gesture-based instruction is sufficient for gaining insight and mastering this particular, highly spatial concept. Understanding these complicated interactions between content domain and effective instruction techniques is a computationally overwhelming problem (Koedinger, Booth & Klahr, 2014) and there is much work to be done to discover guiding principles of when and how to implement movement-based instruction.

The current study provides a promising beginning towards this ambitious goal by highlighting two features of an instructional context that must be considered when teaching children new ideas through hand movements. Though there is some compelling research showing that gesture, in particular, can promote learning, generalization, and retention across a number of different domains and age groups (e.g., Wakefield & James, 2015; Macedonia, Müller, & Friederici, 2011; Ping & Goldin-Meadow, 2008; Valenzeno, Alibali, & Klatzky, 2003; Singer & Goldin-Meadow, 2005; Cook, Mitchell & Goldin-Meadow, 2008; Novack, Congdon, Hemani-Lopez & Goldin-Meadow, 2014; Cook, Duffy & Fenn, 2013; Levine, Goldin-Meadow, Carlson & Hemani-Lopez, 2016), the current findings underscore the need to consider the learner's level of conceptual understanding prior to instruction, as well as the nature of the target concept, before implementing gesture-based instruction. The very same properties of gesture that differentiate it from action and facilitate long-lasting and flexible conceptual change in certain problem solving contexts may make it ineffective in other problem-solving contexts and altogether inaccessible to certain learners, who benefit more from actions-on-objects.

CHAPTER 3 – TEACHING MEASUREMENT FROM ACTION TO ABSTRACTION

General Overview

The findings from Chapter 2 demonstrate that when lower prior knowledge children encountered a difficult linear measurement task, actions-on-objects were far more effective than gesture-based instruction. By contrast, children with higher prior knowledge were able to learn from either action or gesture. Importantly, the performance within this higher prior knowledge group shows that either type of movement *can* be very effective under certain circumstances. Given these intriguing results, Chapters 3 and 4 explore this interaction between prior knowledge and instruction type and ask *why* the children in the read-off strategy group struggled to learn from the gesture-based instruction while their hatch-mark counting peers were successful.

In the discussion in Chapter 2, two possibilities were raised to explain this phenomenon. The first is that the children with lower prior knowledge simply do not understand the representational content of the gesture. This hypothesis is consistent with the idea that children who use the read-off strategy have a relatively sparse conceptual understanding of linear measurement. In other words, the thumb-and-forefinger gesture, intended to represent a unit of spatial extent, may have been seen as a meaningless movement for children with very low prior conceptual knowledge. Children and even adults have difficulty interpreting gestures as meaningful hand movements if they are not given enough context in which to interpret a movement (Wakefield, Novack & Goldin-Meadow, 2016; Novack, Wakefield & Goldin-Meadow, 2015). The second possibility is that children are able to interpret the gesture as a meaningful representation at first, but they subsequently have trouble tracking the transient, iterative, abstract movement throughout the training session. Under this hypothesis, children with lower conceptual knowledge might be experiencing higher demands on their working memory, and could become confused when trying to understand how counting the abstract gesture maps onto a correct final answer. In the remaining chapters, I explore each of these possibilities in turn to try to better understand how and why gesture might prove challenging for low prior knowledge learners.

Introduction

In Experiment 2, I ask whether providing children with the action-based instruction *and* the gesture-based instruction in a single training session might improve learning outcomes by helping children in the read-off group to understand the action-based referent of the gesture. In this four-condition design, one group of children received instruction with only gesture, one group received only action, one received action followed by gesture, and one received gesture followed by action. This design allows me to a) replicate findings from Experiment 1, b) ask whether it is helpful to provide children with multiple representations of units within a single training session and c) ask whether the order of those representations matters for learning outcomes.

The first possibility is that lower prior knowledge learners will learn best when they receive action-based instruction *followed by* gesture-based instruction. This hypothesis is loosely based on the Piagetian idea that across ontogenetic development, children learn best when they begin with more concrete representations and experiences before they move onto more symbolic ones (Piaget, 1953). More recently, researchers coined the term "concreteness fading" to refer to the practice of transitioning from concrete representations to symbolic ones over instructional time (Goldstone & Son, 2005; McNeil & Fyfe, 2012; Fyfe, McNeil, Son & Goldstone, 2014). In this work, researchers show that across different domains and different age groups, introducing learners to a concrete or more real-world example (e.g., a physical balancing scale) before introducing them to a more abstract schematic (e.g., a balanced mathematical equation) is better for learning outcomes than introducing either representation alone, or presenting the representations in the opposite order.

An alternative possibility is that the order of presentation of the action-based instruction and the gesture-based instruction will not matter, but that there will be a main effect of the number of instruction types. For example, it might be useful for children with lower prior knowledge to see multiple exemplars of a 'unit' within one lesson. This hypothesis is inspired by principles of category learning and verb learning in the domain of language. Multiple exemplar training (MET) shows that giving children more than one exemplar of a category helps them to extract the relevant characteristics of that category, remember the rules better, and apply them to novel situations (e.g., Smiley & Huttonlocher, 1995; Gentner, 2003; Luciero, Becera & Valverde, 2007; Horst, Twomey & Ranson, 2013). In the domain of mathematics, there is evidence that providing young children with varied examples of perceptually distinct triangles can better help them to extract the defining features of triangles (Smith et al., 2014).

On the other hand, it is possible that children with lower prior knowledge will do poorly when given multiple representations, particularly if they are unable to make the appropriate conceptual links between the two movement types or if they become overwhelmed by too much information in a short lesson. For example, in order to extract the benefits of multiple exemplars, it could be necessary to present action and gesture simultaneously, or to interweave them, rather than present them sequentially as in the current design. Or, based on some evidence from the language learning literature, children may do best when only presented with one representation that they can focus on mastering. For example, infants learning a new verb learned and extended that learning best when only exposed to one exemplar as compared to four (Maguire, Hirsh-Pasek, Golinkoff & Brandone, 2008), and infants learning spatial prepositions learned best when shown the same exemplar multiple times, rather than different exemplars (Casasola, 2005). Both of these studies suggest a 'less is more' hypothesis in which one good exemplar is better than many different exemplars.

The third possibility is that demonstrating the action-on-objects (i.e., using the unit chips) will not be sufficient to give the read-off children the necessary context to interpret and learn from the gesture. Under this hypothesis, I would expect to find that any of the three instructional conditions that incorporate the gesture are not be as effective as the instructional condition that relies on action alone. Or perhaps that gesture followed by action-on-objects would be successful primarily because of a recency effect – that is, that the training will conclude with the action, allowing the children to ignore any confusion stemming from the gesture instruction. Based on the findings in Chapter 2, I predicted that across all four conditions, children in the higher prior knowledge group would improve after instruction.

Method

*Subjects***.** 117, 1st grade students (68 females; 49 males; mean age at test: 6.97 years, SD = 0.37 years) were recruited and tested at several Chicago area schools. In contrast to the sample reported in Chapter 2, children in this sample were from a broader range of socio-economic backgrounds. Based on a categorical income questionnaire, children in the current study reported ranges from 1 (lowest possible score) to 6 (highest possible score). In the previous study, the range was 4-6. Overall though, the average score reported in the current sample was still quite high (5.38 out of 6, SD = 1.34), and SES was a non-significant predictor in all models and thus is not included in any final analyses. Children whose parents signed a consent form participated in two one-on-one sessions one week apart in a quiet area of their school (Session 1 and Session II).

Session I. To assess pre-test strategy, children were given a 14-question multiple-choice paper and pencil test (see Figure 1, Chapter 2). As described in Chapter 2, the first four test items were of a crayon that was aligned with the "0" point on the ruler ("unshifted problems"). In the 10 subsequent test items the crayons were shifted to different points on the ruler ("shifted problems"). All crayons started and ended at a whole unit. The four answer choices reflected the correct answer, a read-off strategy answer, a hatch-mark strategy answer, and a fourth random choice that did not match any of the other three strategy-related options. This multiple-choice test was re-administered right after training, and is the main outcome of interest.

Based on performance on the pretest shifted crayon items, children were categorized into a particular strategy group if 6 or more of the 10 shifted-object items were answered in a way that was consistent with a single strategy. This criterion was based on probability values of the binomial distribution: on a task with 4 options, answering 6 out of 10 using a particular strategy means that that child is using that strategy more often than would be predicted by chance or random guessing ($p < .01$). By this metric, there were 12 children in the 'correct' group (N=9 males); 46 children in the 'hatch-mark' group' (N=19 males); 57 in the 'read-off' group (N=20 males); and 2 children (N=1 male) whose dominant strategy was 'random' or did not meet criteria for inclusion in one of the other groups. Children who were in the 'correct' group were excluded from further analyses given that the small sample size and limited room for improvement after training, and children in the 'random' group were excluded due to sample size. Four children were excluded for missing data on the follow-up session, and one child was excluded due to a language barrier. Thus, the final sample consisted of 98 children.

Immediately after completing the multiple-choice crayon task (Pre-Test), children in Session I received a set of three tasks that were intended to assess growth in other unit-based tasks following training. Given the generally low rates of transfer reported in Experiment 1, several new tasks were piloted and selected for the current study (Experiment 2). The first task was an image of a crayon with numbered circles below it instead of a ruler. This task was intended to test what children would do in a situation where there are no hatch marks to count. The second task involved asking children how many units long an object was, and then giving them an array of laminated unit chips to find the answer. The unit chips were either .75 inches or 1.5 inches long, and half of each length of unit was pink and half was yellow. The purpose of this task was to see whether participants would spontaneously select equal-sized units, or whether they would be distracted by the irrelevant feature (i.e. the color). To ensure that students were forced to contend with both of these dimensions, there were not enough chips provided of any single type to be able to fully measure the object. The third task was called "Going to the Store" and involved reading a short scenario to children in which a fictional character wants to take the shortest possible path to one of two stores. Participants were told to select the closer store by determining which was the shorter path, and were told to, "Use the ruler if you think it will help." Each participant was given two trials at pre-test and two trials at follow-up: the 'easy trials' had two straight paths, and the 'hard trials' had two zigzag paths.

After completing the baseline transfer tasks, children were randomly assigned to one of four between-subjects training conditions. Assignment was counter-balanced by children's dominant initial measurement strategy. The four training conditions were: Action only (N = 25;

N = 13 read-off); Gesture only (N = 25; N = 13 read-off); Action-then-Gesture (N = 24; N = 12 read-off) and Gesture-then-Action (N = 24; N = 12 read-off). Based on the results from Experiment 1 in which no children improved after training on unshifted objects, all training in Experiment 2 was performed with shifted objects (See Chapter 2, Figure 2 for images of the shifted-object training conditions). As in Experiment 1, all children received a total of 8 training trials with experimenter feedback. For the training conditions with two different types of movement instruction, action and gesture, children received 4 training trials of one type and 4 of the other for a total of 8. During the transition from one movement type to the other, the experimenter stated, “Now we are going to play the same game, but with a different kind of unit”, and then introduced the child to the new movement type. All other training procedures were identical to those reported in Experiment 1. Following training, children received a second version of the multiple-choice crayon measurement task (Posttest).

Session II. Approximately one week after the second session (mean delay = 7.05 days, SD = 0.48 days), each participant received a third version of the multiple-choice crayon task (Follow-Up) followed by a series of generalization tasks aimed at characterizing each child’s ability to transfer his or her understanding of the concept of a “unit”. In one generalization task, children were asked to measure three real-world objects with a “broken” ruler, which started with a jagged edge at the 2.5 or 3.5-unit mark. For the second generalization task, we asked children to find the perimeter of 4 different figures of varying difficulty (see Figure 7 for sample items). In addition, to assess growth across the training session, each child was again asked to do the numbered circles crayon measurement task, the color/size unit measuring task, and the “Going to the Store” task.

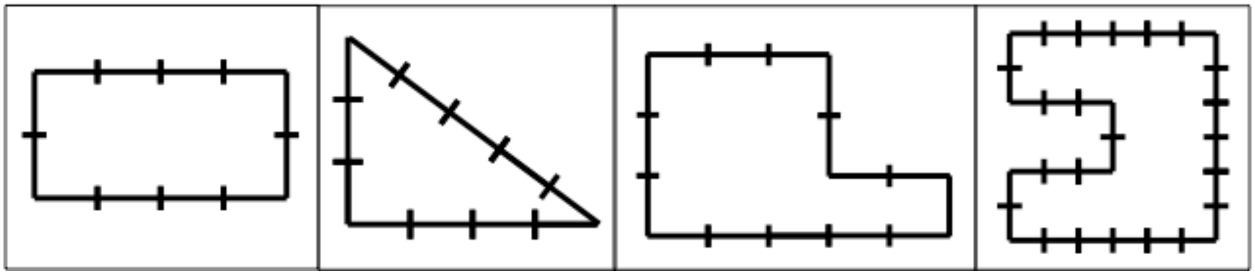

Figure 7. Sample perimeter test items. This task was administered only in Session II.

Results

As expected, performance on the four unshifted items on the multiple-choice crayon test was very high at all three time points for all participants (*M*=3.94, *SD*=0.49 at pre-test; *M*=3.81, *SD*=0.84 at immediate posttest; *M*=3.75, *SD*= 0.97 at the 1-week follow-up). As such, we only carried out formal analyses on children's performance on the ten shifted-item questions.

On the main outcome of interest, the crayon and ruler task, the data were non-normally distributed (children either got most problems right or wrong). Accordingly, the data were fit with a mixed effects binomial logistic regression model that predicted correct performance on each shifted-object test item. All analyses were performed using R (R Development Core Team, 2008). Based on a priori predictions about differences between the higher and lower prior knowledge groups, I built two separate models: one for children who predominantly used the hatch-mark counting strategy at pre-test, and one for those who began by using the read-off strategy. Means and standard errors of the means for the two groups at each session are displayed in Figure 8.

