CLASSROOM PRESCHOOL SCIENCE LEARNING: THE LEARNER,

INSTRUCTIONAL TOOLS, AND PEER-LEARNING ASSIGNMENTS

By

JAMIE M. REUTER

A dissertation submitted to the

Graduate School- New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Education

Written under the direction of

Rochel Gelman

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2017

ABSTRACT OF THE DISSERTATION

Classroom Preschool Science Learning: The Learner, Instructional Tools, and Peer-

Learning Assignments

by JAMIE M. REUTER

Dissertation Director:
Rochel Gelman

The recent decades have seen an increased focus on improving early science education. Goals include helping young children learn about pertinent concepts in science, and fostering early scientific reasoning and inquiry skills (e.g., NRC 2007, 2012, 2015). However, there is still much to learn about what constitutes appropriate frameworks that blend science education with developmentally appropriate learning environments. An important goal for the construction of early science is a better understanding of appropriate learning experiences and expectations for preschool children. This dissertation examines some of these concerns by focusing on three dimensions of science learning in the preschool classroom: (1) the learner; (2) instructional tools and pedagogy; and (3) the social context of learning with peers.

In terms of the learner, the dissertation examines some dimensions of preschool children's scientific reasoning skills in the context of potentially relevant, developing general reasoning abilities. As young children undergo rapid cognitive changes during the preschool years, it is important to explore how these may influence scientific

thinking. Two features of cognitive functioning have been carefully studied: (1) the demonstration of an epistemic awareness through an emerging theory of mind, and (2) the rapid improvement in executive functioning capacity. Both continue to develop through childhood and adolescence, but changes in early childhood are especially striking and have been neglected as regards their potential role in scientific thinking. The question is whether such skills relate to young children's capacity for scientific thinking.

Another goal was to determine whether simple physics diagrams serve as effective instructional tools in supporting preschool children's scientific thinking. Specifically, in activities involving predicting and checking in scientific contexts, the question is whether such diagrams facilitate children's ability to accurately recall initial predictions, as well as discriminate between the