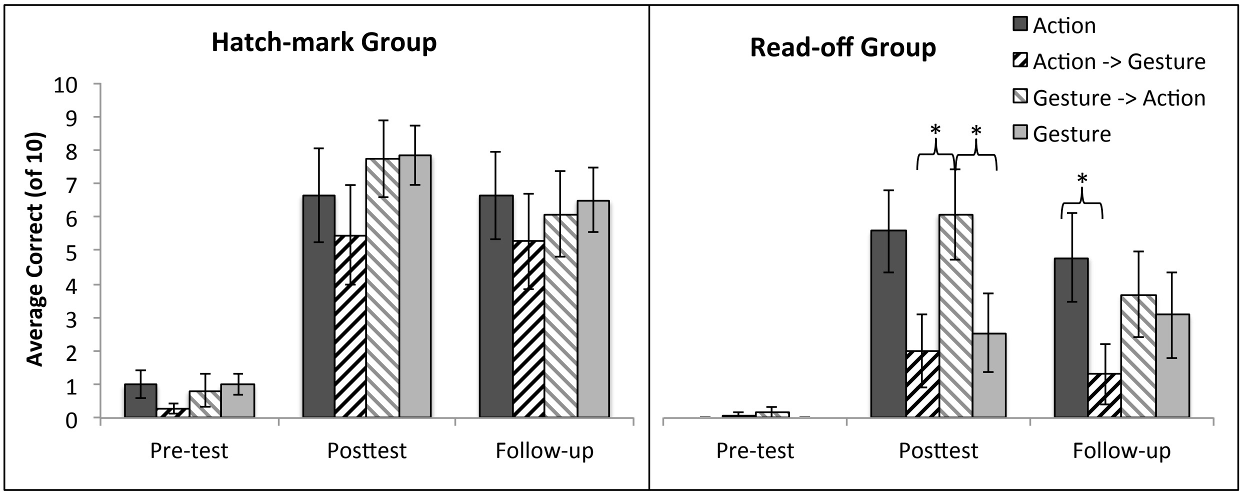

Figure 8. Average performance by starting strategy and training condition across the three sessions. Bars represent +/- 1 standard error of the mean when the data are aggregated by participant.

For the children who began the study by counting hatch-marks, I built a model with training condition, testing session (pre-test; posttest; follow-up) and the interaction between the two as fixed effects. Subject was entered as a random effect, and I controlled for the test item and the gender of the child. An analysis of variance of the model revealed a main effect of testing session ($X^2 = 240.59$, $p < .0001$), which was qualified by a marginal testing session X condition interaction ($X^2 = 15.90$, $p < .05$). There was no main effect of training condition ($X^2 = 1.70$, $p = 0.64$) and gender was not significant ($X^2 = 0.32$, $p = 0.57$). Across the conditions, there were no significant differences by condition within each of the three testing sessions.

For the children who began the study by using the read-off strategy, I first determined whether the general pattern of learning in the current study replicated the results of Experiment 2. I ran a simple model with condition as a fixed effect and subject and question as random effects. Overall, children in the Action group marginally outperformed those in the Gesture group (β= 2.62, p<.08) and significantly outperformed children in the Action-then-Gesture group overall (β= 4.15, p<.001). This pattern was true at both post-test and follow-up.

Next, I built a model with training condition, testing session (pre-test; posttest; follow-up) and the interaction between the two as fixed effects for the read-off group. An analysis of variance of the model showed a main effect of condition (X^2 = 11.03, p<.05) and a main effect of testing session (X^2 = 57.61, p<.0001), which were qualified by a significant condition by testing session interaction (X^2 = 29.96, p<.001). Gender was a non-significant predictor and thus was dropped from further analysis in this group (X^2 = 0.04, p=.83). To explore the condition by training session interaction, I built a model for each training session separately and used the Action-then-Gesture group as the comparison baseline, as they had the overall lowest performance. There were no condition differences at pre-test, but at posttest, the Gesture-then-Action group showed higher performance than the Action-then-Gesture group (β= 20.81, p<.001). Re-leveling the model with Gesture as the baseline showed that children in the Gesture-then-Action group outperformed children in the Gesture group too (β= 20.54, p<.001). Thus, at post-test, the Gesture-then-Action group showed the largest gains, which were significant larger than the gains for either the Action-then-Gesture group or the Gesture group. At follow-up, the Action group showed higher performance than the Action-then-Gesture group (β= 21.83, p<.001). All other group comparisons were non-significant. Thus, there was only one significant difference at Follow-up – the Action Group outperformed the Action-then-Gesture group.

Another set of analyses compared conditions that used one or two representations. For both the hatch-mark counting group and the read-off group, there was no effect of the number of representations (one vs. two) at any time point (all p's > 0.69).

Strategy Analysis. Once again, I performed a descriptive analysis of the kinds of errors children in both groups were making before and after training to ask whether some children were showing qualitative improvements that were not captured by our main outcome (Figure 9). This analysis showed that training led some children in the read-off group to switch their strategy to the more sophisticated, yet still incorrect, hatch mark strategy across all four training conditions. By contrast, none of the children in the hatch-mark group switch to a read-off strategy after training in any instructional condition.

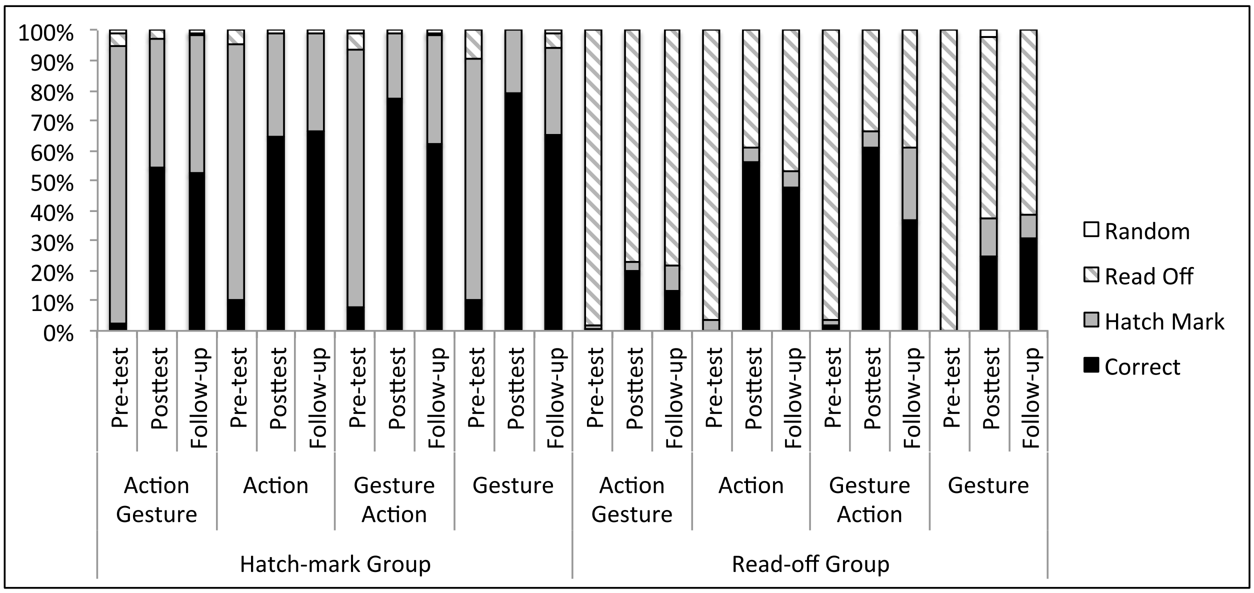

Figure 9. Each trial was coded based on the child's response. This chart, broken down by starting strategy (hatch-mark and read-off), training condition (action-then-gesture; action; gesture-then-action; gesture), and time point (pre-test, posttest, follow-up) shows the distribution of strategy use across the entire study. While some children in the read-off group switched to the hatch-mark strategy after training, children in the hatch-mark group never switched to the read-off strategy.

Transfer Tasks. There were three tasks that were administered at both pre-test and follow-up to assess growth across the session (numbered circles; colored unit chips; 'going to the store'). For these tasks, a set of linear regression models used improvement on the main ruler and crayon outcome (post-test score minus pre-test score); training condition; and the interaction between condition and improvement to predict the change in performance on the generalization task from Session I to Session II (Table 4). For the two tasks that were administered only at follow-up to assess transfer (perimeter and 'broken ruler'), the same fixed effects were used to predict performance on each transfer task (Table 5). While performance on some the tasks was correlated with improvement on the measurement task overall (numbered circles, broken ruler, and perimeter), there were no interactions with training condition for any of the tasks.

A.	Task	Group Means of Δ Score (SD)	Group Comparison	All Participants Relation to Learning
	Numbered Circles	RO = 0.72 (.95) HM = 0.09 (.67)	p = 0.0003***	$X^2_{Learning}$= 6.92 (p < 0.01)** $X^2_{Condition}$= 4.13 (p = 0.12)[1] $X^2_{Interaction}$= 0.74 (p = 0.78)
	Colored Unit Chips	RO = 0.06 (.68) HM = 0.02 (.60)	p= 0.79	$X^2_{Learning}$= 0.00 (p = 0.97) $X^2_{Condition}$= 2.18 (p = 0.16) $X^2_{Interaction}$= 0.80 (p = 0.59)
	'Going to the Store'	RO = 0.06 (0.74) HM = 0.18 (0.61)	p = 0.39	$X^2_{Learning}$= 0.88 (p = 0.16) $X^2_{Condition}$= 1.92 (p = 0.25) $X^2_{Interaction}$= 1.67 (p = 0.31)

Table 4. This shows children's improvement in performance on the transfer tasks that were administered at both pre-test and follow-up.

[1] Though there was no main effect of condition overall, an analysis within the read-off group showed a marginal main effect of condition (X^2 = 3.92, p = 0.07) even after controlling for follow-up score, which was a significant predictor (X^2 = 15.39, p < 0.001). In this analysis children in the Gesture-then-Action group improved marginally more than children in the Action group (p = .13) and significantly more than the Gesture group (p < .05) on the numbered circles task.

B.	Task	Group Means (SD)	Group Comparison	All Participants Relation to Learning
	Broken Ruler	RO = 0.81 (.71) HM = 1.58 (1.25)	p = 0.0002***	$X^2_{Learning}$= 16.18 (p<.001)*** $X^2_{Condition}$= 2.36 (p = 0.46) $X^2_{Interaction}$= 3.36 (p = 0.30)
	Perimeter	RO = 0.38 (.96) HM = 0.93 (1.01)	p= 0.007**	$X^2_{Learning}$= 6.87 (p < 0.05)* $X^2_{Condition}$= 1.21 (p = 0.76) $X^2_{Interaction}$= 0.76 (p = 0.87)

Table 5. This shows children's performance on the tasks that were administered only at follow-up.

Learning Trajectory During Training. The pattern of performance in the read-off group did not map on cleanly to any of the a priori predictions about possible learning outcomes. Given this surprising result, I decided to look at the trajectory of learning *during* the training process to see whether there were any identifiable issues during the training itself. Recall that each child received 8 total training trials. In the Gesture-then-Action and the Action-then-Gesture groups, children saw 4 training trials of one type and the 4 of another. A trained coder re-watched each video and simply marked whether each child answered each training problem correctly on his or her first guess (before using the gesture or action to check the answer). The results are displayed in Figures 10 and 11. Overall, there do not appear to be dramatic differences by condition in performance during training, suggesting that these learning trajectories cannot explain the observed differences at post-test.

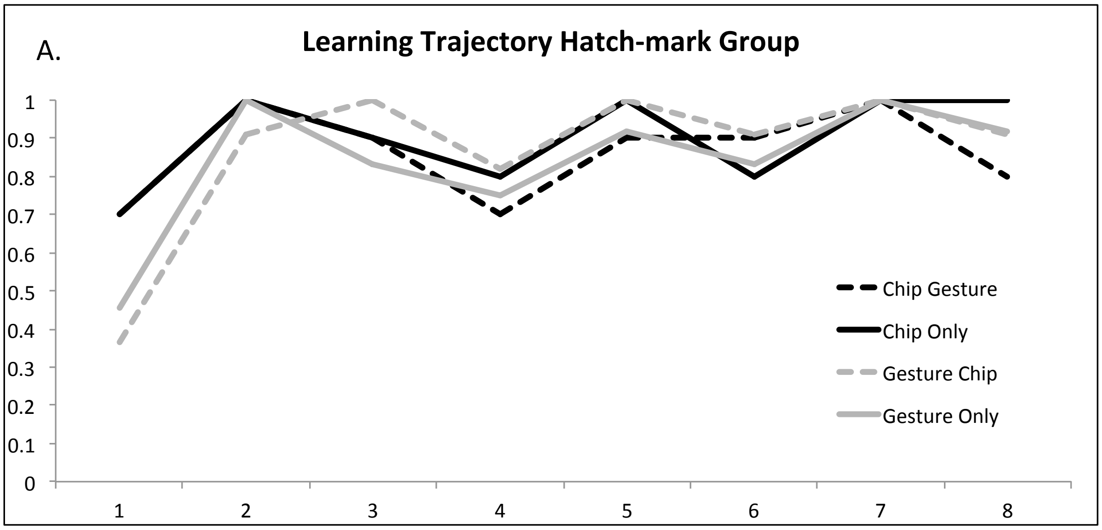

Figure 10. The proportion of hatch-mark counting children in each training condition who answer each of the 8 training trials correctly.

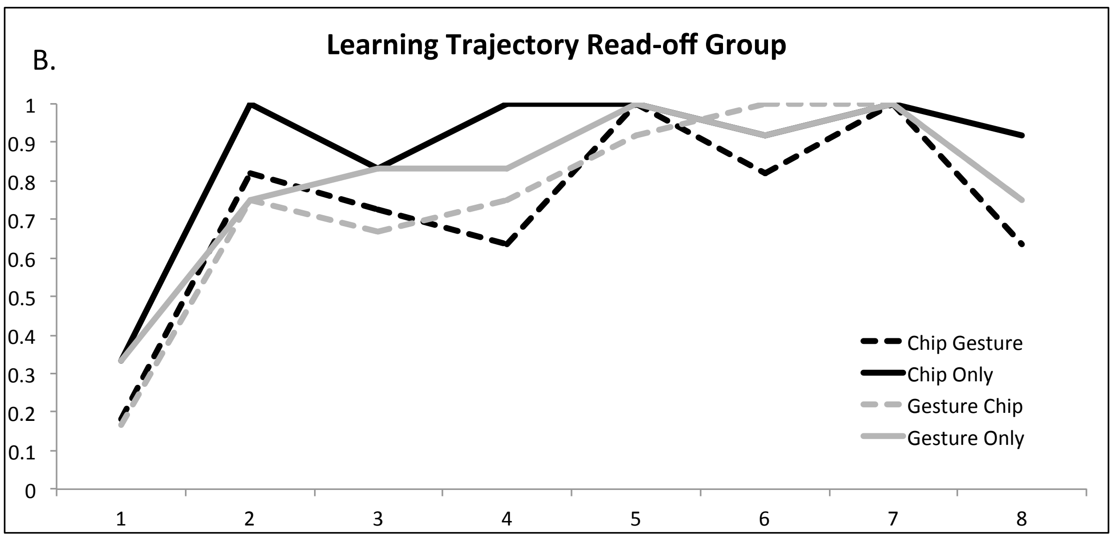

Figure 11. The proportion of read-off strategy children in each training condition who answer each of the 8 training trials correctly.

Finally, to investigate whether there were any qualitative differences during training *within* the two sequential training conditions for the read-off group, a trained coder determined from the videotape whether each child was properly performing the movement on each of the training trials. A 'proper' movement was one that closely matched that of the experimenter and did not require corrections or additional modeling. The results of this coding are displayed in

Figure 12 and reveal that children who were asked to perform gesture during the first half of the training struggled with producing the movement correctly.

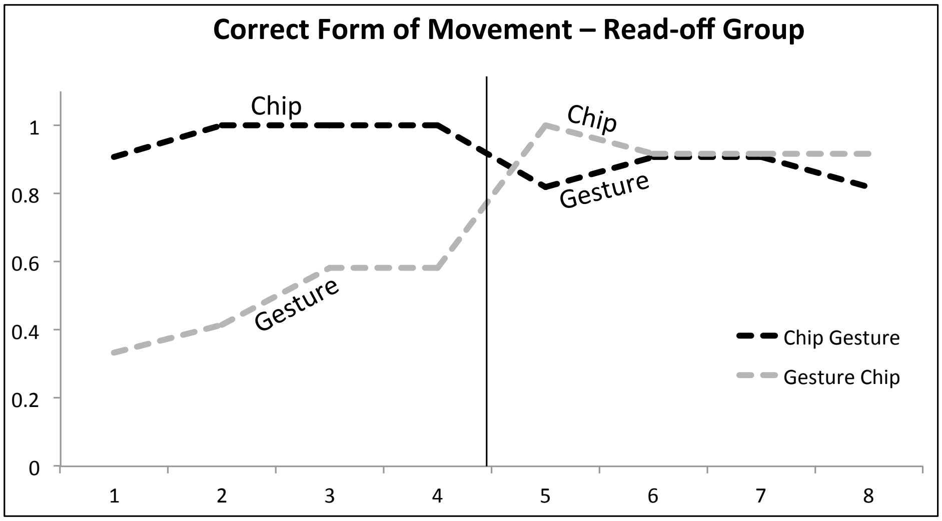

Figure 12. The proportion of children who were correctly performing the movement the experimenter had modeled during the two sequential training conditions of interest. The vertical black line marks the point in training where the training switched from one movement-type to the other.

Discussion

In Experiment 2, I replicated the pattern of results from Experiment 1 and tested a new hypothesis. The replication, in general, showed that children who used a read-off strategy at pre-test learned more from action-based instruction than gesture-based instruction and performed worse on transfer tasks than children in the hatch-mark group. With two new training conditions, I also tested whether children in the read-off group might benefit from seeing a unit represented through both action and gesture in the same training session. Contrary to my predictions, children in the read-off group did not benefit from instruction that began with a concrete, physical demonstration and transitioned to a more abstract, gesture-based demonstration, nor did they benefit from seeing multiple representations overall. Instead, they learned most when the abstract gesture training was followed by the concrete action training. Here, I discuss several possible explanations for the observed pattern of results, and outline some remaining questions that were raised by these data.

The first observation from these data is that overall, two representations (acting on and gesturing about units) are not better for learning than a single representation. This pattern was true for both the lower and higher prior knowledge groups on both immediate learning and transfer tasks. Secondly, within the hatch-mark group, children improved significantly after all types of training, and maintained that improvement at follow-up. These children also showed significantly higher rates of transfer than those in the read-off group, particularly on the two tasks that were administered only at Session II.

Within the read-off group, I found that the order of the presentation of action and gesture mattered for learning outcomes. Children performed significantly worse when they saw action followed by gesture than when it was presented the other way around. One possibility is that these data are the result of a recency effect whereby children in the read-off group were essentially ignoring the first half of training and focusing on the second half of training. Under this hypothesis, children who concluded the training with a concrete action should outperform those who concluded the training with an abstract gesture. However, an investigation of the learning trajectories during training does not support this hypothesis. Children's performance, particularly in the two sequential training groups (the dashed lines in Figure 12), track remarkably closely with one another and with the other training conditions across the training session. If children were overtly ignoring the first part of the training in either of the sequential training conditions, performance during the first part of training should be markedly lower – a pattern I do not observe.

An alternative possibility to explain these data is that the abstract gesture in the Gesture-then-Action training condition is actually serving as a kind of placeholder for children. Under this hypothesis, the abstract gesture is still confusing for children in the read-off group, but when they see the gesture *then* the action, the confusion is resolved. The theory underlying this explanation comes from the 'desirable difficulties' literature, which states that providing children with a challenge makes them more receptive to subsequent information and improves learning and transfer outcomes over and above a scenario with no such challenge (e.g., Bjork, 1994). In this case, gesture is providing the challenge, and immediately following that challenge with a more concrete action really helps children to learn. In the current study, children in the Gesture-then-Action condition showed very high performance at post-test, and though there was a slight decrease in performance on follow-up, an investigation of strategy use shows that many of the children maintained some learning by reverting to a more advanced hatch-mark counting strategy rather than reverting fully back to the read-off strategy. Also, while there was very limited evidence of condition-dependent transfer in this study as a whole, there was one task that did show a pattern within the read-off group. On the numbered circles task, read-off children in the Gesture-then-Action group improved marginally more than children in the Action only group and significantly more than children in the Gesture only group. This pattern is consistent with the hypothesis that the Gesture-then-Action condition was a powerful driver of deeper learning and transfer. Finally, an investigation of the form of children's movements during the training process showed that children who saw the gesture first were really struggling to properly perform the movement. In other words, they appeared to be struggling more or working harder to achieve similar levels of performance on the training, which is also consistent with a 'desirable difficulties' hypothesis. Taken together, the qualitative and quantitative performance during training, performance after training at post-test and follow-up, and improvement in performance on the numbered circles transfer task, are suggestive of a desirable difficulties gesture story. More research is needed to further explore this intriguing hypothesis.

Limitations. One clear limitation of this study was the degree to which the selected transfer tasks did not differentiate learning outcomes by condition (with the exception of the numbered circles task for the read-off group). In contrast to Experiment 1, the majority of tasks reported here did correlate with improvement on the main task. However, once again, overall rates of improvement and performance were quite low on the transfer tasks. This particular limitation means that it can be difficult to capture qualitative differences in learning or differences in 'deep learning' that cannot be detected by performance on the main outcomes of interest. It also could be the case that instruction on the main task needs to be more prolonged and involve deeper discussion of units in order to transfer to other related tasks.

It remains unclear from the current study how generalizable the findings are to other domains of math and to other types of instruction. The results show that in this paradigm, neither the higher prior knowledge group nor the lower prior knowledge group benefitted drastically more from multiple types of movement instruction in the same session. It is possible, however, that benefits would start to emerge after a higher dosage of training (i.e., 8 trials of each, 16 trails total), or with multiple exemplars of another type (i.e., actions paired with a more representationally transparent gesture). Relatedly, although we did not find the predicted benefit of concrete-to-abstract training within the read-off group, it is possible that such a benefit would emerge if the dosage of training were increased, or if the transition from concrete to abstract was spread out over a longer period of ontogenetic development (i.e., concrete training followed by abstract training several weeks later). These are open questions for future investigation.

Conclusions. In the current study, children in the hatch-mark counting group improved dramatically after all four training conditions. Children in the read-off group did best if they received the Gesture-then-Action condition or the Action only training condition. Furthermore, there was a trend whereby children in the read-off group who received Gesture-then-Action training showed higher rates of transfer than those in the Action alone training. One potential interpretation of these data is that gesture is serving as a placeholder that challenges children's assumptions and creates a cognitive dissonance. In the Gesture-then-Action training condition, this dissonance is immediately resolved. In the Gesture only condition and the Action-then-Gesture condition, this dissonance and potential confusion is carried into post-test, which may reflect the significantly lower scores in these two training conditions. While future work is needed to confirm this hypothesis, one clear overarching finding is that the abstract gesture poses a challenge for children in the lower prior knowledge read-off group in a lesson about spatial units of measurement. In Experiment 3, I ask whether this challenge stems from the fact that the gesture is abstract and representational, or whether it stems from the fact that it is iterative and does not leave a permanent trace in the external environment. In Experiment 3, I also collect spatial and verbal working memory measures to see if they might explain some of the individual variability in learning outcomes.

CHAPTER 4 – ACTION, GESTURE, AND THE ROLE OF WORKING MEMORY

Introduction

Both Experiment 1 and Experiment 2 emphasize that gesture can be difficult for lower prior knowledge learners. Across two studies and eight different training conditions, the only training condition with a gesture component that was successful in improving learning outcomes was a training condition with gesture that was immediately followed by an action demonstration. From these findings, it is clear that children with lower prior knowledge have difficulty interpreting or using the thumb-and-forefinger gesture, but it remains unclear precisely *why*, particularly when it poses no such issues for children with higher prior knowledge. In the final study of this dissertation, I deconstruct some of the features of the thumb and forefinger gesture that might prove challenging to children. Specifically, I ask whether training with a single iterated unit chip, which is similar to gesture in many ways but is not abstractly representational, will behave more like the multiple unit chip instruction, or more like the iterated, abstract gesture condition. In doing so, I hope to determine whether the gesture is difficult for lower prior knowledge learners *because* it is abstractly representational, or whether the difficulty stems from the other features of the gesture instruction that differentiate it from action (i.e., that it is iterated and does not leave a trace).

One possibility is that the iterated unit chip training condition will have a similar effect on learning outcomes as the gesture condition. That is – children in the read-off group might struggle to learn from the iterated unit chip. If so, it would indicate that lower prior knowledge learners may struggle to learn from gesture because it does not permanently change the environment. In fact, changing aspects of the external environment or 'offloading cognition' onto the environment while problem solving is one of the arguments put forth for why physical mathematical manipulatives are so useful for learners (Mix, 2010). Gestures, which do not interact directly with the environment, do not create this type of physical change. And when a learner is just beginning to learn a new idea, a movement that is transient might not give the learner the opportunity to process and remember that information as effectively as if it had created more lasting change. While it has mostly been characterized in adults, the 'time-based resource sharing model' of working memory simply states that representations in working memory decay over time unless they are refreshed (Barrouillet, Bernardin, & Camos, 2004). Under this model, children with lower prior knowledge who have larger conceptual gaps may simply have more trouble learning from any instruction that is transient and therefore more taxing on working memory.

The alternative possibility is that children in the read-off group will learn more successfully from the iterated unit chip than from the gesture. This pattern of results would suggest that gesture is difficult for children with lower prior knowledge because its representational content is opaque, not because it is iterative and transient. There is research to support this possibility as well. Production of spontaneous gestures during learning has been linked to improvements in working memory performance for the speaker (Goldin-Meadow, Nusbaum, Kelly & Wagner, 2001; Ping & Goldin-Meadow, 2010; Wagner, Nusbaum & Goldin-Meadow, 2004), but we know that this effect is only true for meaningful and not meaningless hand movements (Cook, Yip & Goldin-Meadow, 2012). Furthermore, we know from some recent work that young children and even adults do not always interpret representational hand movements as gestures, but instead may see them as meaningless hand movements or movements for their own sake if they are not provided with enough context (Wakefield, Novack, Goldin-Meadow, 2016; Novack, Wakefield & Goldin-Meadow, 2016). Thus, it is possible that if children in the read-off group see the thumb-and-forefinger gesture as meaningless or irrelevant movement, they will not glean any of the working memory benefits that are normally associated with gesture-based instruction and will have worse learning outcomes as a result. In fact, it is possible that asking a child to produce a gesture they do not understand is more detrimental for learning than incorporating no movement into instruction at all, as it might provide the child with a sort of dual task. This particular part of the hypothesis is not directly tested here, as all of the current training conditions contain movement components, but it may be a comparison worthy of future research.

In both of the hypotheses outlined here (action > iterated chip = gesture; action = iterated chip > gesture), working memory capacity plays a role in helping to explain why gesture, and potentially the iterated chip, might be difficult for some learners. Because of this, Experiment 3 includes measures of both spatial and verbal working memory, to explore whether this particular measure of individual difference can explain any of the variability we see in learning outcomes within each of the instruction types or across the instruction sessions more broadly. For example, if it is the case that gesture-based instruction is taxing on spatial working memory for some children, we might expect only those students with relatively high spatial working memory to be successful in that instructional condition. Interest in this particular measure of individual difference comes from a large body of literature that establishes a strong relationship between working memory capacity and math performance in adults (e.g., Ashcraft & Kraus, 2007; Beilock & Carr, 2005) and between working memory and math learning outcomes in children (Swanson & Beebe-Frankenberger, 2004), even controlling for other factors like IQ and domain knowledge (Alloway, 2009). In particular, children with lower mathematical ability and slower learning trajectories appear to have specific deficits in inhibitory control and working memory (Bull & Scerif, 2001). There is also some work with adults looking at the interaction between working memory, domain knowledge and performance on a non-mathematical task that shows that working memory predicts performance over and above domain knowledge (Hambrick & Engle, 2002). Taken together, these findings identify working memory as a predictor of interest in the current mathematical learning task.

<u>Methods</u>

Subjects. 94, 1st grade students (46 females; 48 males; mean age at test: 6.99 years, SD = 0.34 years) were recruited and tested at several Chicago area schools. As with Experiment 2, children in this sample were from a broader range of socio-economic backgrounds. Based on a categorical income questionnaire, children in the current study reported ranges from 2 to 6 (6 is the highest possible score). In Experiment 1, the range was 4-6; in Experiment 2, the range was 1-6. Overall, the average score reported in the current sample was still quite high (5.30 out of 6, SD = 1.28), and SES was a non-significant predictor in all models and thus is not included in any final analyses. Children whose parents signed a consent form participated in two one-on-one sessions with an experimenter, which took place one week apart in a quite area of their school (Session 1 and Session II).

Session I. To assess pre-test strategy, children were given a 14-question multiple-choice paper and pencil test (see Figure 1, Chapter 2). As described previously, the first four test items were of a crayon that was aligned with the "0" point on the ruler ("unshifted problems"). In the 10 subsequent test items the crayons were shifted to different points on the ruler ("shifted problems”). All crayons started and ended at a whole unit. The four answer choices reflected the correct answer, a read-off strategy answer, a hatch-mark strategy answer, and a fourth random choice that did not match any of the other three strategy-related options. This multiple-choice test was the main outcome of interest.

Children were categorized into a particular strategy group if 6 or more of the 10 shifted-object items were answered in a way that was consistent with a single strategy. This criterion was based on probability values of the binomial distribution: on a task with 4 options, answering 6 out of 10 using a particular strategy means that that child is using that strategy more often than would be predicted by chance or random guessing ($p < .01$). By this metric, there were 22 children in the ‘correct’ group (N=16 males); 38 children in the ‘hatch-mark’ group’ (N=18 males); 30 in the ‘read-off’ group (N=11 males); and 4 children (N=3 males) whose dominant strategy was ‘random’ or did not meet criteria for inclusion in one of the other groups. Children who were in the ‘correct’ group were excluded from further analyses given limited room for improvement after training, and children in the ‘random’ group were excluded due to the small sample size. One additional child from the read-off group was excluded for missing data on the follow-up session. Thus, the final sample consisted of 67 children.

Immediately after completing the multiple-choice crayon task (Pre-Test), children in Session I received two other unit-based tasks that were also administered at Session II to capture growth across the whole experiment. The ‘numbered circles’ crayon task, which captured some differences by training condition in Experiment 2, was once again administered at pre-test and posttest. Two trials of the perimeter task were also administered before and after training, as that task has consistently correlated with learning outcomes in both of the two previous studies.

After completing the two baseline transfer tasks, children were randomly assigned to one of three between-subjects training conditions. Assignment was counter-balanced by children's dominant initial measurement strategy. The three training conditions were: Action (N = 21; N = 9 read-off); Gesture (N = 23; N = 10 read-off); and Iterated Action (N = 23; N = 10 read-off). Once again, all training in Experiment 3 was performed with shifted objects (see Figure 13 for images of the training conditions). All children received a total of 8 training trials with experimenter feedback. Following training, children received a second version of the multiple-choice crayon measurement task (Posttest).

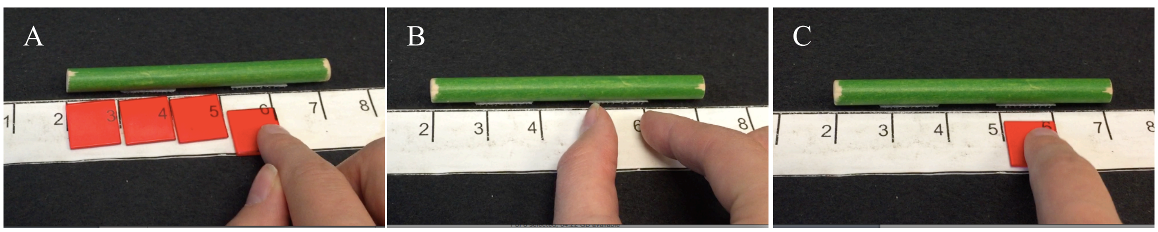

Figure 13. Still shots of the three training conditions: A) Action, B) Gesture, and C) Iterated Action.

Session II. One week after the second session (mean delay = 7.09 days, SD = 0.87 days), each participant received a third version of the multiple-choice crayon task (Follow-Up) followed by a verbal working memory and a spatial working memory assessment. The verbal working memory task was a Letter Span task in which children were read a list of letters (e.g., "B-R") and then asked to repeat the list back in the same order. The length of the list increased by one letter every two trials. To move to the next level, a child had to get at least one of the two trials correct. The child's final score was calculated by adding up the total number of items answered correctly (out of a maximum of 16 trials over 8 levels). The spatial working memory measure was a Corsi Block task (Corsi, 1972; Milner, 1971; Pagulayan et al., 2006). In this task, the experimenter uses her finger to tap the tops of cubes in a predetermined order. The child is then asked to repeat the tapping sequence in the same order. For ease of administration, the experimenter can see small numbers printed on the cubes, while the child sees only plain cubes (see Figure 14 for an image of the testing apparatus). The first level in this task is a sequence of three taps and each successive level increases by one tap. In order to move to the next level, the child has to answer at least one of the three trials correctly at a given length. If the child got all three trials of a certain block length incorrect, no further trials were administered. The child's final score is calculated by adding up the total number of items answered correctly (out of a maximum of 18 across 6 levels). In addition to the verbal and spatial working memory measures, each child was again asked to do the numbered circles crayon measurement task and the perimeter task to assess growth across the session.

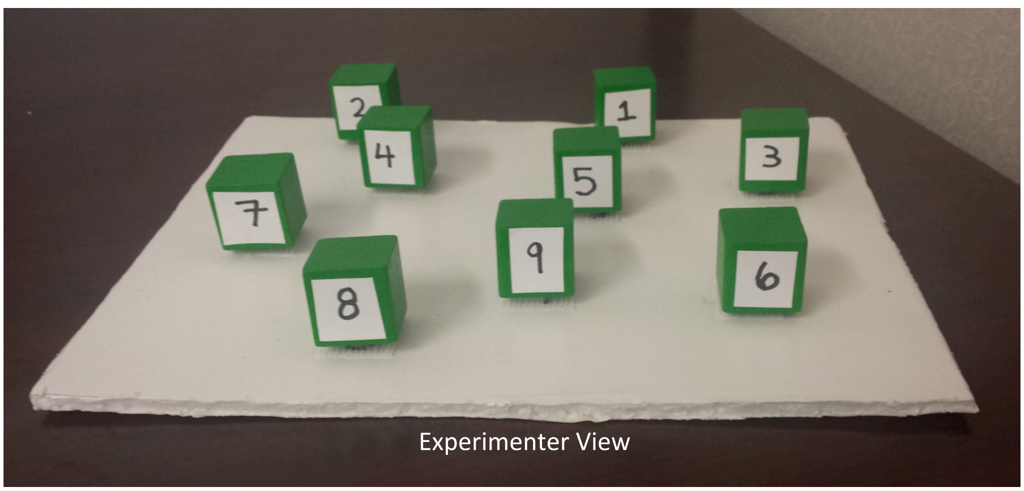

Figure 14. This figure shows the experimenter view of the Corsi block task. The child's view does not have any visible numbers.

<u>Results</u>

Once again, performance on the four unshifted items on the multiple-choice crayon test was very high at all three time points for all participants (*M*=3.94, *SD*=0.49 at pre-test; *M*=3.76, *SD*=0.95 at immediate posttest; *M*=3.91, *SD*= 0.59 at the 1-week follow-up). As such, formal analyses focus on children's performance on the ten shifted-item questions.

On the main outcome of interest, the crayon and ruler task, the data were non-normally distributed (children either got most problems right or most problems wrong). Accordingly, the data were fit with a mixed effects binomial logistic regression model that predicted correct performance on each shifted-object test item. All analyses were performed using R (R Development Core Team, 2008). Based on a priori predictions about difference between the higher and lower prior knowledge groups, I built two separate models: one for children who predominantly used the hatch-mark counting strategy at pre-test, and one for those who began by using the read-off strategy. Means and standard errors of the means for the two groups at each session are displayed in Figure 15.

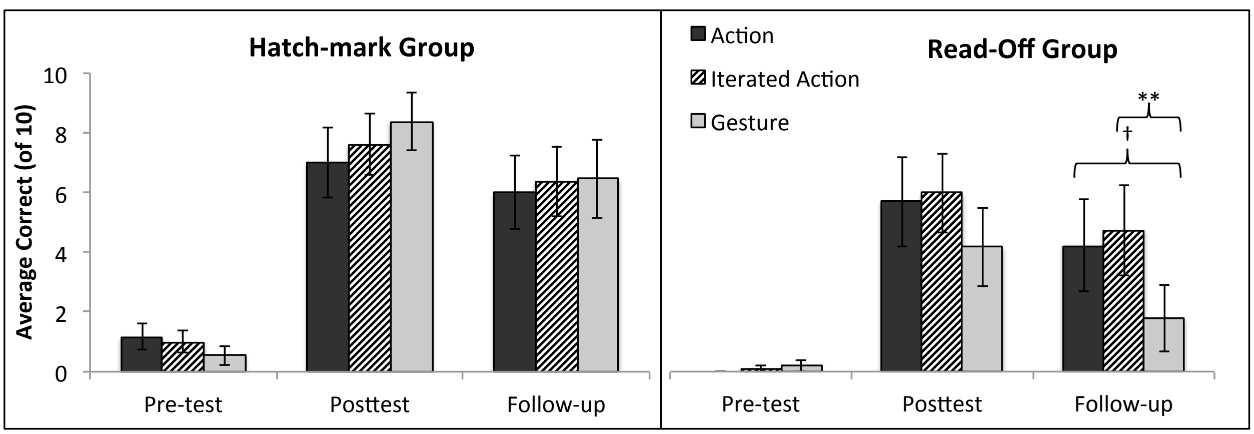

Figure 15. Average performance by starting strategy and training condition across the three sessions. Bars represent +/- 1 standard error of the mean when the data are aggregated by participant.

For the children who began the study by counting hatch-marks, I built a model with training condition, testing session (pre-test; posttest; follow-up) and the interaction between the two as fixed effects. Subject was entered as a random effect, and I controlled for the test item and the gender of the child. An analysis of variance of the model revealed a main effect of testing session ($X^2 = 266.36$, $p<.0001$), which was qualified by a significant testing session X condition interaction ($X^2 = 14.49$, $p<.01$). There was no main effect of training condition ($X^2 = 0.10$, $p=0.95$) and gender was not significant ($X^2 = 0.24$, $p=0.63$) and thus was dropped from future analysis. To explore the condition by training session interaction, I built a model for each training condition separately and used posttest as a baseline comparison group. Question and subject were both included as random effects. This analysis showed that performance in all three groups dropped from posttest to follow-up, but the drop was smallest in the Action group and largest in the Gesture group ($\beta= -0.76$, $p < .05$ for Action; $\beta= -1.04$, $p < .01$ for Iterated Action; $\beta= -2.84$, $p < .00001$ for Gesture). Despite these minor differences in learning trajectory within the conditions, there were no significant differences by condition within each of the three testing sessions (all p's > 0.28), and learning across all conditions was quite high.

For the children who began the study by using the read-off strategy, I built a similar model with training condition, testing session (pre-test; posttest; follow-up) and the interaction between the two as fixed effects. An analysis of variance of the model showed a main effect of testing session ($X^2 = 80.77$, $p<.0001$). Gender was a non-significant predictor and thus was dropped from further analysis in this group ($X^2 = 2.40$, $p=.12$). The interaction between training condition and testing session was not statistically significant ($X^2 = 6.19$, $p=.19$), but it was trending towards significance as was reported in Experiments 1 and 2. To look at performance by condition within each testing session, I built a model for each session separately and used the Gesture group as the comparison baseline, as they had the overall lowest performance. There were no condition differences at pre-test or posttest, but at follow-up, the Action group and the

Iterated Action group showed higher performance than the Gesture group (β= 6.3, p = .07 and β= 7.8, p < .05, respectively).

Strategy Analysis. Once again, I performed a descriptive analysis of the kinds of errors children in both groups were making before and after training to ask whether some children were showing qualitative improvements that were not captured by the main outcome (Figure 16). This analysis showed that training led some children in the read-off group to switch their strategy to the more sophisticated, yet still incorrect, hatch mark strategy. This effect was most pronounced within the Gesture group and the Action group. Very few read-off strategy users in the Iterated Action condition switched to a hatch-mark counting strategy. Consistent with Experiment 1 and Experiment 2, none of the children in the hatch-mark group switch to a read-off strategy after training in any instructional condition.

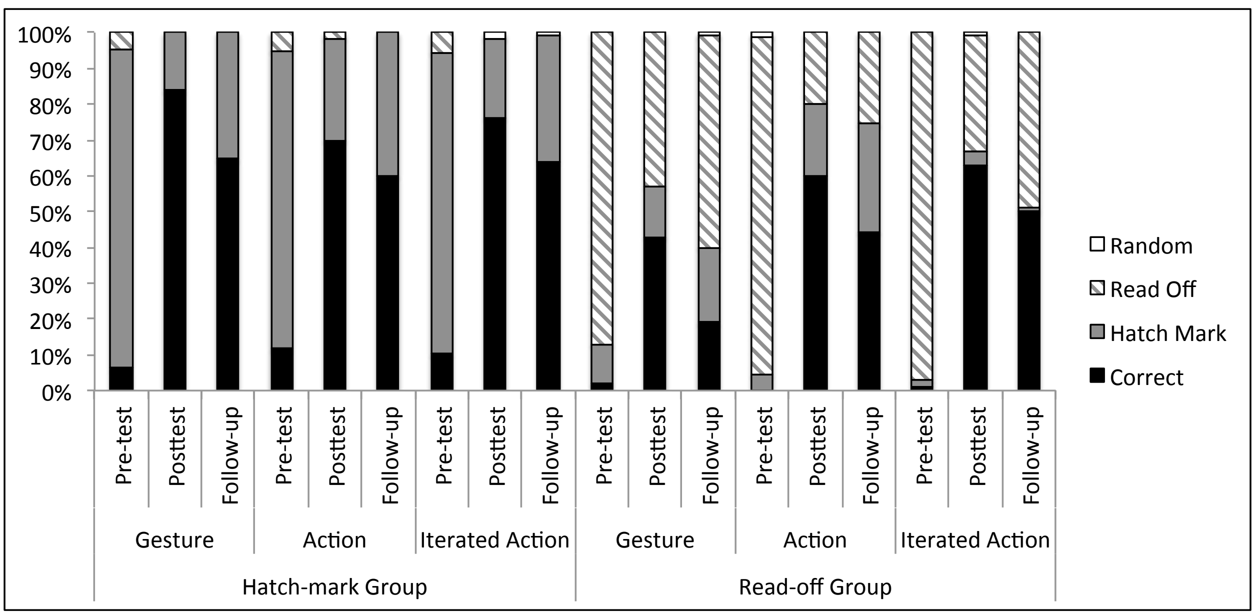

Figure 16. Each trial was coded based on the child's response. This chart shows the distribution of strategy use across the entire study.

Transfer Tasks. For the tasks administered at both Session I and Session II (numbered circles and perimeter), I ran a set of linear regression models with improvement on the main ruler and crayon outcome (post-test score minus pre-test score); training condition; and the interaction between condition and improvement to predict the change in performance on the transfer task from Session I to Session II. In the numbered circles task, improvement was correlated with learning outcomes, but was not related to training condition in any way. In the perimeter task, improvement from pre-test to follow-up was marginally correlated with learning outcomes overall, and there was a significant main effect of training condition whereby children in the two action-based condition (Action and Iterated Action) improved significantly more than the children in the Gesture condition ($\beta = 0.46$, $p < .05$ for Action compared to Gesture and $\beta = 0.47$, $p < .01$ for Iterated Action compared to Gesture).

A.	Task	Group Means of Δ Score (SD)	Group Comparison	All Participants Relation to Learning
	Numbered Circles	RO = 0.69 (1.00) HM = 0.18 (.65)	$p = 0.015$*	$X^2_{Learning} = 4.27$ ($p < 0.05$)* $X^2_{Condition} = 0.78$ ($p = 0.58$) $X^2_{Interaction} = 0.033$ ($p = 0.98$)
	Perimeter	RO = 0.17 (.60) HM = 0.026 (.64)	$p = 0.34$	$X^2_{Learning} = 0.10$ ($p = 0.11$) † $X^2_{Condition} = 3.33$ ($p < 0.05$) * $X^2_{Interaction} = 0.63$ ($p = 0.40$)

Table 6. This shows children's change in performance on the tasks that were administered at both pre-test and follow-up. On the numbered circle task, children in the hatch-mark counting group were performing near ceiling at pre-test, which is why improvement is significantly greater in the read-off group.

Working Memory. Working memory tasks were administered for three reasons. First, to determine whether there were any global differences in working memory capacity based on children's starting strategy. Second, to see whether spatial working memory, verbal working memory, or both were predictive of learning or retention irrespective of condition. And finally, to see whether a certain type of working memory (spatial vs verbal) predicted performance differentially based on training condition.

For the first analysis, I ran a simple model using starting strategy (read-off and hatch mark) to predict performance on each of the working memory measures. I used hatch-mark counting as the baseline strategy group, and also controlled for both SES and Gender in the models. For the letter span task, there was no relation to strategy (β= -0.45, p = .21), but there was a significant relation to strategy for the Corsi Block Task (β= -2.15, p < .001), whereby children in the hatch-mark counting group had higher spatial working memory than children in the read-off group. See Figure 17 for means and standard error of the means for both tasks. Gender and SES were not significant predictors in either model (all p's > 0.29).

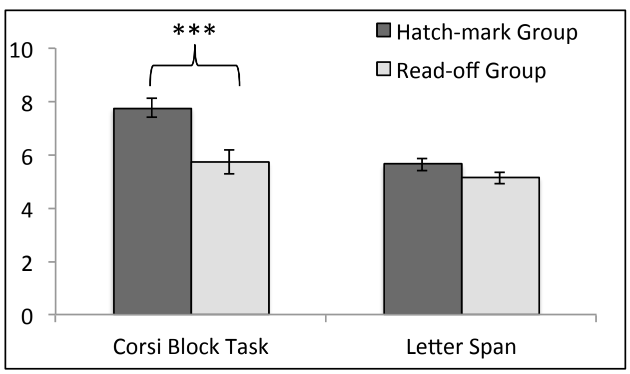

Figure 17. Mean performance on the two working memory tasks. Error bars are +/- 1 SE of the mean.

Beyond these overall differences by starting strategy, there was a large distribution of scores within each type of working memory task, suggesting a big range of individual differences (Figure 18). For the second analysis, I tested whether these individual differences in either spatial or verbal working memory were predictive of whether a child was a learner at posttest or retained learning at follow-up, irrespective of starting strategy.

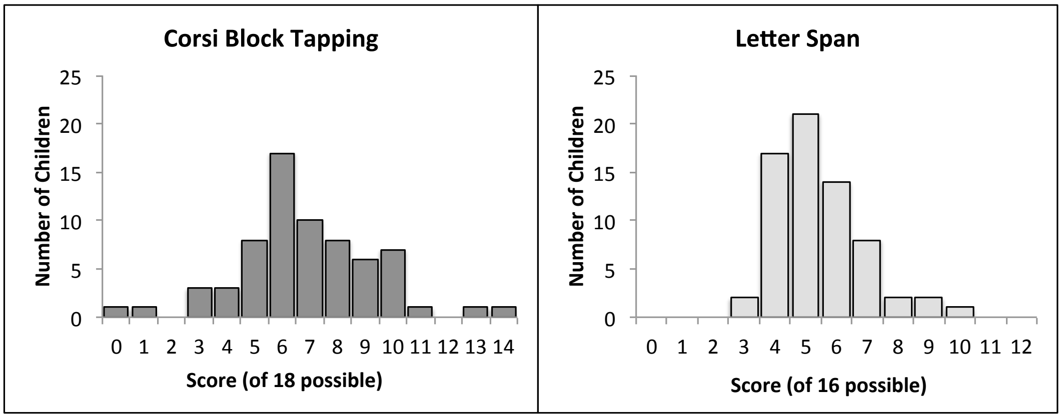

Figure 18. Distribution of scores on spatial and verbal working memory measures.

I first coded each child as a 'learner' or 'non-learner' at each session to create a binomial outcome measure of learning. A participant was considered a 'learner' if he or she answered 6 or more of the 10 crayon measurement items correct at a given session. Again, this coding scheme was appropriate given the bimodal distribution of the data – most children answered either all or none of the questions correctly. I then built a series of simple regression models using either spatial or verbal working memory to predict learning status at post-test or retention status at follow-up. Figure 19 displays the average scaled scores for learners and non-learners on each of the working memory tasks at posttest and follow-up. To scale the scores, I used a standard method – subtracting the mean score from each participant's score and dividing by the standard deviation, to give a mean of 0 and a standard deviation of 1 for each working memory measure across all participants. The scaling process allows me to compare effect sizes between the two working memory measures, which had slightly different levels of average difficulty and slightly different total possible scores. These scaled scores are used in the remainder of the working memory analyses. Overall, this first analysis shows that spatial working memory is marginally predictive of learning and significantly predictive of retention (β= 0.54, p = .067 at post and β=

0.71, $p < .01$ at follow-up). Verbal working memory was not predictive of learning and was only marginally predictive of retention (β= 0.17, $p = .49$ for post and β= 0.427, $p = .059$ for follow-up). When both measures are entered into the same model, only spatial working memory is predictive of performance (Corsi: β= 0.09, $p = .08$ for post and β= 012, $p < .05$ for follow-up; Letter Span: β= 0.02, $p = .78$ for post and β= 0.09, $p = .17$ for follow-up).

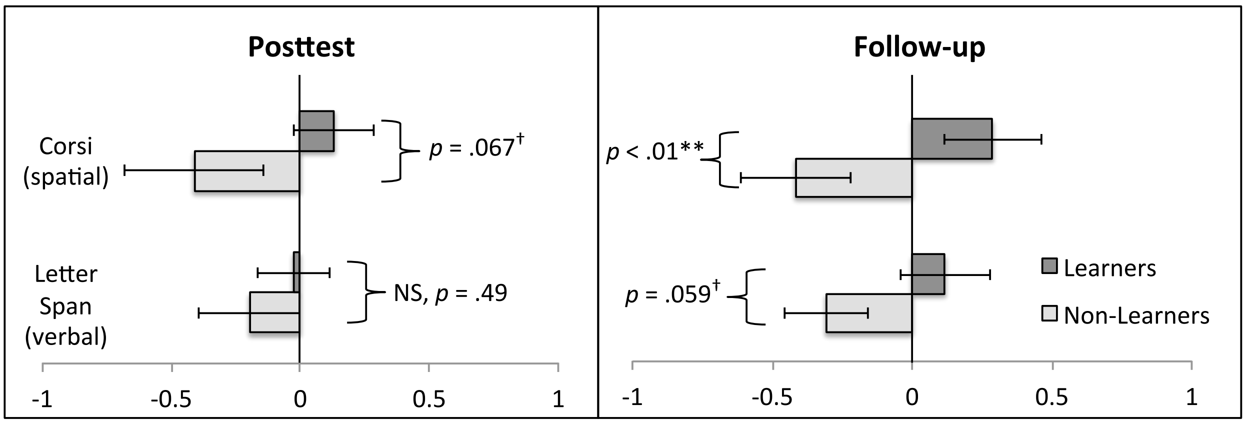

Figure 19. This figure displays the mean scores on each working memory task for learners and non-learners at posttest and follow-up. Error bars are +/- 1 SE of the mean.

The final set of analyses investigates whether spatial or verbal working memory differentially predict learning or retention *within* each training condition (Figure 20). Given the small sample size for this final question, findings should be considered preliminary. The analysis reveal that spatial working memory marginally predicts learning in the action condition (β= 0.15, $p = .111$) and retention in the gesture group (β= 0.14, $p = .07$). Verbal working memory marginally predicts retention in the iterated chip group (β= 0.186, $p = .055$). While these findings are marginal due to the small sample sizes, the patterns are suggestive of the idea that the different training conditions rely on different aspects of working memory.

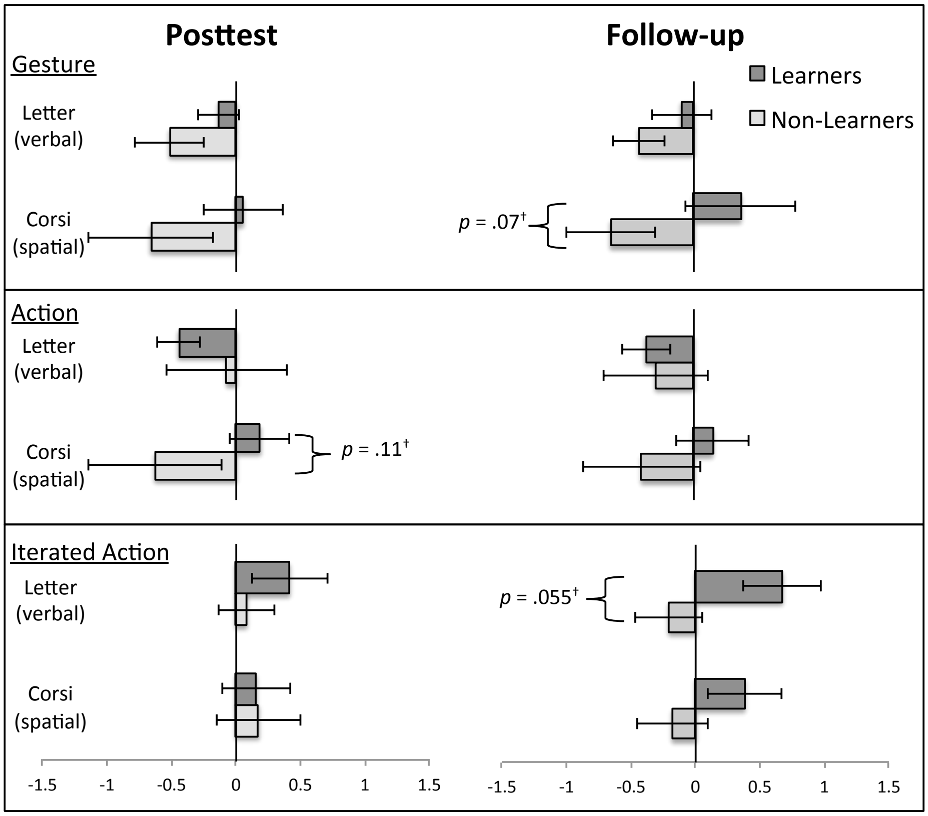

Figure 20. This figure displays the mean scores on each working memory task for learners and non-learners at posttest and follow-up within each training condition. Error bars are +/- 1 SE of the mean.

Discussion

The findings from Experiment 3 are threefold. First, though there were non-significant differences in learning on immediate posttest, children with lower prior knowledge performed significantly better at follow-up after Action instruction as compared to Gesture instruction. This replicates the general pattern of results reported in Experiment 1 and Experiment 2. Second, within the read-off group, children in the newer condition of interest, Iterated Action, performed significantly better than those in the Gesture training and did not perform any differently from those in the Action condition. Finally, Experiment 3 explores the role of both verbal and spatial working memory and suggests some exciting areas for future research.

Though the general pattern of results for children in the read-off group is similar in this study to that reported in the previous two studies, one notable difference is that the effect does not emerge as a significant difference until the follow-up session. It is not clear at this juncture why the group differences emerge primarily at Session II, though it could be related to the slightly lower SES than Experiment 1 -- the learning differences have been attenuated in the more diverse SES samples, combined with a smaller sample size than Experiment 2 -- a larger sample could make the trending differences at posttest statistically significant. The alternative possibility is that for reasons unexplained by the demographic information collected (age, SES, school, gender), children in this sample were more familiar with gesture on average or more likely to learn from it for some other unidentified reason (12% learned from the Gesture training in Experiment 1; 25% in Experiment 2; 44% in Experiment 3). Nevertheless, there was still remarkable variability within the group, in that some read-off children learned well from gesture, but the majority of children continued to struggle with it.

The primary training condition of interest in the current study was the Iterated Action condition. This condition was designed to share some features with the Gesture training – it was iterated and did not leave a permanent trace -- and some features with the Action training – there was a concrete unit chip to touch and manipulate rather than just the abstract idea of a unit represented by gesture. Children learned quite well from this training condition and maintained that learning over time more so than children in the Gesture condition. This result is consistent with the idea that it is the specific, representational features of the thumb-and-forefinger gesture that make it difficult for children, and not the other properties of the movement (that it is iterative and leaves no trace). From an educational perspective, this finding emphasizes the importance of considering the exact form of the gesture being used in instruction. It raises the possibility that a different representational gesture, perhaps one that traces the unit space from left to right, or one that simply indicates the space between two lines, could be more effective in this learning situation. From a theoretical perspective, this finding suggests that the key in selecting an effective gesture for any learning scenario might rest on whether the representational content of the gesture is accessible to a given learner. This dissertation underscores that a child's prior conceptual knowledge is a strong predictor of this access to representational content.

Beyond the immediate learning outcomes, there were also two transfer tasks administered at both pre-test and follow-up to assess growth across the sessions. Improvement on both tasks correlated at least marginally with learning outcomes, but only the perimeter task reflected differences based on training condition. Children improved significantly more after training in both of the action-based conditions (Action and Iterated Action) than they did in the Gesture condition even after controlling for differences in the rates of learning. Despite the fact that the perimeter task has been administered in all three experiments in various forms, this is the first time that it has reflected such significant condition differences. However, revisiting the results from Experiment 1, the only other time when the perimeter task was administered twice to assess growth, reveals that there was a marginal effect of condition after controlling for learning in that study as well ($p = 0.17$). The data show a similar trend whereby children in both of the action conditions improve marginally more after training than children in both of the gesture conditions. These patterns seem somewhat contradictory to some of my own previous work extolling the benefits of gesture-based instruction to promote transfer (Novack, Congdon, Hemani-Lopez & Goldin-Meadow, 2014). One possible explanation that arises from these data is that action instruction is more appropriate for promoting transfer in this scenario because units are concrete, spatial entities rather than abstract ideas, like mathematical equivalence. Overall though, transfer rates remain quite low across all three of these studies, so future work is needed to investigate the strength and generalizability of this effect.

Finally, the current study addressed three questions about the relationship between spatial working memory and verbal working memory and performance on a linear measurement task. The first finding was a strong and significant interaction between working memory and starting strategy. There were no differences in verbal working memory between children in the read-off group and the hatch-mark group, but there was a large difference in spatial working memory whereby hatch-mark counting children had higher spatial working memory scores than their read-off strategy peers. This is consistent with some previous literature showing that verbal working memory is less predictive of math achievement than visuo-spatial working memory (Holmes & Adams, 2006). This finding also seems parsimonious with the argument that linear measurement is a spatial task, which could mean that higher spatial working memory should correlate with a better understanding of the task. To date, this finding is simply correlational and it would be interesting to probe the causal link between spatial working memory and higher conceptual knowledge of linear measurement.

Additional analyses explored the relationship between individual differences in spatial and verbal working memory and learning status at posttest and follow-up. The results showed that spatial working memory was a predictor of learning status at both sessions and that verbal working memory was a marginal predictor of learning status at follow-up irrespective of training condition and starting strategy. Again, spatial working memory could be a stronger predictor overall because linear measurement is a spatial task. It is less clear why individual differences in verbal working memory would marginally relate to performance at follow-up and not during learning itself. One possibility is that verbal working memory is serving as a proxy for verbal skills more broadly (e.g., Baddeley, 2003), and higher verbal skills could play a role in remembering a lesson over time, though this specific link is not well established by previous research. Furthermore, when both factors were entered into the same model, only spatial working memory was predictive of learning outcomes, meaning the link between verbal working memory and learning outcomes is tenuous at best. In the final set of analyses looking at how the two types of working memory might contribute differentially to learning based on the type of training, spatial working memory was marginally related to success in the Gesture condition and the Action condition, whereas verbal working memory marginally contributed to success in the Iterated Action condition. Due to the relatively small sample sizes, this is only preliminary evidence that spatial and verbal working memory may contribute to success differentially depending on the features of the training condition. More work is needed to explore this intriguing possibility.

Conclusions. In the current study, children in the hatch-mark counting group improved dramatically after all three training conditions. Children in the read-off group did best in the Action and Iterated Action conditions and did not do as well in the Gesture training condition. This result suggests that gesture may be difficult for some learners *because* its representational meaning is unclear, and not because it is iterative or leaves no permanent change in the environment. Furthermore, success in the Gesture training condition was associated with higher spatial working memory, suggesting that gesture may be drawing heavily on spatial working memory resources during training (resources that are significantly lower in the read-off group). Interestingly, when there was evidence of transfer to another unit-based task, it was strongest in the action-based training conditions, suggesting that this particular domain of mathematics, linear measurement, could be better suited to action-based instruction. Overall, Experiment 3 provides convincing evidence that lower prior knowledge in the domain of linear measurement means that children struggle with the representational content of a gesture that is perfectly clear to their higher prior knowledge peers. This difficulty cannot be explained by the fact that the gesture is iterative and does not leave a trace, as children in the read-off group had no trouble learning from an Iterative Action condition. Thus, Experiment 3 underscores the importance of considering the *type* of gesture used during instruction in addition to the child's prior knowledge of a domain.

CHAPTER 5 – GENERAL DISCUSSION

The goal of this dissertation has been to understand how individual differences in a learner's prior knowledge might predict their propensity to learn from action versus gesture during instruction. Across three experiments, I demonstrated that children with lower prior conceptual knowledge learn better from action-based instruction, while their higher prior knowledge peers can learn equally well from actions and gestures. In doing so, I also explored some of the features of gesture that make it challenging for naïve learners, and found at least one example where those same challenging features, when presented in the right context, actually provided a 'desirable difficulty' for learners that led to deeper learning. In this final chapter, I discuss the direct theoretical and practical implications of these findings; I explore the generalizability of the effects reported; and I review open questions for future research.

Part I: Practical Implications

Linear measurement is an important foundational mathematical concept, yet children in the United States struggle more with measurement than they do with any other subdomain of mathematics (TIMSS, 2011; Foy, Arora & Stanco, 2013). This difficulty is most pronounced on standardized test items where children are asked to measure objects that are shifted away from the start edge of the ruler. Alarmingly, difficulties on these problems can persist as late as middle school, when up to 38% of 8^{th} grade students will answer a shifted measurement problem incorrectly (Lindquist & Kouba, 1989).

In the current dissertation, a total of 333 first grade students were assessed for their understanding of linear measurement on shifted-object items. Only 18% of these students answered the majority of items correctly, while 43% used a hatch-mark counting strategy, 37%

used a read-off strategy, and 2% did not have a consistent strategy at all. These results are consistent with research showing that children rarely, if ever, see shifted-object measurement items in their classrooms. In fact, the majority of classroom exercises ask students to align an object with the start of a ruler, and read off the number that corresponds to the right edge of the object (Smith, Males, Dietiker, Lee, & Mosier, 2013). While this technique is effective for getting the proper result, this type of exercise leads to a shallow, procedural understanding of measurement rather than a rich, conceptual understanding of units of measure.

The current dissertation shows that after one brief training session, children *can* learn to solve shifted-object measurement problems through exposure and training on these particularly difficult problems. Across all training conditions and across both the hatch-mark and read-off groups, 60% of children who got shifted-object training significantly improved from pre-test to posttest, while only 10% of the children in more traditional, unshifted training conditions (administered only in Experiment 1) showed improvement. Shifted-object items cause children to reevaluate their preexisting strategies because children can discover for themselves that their own intuitions do not lead to a correct answer. This process, known as disconfirming evidence, has been established in some of my own previous work as a powerful driver of learning on linear measurement tasks (Kwon, Ping, Congdon & Levine, 2016).

Another major contribution of the current dissertation is a better characterization of the two major misconceptions in linear measurement, reading-off and hatch-mark counting. First, the findings clearly demonstrate that children in the read-off group are further behind in their conceptual understanding of linear measurement. Across all studies, children in the hatch-mark counting group outperformed those in the read-off group at pre-test and on all other unit-based tasks at baseline and at follow-up. Children in the hatch-mark group also learned more from instruction overall -- 64% of children in the hatch-mark group significantly improved from pre-test to posttest but only 36% of children in the read-off group showed significant improvement after training. In addition, there was evidence that children in the read-off group sometimes switched to a hatch-mark counting strategy after training, but the reverse pattern was never observed. Finally, Experiment 3 is the first study that characterizes working memory differences based on linear measurement strategy. While the two groups were matched on their verbal working memory on average, children who used a hatch-mark counting strategy at pre-test had significantly higher spatial working memory.

Despite these big gaps in pre-test performance and learning outcomes between the two strategy groups, the general implications of this dissertation are positive. In each experiment presented here, children were given a very short (3-4 minute) training session, and overall, there were rapid and long-lasting gains in understanding. Moreover, Experiment 1 offered an in-depth look at children's verbal explanations after training to show that the training caused an improved conceptual understanding of units and spatial extent, rather than of some procedural trick for 'getting the right answer'. Given that a strong conceptual understanding of units is central to a number of other mathematical topics like fractions, place-value, and division (Sophian, 2007), catching and addressing unit-based misconceptions early in development might be a particularly powerful way to positively affecting a child's long-term success in mathematics. This dissertation suggests that even a very brief, well-designed lesson could serve as such a catalyst for far-reaching conceptual change.

Part II: Theoretical Implications

There is ample experimental evidence demonstrating the benefits of learning through actions-on-objects. In the domain of mathematics, manipulatives allow children to "offload cognition" onto the environment, encourage the formation useful conceptual metaphors (Manches & O'Malley, 2012), direct attention to the relevant components of a complex problem (Mix, 2010), and engage young learners with limited attention spans (Peterson and McNeil, 2008). There is also ample research showing that children can learn new ideas through gesture across a variety of academic domains including algebra, chemistry, geometry and word learning (e.g. Wakefield & James, 2015; Macedonia, Müller, & Friederici, 2011; Ping & Goldin-Meadow, 2008; Valenzeno, Alibali, & Klatzky, 2003; Singer & Goldin-Meadow, 2005). Producing gesture may even help children to transfer their knowledge to new contexts (Cook, Duffy & Fenn, 2013) and better retain newly learned information across time (e.g., Cook, Mitchell & Goldin-Meadow, 2008; Levine, Goldin-Meadow, Carlson & Hemani-Lopez, 2016).

In the current dissertation, I directly compare these similar, though not identical types of movement within the same instructional paradigm. In doing so, I demonstrate three main findings. The first finding is that actions-on-objects and gestures have very different effects on learning and cognition despite their similarities (they both engage the motor system, they can direct a learner's attention to relevant components of a complex problem, and they even look somewhat similar to an outside observer). Second, I show that the prior conceptual knowledge of the learner is a strong predictor of the efficacy of each type of movement. And finally, I pull apart some of the features that differentiate gesture from action in a linear measurement context, and I show that it is the representational content of the gesture that makes it challenging for lower prior knowledge learners, rather than the fact that the gesture is iterative and transient. In

Experiment 2, I demonstrate that this challenge may not necessarily be a bad thing, given that children seem to show deeper learning when they see gesture followed by action than when they see action alone. Given the right context, the same features that make gesture challenging may also be responsible for its power.

One possibility raised by the current findings is that gesture might not really be true 'gesture' if the learner performing it does not understand its meaning. There is some recent work showing that without enough contextual cues, like the presence of objects, children are likely to interpret another person's gestures as movements for their own sake, or movements with no communicative or instructive purpose (Wakefield, Novack & Goldin-Meadow, 2016). Even adults are less likely to interpret something as a purposeful gesture if it is not accompanied by speech or if there are no objects present (Novack, Wakefield & Goldin-Meadow, 2016). Based on this work, it is likely that children in the read-off strategy group did not see the gesture as a gesture, per se, but rather as a meaningless or irrelevant hand movement. Irrelevant movements do not show the same learning benefits as meaningful, relevant gestures (Brooks & Goldin-Meadow, 2015). This hypothesis is further supported by coding in Experiment 2, which showed that children who were asked to produce the thumb-and-forefinger gesture without having ever seen the action, were very likely to perform the movement incorrectly or hesitantly. This type of gesture analysis, which focuses on the form and trajectory of the hand rather than on the intended representation, has recently been used as a way to mark the moment of insight when a child learns how to properly solve a mathematical equivalence problem (Harden, 2015). In this work, the physical properties of the child's own hand movements reflect the child's own understanding of what the movements, or gestures, are intended to represent. Taken together, there is reason to believe that children with lower prior knowledge in the current study are conceptualizing the gesture differently than their higher prior knowledge peers, which explains the lower learning rates in the gesture-based training conditions.

Part III: Beyond Measurement

Overall, the current dissertation shows an advantage for action-based instruction during both learning and transfer on a linear measurement task. How can this finding be reconciled with previous work showing the power of gesture to promote learning (see Novack & Goldin-Meadow, 2015 for a review), and some of my own work showing that abstract gesture may promote transfer even more so than action (Novack, Congdon, Hemani-Lopez & Goldin-Meadow, 2014)?

First, it is helpful to set aside the question of transfer for a moment to consider immediate learning outcomes in any study that has directly compared action and gesture within the same learning paradigm. In Novack et al., 2014, a child's prior knowledge, as predicted by the number of speech-gesture mismatches (see Church & Goldin-Meadow, 1986 for a definition of a mismatch), did not significantly interact with training condition to predict learning outcomes on a mathematical equivalence task (Novack et al., 2014). However, as an author on this paper, I used the child's prior knowledge as an a priori hypothesis to explore the learning patterns further and found that children with lower prior knowledge, who did not produce any pre-test mismatches, learned significantly less from the abstract gesture instruction session than their higher prior knowledge mismatching peers. Both groups of children learned equally well from action-based instruction. In a replication of that study, which also tested the effect of sequential action and gesture instruction in the domain of mathematical equivalence, I found the same pattern. Children with lower prior knowledge did not learn as much from the abstract gesture condition as their higher prior knowledge peers (Congdon, Novack & Goldin-Meadow, in prep). Moreover, that same study showed that lower prior knowledge children learned more from gesture-*then*-action than they did from action-*then*-gesture, which mirrors the pattern of results reported in Experiment 2 of the current dissertation. Overall, these findings suggest a robust and generalizable phenomenon whereby children with lower prior knowledge struggle to learn from gesture, but if that gesture is immediately followed by an action demonstration, the learning rates recover nicely. Even outside of the domain of mathematics, there is some recent evidence from a linguistic learning task that children with higher phonological knowledge showed a learning boost from gesture-based instruction, while their peers with lower phonological knowledge did not (Wakefield & James, 2015).

The patterns of results in the domain of mathematical equivalence start to diverge from that of the current dissertation when you look beyond immediate learning to consider the question of transfer. In the domain of linear measurement, there are either no differences in transfer based on movement type, or there is a slight advantage of action-based instruction. The one exception to this rule is the sequential gesture-then-action training condition in Experiment 2, in which gesture seems to set the stage for action and together they promote learning *and* transfer. By contrast, children who learn from gesture (or gesture-then-action) in each of the mathematical equivalence paradigms show stronger transfer than children who learn from action alone (Novack et al., 2014; Congdon et al, in prep).

There are many differences between the current dissertation and the mathematical equivalence paradigm. For example, prior knowledge is characterized differently (mismatching versus problem-solving strategy) and transfer is assessed differently across the two sets of studies (different formats of the same problem versus entirely new unit-based tasks). These procedural differences could be partially responsible for the dissociation in patterns of transfer after action and gesture instruction that are reported above. If so, we might start to see similar patterns of transfer with minor adjustments in the procedure. However, the paradigms also differ along two theoretically important dimensions: content domain and age of the learner. As such, one possibility is that action-based instruction is better suited for learning and transfer on a concrete, spatial problem like linear measurement. Gesture-based instruction, though difficult for some learners, could be better suited for learning and transfer on a more abstract, symbolic problem like mathematical equivalence.

Alternatively, it could be the case that action is generally more appropriate for younger children and gesture is more appropriate for older children, irrespective of domain. In the current dissertation, all children were in first grade, whereas children in the mathematical equivalence studies were in third or fourth grade. Recall Piaget's claim that children must first learn in concrete situations before they can succeed with symbolic representations (Piaget, 1951). Piaget argued for a stage-theory of cognitive development, which meant that in his view, this transition from concrete to abstract representations was happening over ontogenetic, developmental time. There is some evidence from current research that also supports this general age-related hypothesis. For example, we know that young children can interpret other people's failed actions before they can interpret other people's gestures (Novack, Goldin-Meadow & Woodward, 2015), and that neural correlates of gesture processing change over developmental age, potentially reflecting more gesture-processing expertise with age (Wakefield, James & James, 2013). More research is needed to test these hypotheses.

Part IV: Open Questions

One question raised by the current dissertation is whether a different iconic gesture would have been more effective than the thumb-and-forefinger gesture. After an informal piloting study with adults, this particular gesture was selected because it was spontaneously produced to describe a small measure of spatial extent. In addition, there is some previous work showing that children as young as 2.5 years old can map this exact gesture to the size of an object (Novack, Filippi, Goldin-Meadow & Woodward, 2016). In the current dissertation, children occasionally spontaneously produced the thumb-and-forefinger gesture when explaining their reasoning during the experiment, even when they were not in the gesture-based training condition. Finally, after this gesture produced interesting individual variability in learning outcomes in Experiment 1, it seemed logical to continue to use the same iconic gesture across all subsequent experiments.

From a practical perspective, if an educator were tasked with selected the 'best' possible gesture for teaching children about units of measure, there is reason to think that the thumb-and-forefinger gesture would be a poor choice. To start, we know from Experiment 2 that the physical form of the gesture was difficult for some students to produce, particularly those who did not seem to have clear access to its representational content. Second, there is a possibility that the gesture misled children into focusing on the lines of the ruler rather than the space, as both the thumb and forefinger lie directly on a hatch-mark to frame the unit space. Some support for this argument comes from the fact that some read-off children did switch to a hatch-mark strategy after training. Future work could investigate whether another form of iconic gesture, such as a single point to the unit space or a sweeping point along the extent of the space, would create this same shift from the read-off strategy to the hatch-mark strategy.

Despite these potential pitfalls, there is also reason to believe that the thumb-and-forefinger gesture *was* a good selection. For example, some recent descriptive work investigating conceptual learning trajectories in the domain of linear measurement argues that the most advanced state of conceptual understanding is one that incorporates all of the components of the ruler, including the hatch-marks (Barrett et al., 2012). The authors argue that asking children to ignore the lines on the ruler to focus on the spaces leads to an incomplete conceptual understanding. By this metric, the thumb-and-forefinger gesture, which incorporates the lines, the space, and a counting routine (the child says "one" while touching two hatch marks and framing the first unit and "two" while framing the second unit, etc…), could actually be a very powerful way to learn about the units of a ruler, so long as the child has the necessary context to interpret the movement. In other words, this more challenging gesture may have been driving the desirable difficulties effect reported in Experiment 2, an effect we might not have observed with a less complex gesture (such as a point to the blank space). Lastly, children in the read-off group switched to the hatch-mark strategy across *all* training conditions, and arguably did so even more often in the action-based training conditions, which makes it difficult to claim that the gesture was causing this strategy shift. Future research is required to investigate the efficacy of other forms of iconic 'unit' gestures.

Part V: Broader Impact and Conclusions

Overall, this dissertation has practical implications for how to improve children's understanding of linear measurement through exposure to shifted-object problems and action-based training. It also contributes to the existing literature on two of the most pervasive linear measurement misconceptions, reading off and counting hatch marks, by characterizing children's pre-test performance on number of different tasks and testing the effect of various interventions on children's learning trajectories. Finally, from a theoretical perspective, this dissertation serves as a powerful demonstration of why it is crucial to consider the characteristics of the learner when evaluating the efficacy of gesture and action in an instructional context. Children with lower prior knowledge and lower spatial working memory need additional scaffolding to extract the full benefit of gesture's power.

Future work will investigate other factors that influence this effect including the age of the learner, the content domain of the target concept, and the exact form of the representational movements. I also plan to test the effect of *doing* the actions versus *seeing* the actions, and I would like to compare action and gesture on a task that is more skill-based, rather than based on insight. In considering all of these dimensions, my ultimate goal is to better understand the powers and limitations of actions and gestures while building a unified theoretical framework to make accurate a priori predictions about the efficacy of these movement-based interventions.

REFERENCES

Alexander, K.L, Entwisle, D.R. & Olson, L.S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis,* 23(2), 171-191.

Alibali, M. W., Flevares, L. M., & Goldin-Meadow, S. (1997). Assesing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology, 89*, 183-193.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468-523.

Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Jacobs, S. A., Maritinex, C. J., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition and Instruction, 32*, 65-100.

Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science, 22*, 1138-1144.

Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, *25*(2), 92-98.

Andres, M., Seron, X., & Olivier, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience*, 19, 563–576 .

Ashcraft, M. H., & Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*, *14*(2), 243-248.

Atit, K., Resnick, K., Shipley, T. F., Ormand, C. J., Manduca, C. A., Goksun, T., & Tikoff, B. (2013). Spatial gestures point the way: A broader understanding of the gestural referent. *Proceedings of the 35th Annual Cognitive Science Society*.

Baddeley, A. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829-839.

Ball, D. (1992). Magical hopes: Manipulatives and the reform of math education. *American Educator*, 16, 14-18.

Barr, R. (2010). Transfer of learning between 2D and 3D sources during infancy: Informing theory and practice. *Developmental Review*, *30*(2), 128-154.

Barrett, J., Clements, D., Sarama, J., Cullen, C., McCool, J., Witkowski, C., & Klanderman, D. (2012). Evaluating and Improving a Learning Trajectory for Linear Measurement in

Elementary Grades 2 and 3: A Longitudinal Study. *Mathematical Thinking and Learning*, 14(1), 28-54.

Barrouillet, P., Bernardin, S. & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology. General* 133 (1): 83–100.

Beattie, G., Webster, K., & Ross, J. (2010). The Fixation and Processing of the Iconic Gestures That Accompany Talk. *Journal of Language and Social Psychology, 29*, 194-213.

Beaudoin-Ryan, L., & Goldin-Meadow, S. (2014). Teaching moral reasoning through gesture. *Developmental Science, 17*, 984-990.

Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101-105.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psycholgical Science, 21*, 1605-1610.

Bjork, R.A. (1994). *Memory and metamemory considerations in the training of human beings*. In J. Metcalfe & A. Shimamura (Eds.), Metacognition: Knowing about knowing (pp. 185-205). Cambridge, MA: MIT Press

Blume, G. W., Galindo, E., & Walcott, C. (2007). *Performance in measurement and geometry from the viewpoint of Principles and Standards for School Mathematics*. Results and interpretations of the 2003 Mathematics Assessment of the National Assessment of Educational Progress, 95-138.

Brooks, N., & Goldin-Meadow, S. (2015). Moving to learn: How guiding the hands can set the stage for learning. *Cognitive Science,* 1-19.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, MA: Harvard University Press.

Bruner. J. S., Olver, R. R., & Greenfield, P.M. (1966). *Studies in cognitive growth*. New

York: Wiley.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273-293.

Butler, A. J., & James, K. H. (2013). Active learning of novel sound-producing objects: motor reactivation and enhancement of visuo-motor connectivity. *Journal of Cognitive Neuroscience*, *25*(2), 203-218.

Butler, A. J., James, T. W., & James, K. H. (2011). Enhanced multisensory integration and motor reactivation after active motor learning of audiovisual associations. *Journal of Cognitive Neuroscience*, *23*(11), 3515-3528.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147.

Casasola, M. (2005). When less is more: how infants learn to form an abstract categorical representation of support. *Child Development*, 76, 279–290.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*, 706-723.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index on transitional knowledge. *Cognition, 23*, 43-71.

Clark, A. (2007). Re-inventing ourselves: The plasticity of embodiment, sensing, and mind. *Journal of Medicine and Philosophy*, *32*(3), 263-282.

Cohen, R.L. (1983). The effect of encoding variables on the free recall of words and action events. *Memory & Cognition, 11,* 575-582.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin-Meadow, S. (under review). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention.

Congdon, E.L., Novack, M.A. & Goldin-Meadow (in prep) Learning math through action and gesture: Do children benefit from stepwise instruction?

Cook, S. W., Duffy, R. G., & Fenn, K. M. (2013). Consolidation and transfer of learning after observing hand gesture. *Child Development, 84*, 1863-1871.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047-1058.

Cook, S. W., Yip, T. K. Y., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *27*(4), 594–610.

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(02), 891B.

Cronbach, L. & Snow, R. (1977). *Aptitudes and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington.

DeLoache, J. S. (2000). Dual representation and children's use of scale models. *Child Development*, 71, 329-338.

Dewey, John (1916). *Democracy and education: An introduction to the philosophy of education.* New York: The Macmillan Company.

Ekman, P., & Friesen, W. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49-98.

Engelkamp, J. (1991). *Memory of Action Events: Some Implications for Memory Theory and for Imagery*. In C. Cornoldi & M. A. McDaniel (Eds.), *Imagery and Cognition* (pp. 183-219). New York: Springer-Verlag.

Friedman, M. (1978). The manipulative materials strategy: The latest pied piper? *Journal for Research in Mathematics Education,* 9, 78-80.

Fuson, K. C. & Briars, D. J. (1990). Using a base-ten blocks learning/teaching approach for first – and second-grade place-value and multi-digit addition and subtraction. *Journal for Research in Mathematics Education, 21*, 180-206.

Foy, P., Arora, A. & Stanco, G.M. (Eds.). (2013). *TIMSS 2011 User Guide for the International Database*. Lynch School of Education, Boston College: TIMSS and PIRLS International Study Center.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review*, *26*(1), 9-25.

Gerson, S. A., Bekkering, H., & Hunnius, S. (2015). Short-term motor training, but not observational training, alters neurocognitive mechanisms of action processing in infancy. *Journal of Cognitive Neuroscience, 27(6),* 1207-1214.

Gentner, D. (2003). *Why we're so smart*. In D. Gentner & S. Goldin-Meadow (Eds.), Language in mind: Advances in the study of language and thought (pp. 195–236). Cambridge, MA: MIT Press.

Ginsburg, Herbert P. (1989). *Children's Arithmetic- How They Learn It and How You Teach It*. Austin, Texas: Pro·ed.

Glenberg, A. M. (2008). Embodiment for education. *Handbook of cognitive science: An embodied approach*, 355-372.

Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard University Press.

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

Goldin-Meadow, S., Cook, S. W., & Mitchell, Z. (2009). Gestures gives children new ideas about math. *Psychological Science, 20*, 267-271.

Goldin-Meadow, S., Kim, S., & Singer, M. (1999). What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology, 91*, 720-730.

Goldin-Meadow, S., Nusbaum, H., Kelly, S.D. & Wagner, S. (2001) Explaining math: Gesturing lightens the load. *Psychological Science*. 12 (6):516–522.

Goldstone, R. L. & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46*, 414-466.

Goldstone, R. L. & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences*, 14, 69-110.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics and Cognition, 14*, 53-82.

Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., & Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition, 144*, 14-28.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, *44*(4), 339-387.

Harden, H. L. (2015). *From movement to gesture: How the body reflects and enhances learning*. Symposium talk presented at the annual inter-Science of Learning Centers Conference, San Diego, CA.

Hoff, E. (2003), The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, 74: 1368–1378.

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage, 49*, 875-884.

Holmes, J & Adams, J.W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology,* 26(3).

Horst, J., Twomey, K., & Ranson, S. (2013). Encountering multiple exemplars during fast mapping facilitates word learning. *Cognitive Science Society*.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gesture as simulated action. *Psychonomic Bulletin & Review*, 15, 495-514.

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, *123*(3), 284.

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

James, K. H. (2010). Sensori‐motor experience leads to changes in visual processing in the developing brain. *Developmental Science*, *13*(2), 279-288.

James, K. H., & Swain, S. N. (2011). Only self‐generated actions create sensori‐motor systems in the developing brain. *Developmental Science*, *14*(4), 673-678.

Kalyuga, S. (2005). *Prior knowledge principle in multimedia learning*. In R. E. Mayer (Ed.), Cambridge handbook of multimedia learning (pp. 325-337). New York: Cambridge University Press.

Kamii, C., & Clark, F. (1997). Measurement of length: The need for a better approach to teaching. *School Science and Mathematics*, 97, 116-121.

Kamii, C., Lewis, B. A., & Kirkland, L. (2001). Manipulatives: When are they useful? *Journal of Mathematics Behavior,* 20, 21-31.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2005). Relevant concreteness and its effects on learning and transfer. In B. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the XXVII Annual Conference of the Cognitive Science Society*.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2006a). Effects of concreteness on representation: An explanation for differential transfer. In R. Sun & N. Miyake (Eds.), *Proceedings of the XXVIII Annual Conference of the Cognitive Science Society.*

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2006b). Do children need concrete instantiations to learn an abstract concept? In R. Sun & N. Miyake (Eds.), *Proceedings of the XXVIII Annual Conference of the Cognitive Science Society.*

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives*, *3*(3), 151-155.

Kastens, K. A., Agrawal, S., & Liben, L. (2007). Research methodologies in science education: The role of gestures in geoscience teaching and learning. *Journal of Geoscience Education: Research Methodologies in Science Education column*.

Kelly, S. D., Healy, M., Ozyurek, A., & Holler, J. (2014). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review*.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Chicago, IL: The University of Chicago Press.

Kita, S., & Ozyurek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16-32.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, *342*(6161), 935-937.

Kontra, C., Goldin‐Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science*, *4*(4), 731-739.

Kontra, C., Lyons, D., Fischer, S., & Beilock, S. L. (2012). *Embodied Physics Learning* Poster session presented at the annual Inter-Science of Learning Center meeting, San Diego, CA.

Kwon, M., Ping, R., Congdon, E.L., Levine, S.C. (under review) *Overturning children's misconceptions about ruler measurement units: The power of disconfirming evidence.*

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

Landry, S.H., Denson, S.D., & Swank, P.R. (1997) Effects of medical risk and socioeconomic status on the rate of change in cognitive and social development for low birth weight children. *Journal of Clinical and Experimental Neuropsychology*, 19(2), 261-274.

Lee, J. (2002). Racial and Ethnic Achievement Gap Trends: Reversing the Progress Towards Equity? *Educational Researcher*, 31(1), 3-12.

Lehrer, R., Jenkins, M., Osana, H. (1998). *Longitudinal study of children's reasoning about space and geometry*. In R. Lehrer and D. Chazan (Eds.), Designing learning environments for developing understanding of geometry and space (pp. 137-167). Mahwah, NJ: Lawrence Erlbaum.

Levine, S.C., Kwon, M., Huttenlocher, J., Ratliff, K., & Dietz, K. (2009). *Children's understanding of ruler measurement: A training study*. Proceedings of the 31st Cognitive Science Society Meeting, Liels Taatgen and Hedderik van Rijn (Eds), Amsterdam.

Levine, L. C., Goldin-Meadow, S., Carlson, M., & Hemani- Lopez, N. (under review). Mental transformation skill in young children: Action and gesture training promote learning over different time frames.

Lindquist, M. M., & Kouba, V. L. (1989). *Measurement*. In M. M. Lindquist (Ed.), Results from the fourth mathematics assessment of the National Assessment of Educational Progress (pp. 35-43). New York, NY: Teachers College Press.

Luciano, C., Becerra, I. G. & Valverde, M. R. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of Experimental Analysis and Behavior*, 87: 349–365.

Macedonia, M., Muller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32*, 982-998.

Maguire, M. J., Hirsh‐Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: Fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628-634.

Manches, A. & O'Malley, C. (2012). Tangibles for learning: a representational analysis of physical manipulation. *Pers Ubiquit Comput,* 16, 405-419.

Martin, W. G., & Strutchens, M. E. (2000). *Geometry and measurement*. Results from the seventh mathematics assessment of the National Assessment of Educational Progress, 193-234.

McClelland, J. L. (1994). The organization of memory. A parallel distributed processing perspective. *Revue Neurologique*, *150*, 570-570.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: The University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press.

McNeil, N., & Fyfe, E. (2012). "Concreteness fading" promotes transfer of mathematical knowledge. *Learning and Instruction* 22(6). 440-448

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, 19, 171-184.

McNeil, N. M., & Uttal, D. H. (2009). Rethinking the use of concrete materials in learning: Perspectives from development and education. *Child Development Perspectives*, 3(3), 137-139.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, *27*, 272–277.

Mix, K. S. (2010). *Spatial tools for mathematical thought*. In K. S. Mix, L. B. Smith & M. Gasser (Eds.) In Space and Language. New York: Oxford University Press.

Montessori, M. (1909). *The Montessori Method*. Radford, VA: Wilder Publications.

Montessori, M. (1995). *The Absorbent Mind*. Macmillan.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology*, 411-424.

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics, 47*, 175-197.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J. & Chrostowski, S.J. (2004). *Findings From IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002-1005.

Nilsson, L.-G., & Cohen, R.L. (1988). *Enrichment and generation in the recall of enacted and non-enacted instruction*s. In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds.), Practical aspects of memory: Current research and issues (pp. 427-432). Chichester: John Wiley & Sons.

Novack, M. A., Congdon, E. L., Hemani-Lopez, N., & Goldin-Meadow, S. (2014). From action to abstraction using the hands to learn math. *Psychological Science*, *25*(4), 903-910.

Novack, M.A., Filippi, C, Goldin-Meadow, S. & Woodward, A (2016). *Actions speak louder than words when you are 2 years old*. Manuscript in preparation.

Novack, M., & Goldin-Meadow, S. (2015). Learning from gesture: how our hands change our minds. *Educational Psychology Review*, *27*(3), 405-412.

Novack, M. A., Goldin-Meadow, S., & Woodward, A. L. (2015). Learning from gesture: How early does it happen? *Cognition*, *142*, 138-147.

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition, 146*, 339-348.

Pagulayan, K.F., Busch, R.M., Medina, K.L, Bartok, J.A. & Krikorian, R. (2006). Developmental Normative Data for the Corsi Block- Tapping Task, *Journal of Clinical and Experimental Neuropsychology*, 28(6), 1043-1052.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3*, 359-400.

Petersen, L. A., & McNeil, N. M. (2008). *Using perceptually rich objects to help children represent number: Established knowledge counts*. In V. Sloutsky, B. Love, & K. McRae (Eds.), Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society (p. 1567- 1572). Mahwah, NJ: Erlbaum.

Peterson, S. K., Mercer, C. D. & O'Shea, L. (1988). Teaching learning disabled students place value using the concrete to abstract sequence. *Learning Disabilities Research, 4,* 52-56.

Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.

Piaget, J. (1953). How children form mathematical concepts. *Scientific American*, 189 (5), 74-79.

Piaget, J. (1972). *Development and learning*. Lavatelly, C. & Stendler, F. (Eds.) Reading in child behavior and development. New York: Hartcourt Brace Janovich.

Ping, R., Decatur, M.-A., Larson, S. W., Zinchenko, E., & Goldin-Meadow, S. (under review). Unpacking the gestures of chemistry learners: What the hands tell us about correct and incorrect conceptions of stereochemistry.

Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychology, 44*, 1277-1287.

Ping, R.M. & Goldin-Meadow, S. (2010) Gesturing saves cognitive resources when talking about non-present objects. *Cognitive Science*, 34, 602–619.

Ping, R., Goldin-Meadow, S., & Beilock, S. L. (2014). Understanding gesture: Is the listener's motor system involved? *Journal of Experimental Psychology: General, 43*, 195-204.

Program for International Student Assessment. Retrieved from https://nces.ed.gov/surveys/pisa /pisa2012/index.asp

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rader, N. d. V., & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. *Language Sciences, 34*, 559-568.

Ramscar, M., Dye, M., Popick, H.M., O'Donnell-McCarthy, F. (2011). The Enigma of Number: Why Children Find the Meanings of Even Small Number Words Hard to Learn and How We Can Help Them Do Better. *PLoS ONE 6*(7): e22501.

Raymond, G. & Gibbs, J. (2006). *Embodiment and cognitive science*. New York: Cambridge University.

Reardon, S. (2002). *The widening academic achievement gap between the rich and the poor: New evidence and possible explanations*. In G. J. Duncan, R. J. Murnane (Ed.), Whither Opportunity?: Rising Inequality, School, and Children's Life Chances (pp. 91-111). Russell Sage Foundation.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.*Classical conditioning II: Current research and theory*, *2*, 64-99.

Resnick, L. B. & Omanson, S. F. (1987). *Learning to understand arithmetic*. In R. Glaser (Ed.) Advances in instructional psychology (Vol. 3), 41-95. Hillsdale, N. J.: Erlbaum.

Rohlfing, K. J., Longo, M. R., & Bertenthal, B. I. (2012). Dynamic pointing triggers shifts of visual attention in young infants. *Developmental Science, 15*, 426-435.

Roth, W.-M. (2007). *Making use of gestures, the leading edge in literacy development*. In W. Saul (Ed.), Communicating science: Examining the discourse.

Roth, W.-M., & Lawless, D. (2002). Science, culture, and the emergence of language. *Science Education*.

Saltz, E. (1988) *The role of motoric enactment (M-processing) in memory for words and sentences.* In M.M. Gruneberg, P.E. Morris & R.N. Sykes (Eds.), Practical aspects of memory: Current research and issues (pp. 427-432). Chichester: John Wiley & Sons.

Sarama, J. & Clements, D. (2009). Concrete computer manipulatives in mathematics education. *Child Development Perspectives, 3(3)* 145-150.

Schachner, A., & Carey, S. (2013). Reasoning about 'irrational' actions: When intentional movements cannot be explained, the movements themselves are seen as the goal. *Cognition, 129*, 309-327.

Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude–treatment–interaction study. *Learning and Instruction*, *19*(1), 28-42.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychological science*, *14*(3), 237-250.

Siegler, R.S. & Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/cross-sectional study of class inclusion. *Child Development*, 77(4), 997-1015.

Siler, S.A. & Willows, K.J (2014). Individual differences in the effect of relevant concreteness on learning and transfer of a mathematical concept. *Learning and Instruction, 33,* 170-181.

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

Smiley, P., & Huttenlocher, J. (1995). *Conceptual development and the child's early words for events, objects, and persons*. In M. Tomasello & W.E. Merriman (Eds.), Beyond names for things: Young children's acquisition of verbs (pp. 21–62). Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, J., Dietiker, L., Lee, K., Males, L.M., Figuera, H., Mosier, A., … Sisman, G. (2008). *Framing the analysis of written measurement curricula*. Paper presented at the Annual Meeting of the American Educational Research Association.

Smith, J. P., Males, L. M., Dietiker, L. C., Lee, K., & Mosier, A. (2013). Curricular Treatments of Length Measurement in the United States: Do They Address Known Learning Challenges?. *Cognition and Instruction*, *31*(4), 388-433.

Smith, L. B. (2005). Cognition as a dynamic system: Principles from embodiment. Developmental Review, 25(3), 278-298.

Smith, L., Ping, R., Matlen, B., Goldwater, M., Gentner, D., & Levine, S.C. (2014) Mechanisms of Spatial Learning: Teaching Children Geometric Categories. *Spatial Cognition*, Bremen, Germany.

Snow, R. (1989). *Aptitude-Treatment Interaction as a framework for research on individual differences in learning*. In P. Ackerman, R.J. Sternberg, & R. Glaser (ed.), Learning and Individual Differences. New York: W.H. Freeman

Solomon, T. L., Vasilyeva, M., Huttenlocher, J., & Levine, S. C. (2015). Minding the gap: Children's difficulty conceptualizing spatial intervals as linear measurement units. *Developmental Psychology*, *51*(11), 1564.

Sommerville , J. A. Woodward , A. L. (2010). *The link between action production and action processing in infancy.* In F. Grammont D. Legrand P. Livet (Eds.), Naturalizing intention in action. (pp. 67-90). Cambridge, MA: MIT Press.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-B11.

Sophian, C. (2003). Learning about "one": Units as a cornerstone for Head Start mathematics. *Teaching Children Mathematics, 10*(4), 210-215.

Sophian, C. (2007). *The Origins of Mathematical Knowledge in Childhood***.** New York:

Lawrence Erlbaum Associates.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*, 498-505.

Stevenson, H. W. & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Simon & Schuster.

Stieff, M. (2011). When is a molecule three dimensional? A task-specific role for imagistic reasoning in advanced chemistry. *Science Education, 95*, 310-336.

Suydam, M.N. & Higgins, J.L. (1977). *Activity-based learning in elementary school mathematics: Recommendations from research*. ERIC Information Analysis Center for Science, Mathematics, and Environmental Education, Columbus, OH.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *96*(3), 471.

The National Science Education Standards (1996). Washington, D.C.: National Academy Press.

TIMSS 2011 Assessment. Copyright © 2012 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, Chestnut Hill, MA and International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat, Amsterdam, the Netherlands.

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations.*Child Development Perspectives*, *3*(3), 156-159.

Uttal, D. H., Scudder, K. V., & DeLoache, J. S. (1997). Manipulatives as symbols: A new perspective on the use of concrete objects to teach mathematics. *Journal of Applied Developmental Psychology*, 18, 37-54.

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

Wagner, S.M., Nusbaum, H., Goldin-Meadow, S. (2004) Probing the Mental Representation of Gesture: Is Handwaving Spatial? *Journal of Memory & Language*. 50(4): 395–407.

Wakefield, E. M., Congdon, E., Novack, M., Goldin-Meadow, S., & James, K. H. (under review). Learning math through your hands: The neural effects of gesture-based instruction in 8-year-old children.

Wakefield, E. M., & James, K. H. (2011). Effects of sensori-motor learning on melody processing across development. *Cognition, Brain, Behavior. An Interdisciplinary Journal, 15*, 505-534.

Wakefield, E. M., & James, K. H. (2015). Effects of learning with gesture on children's understanding of a new language concept. *Developmental Psychology, 5*, 1105-1114.

Wakefield, E. M., James, T. W., & James, K. H. (2013). Neural correlates of gesture processing across human development. *Cognitive Neuropsychology, 30*, 58-76.

Wakefield, E.M., Novack, M.A. & Goldin-Meadow, S. (in prep) Unpacking the ontogeny of gesture understanding: How movement becomes meaningful across development.

Wearne, D. & Hiebert, J. (1988). A cognitive approach to meantinful mathematics instruction: thesting a local theory iusing decimanl numbers. *Journal for Research in Mathematics Education, 19*, 371-384.

Wilensky, U. (1991). *Abstract meditations on the concrete and concrete implications for mathematics education*. In I. Harel & S. Papert (Eds.) Constructionism. Norwood N.J.: Ablex Publishing Corp.

Wilson , M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, 9, 625-636.

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage, 47*, 1992-2004.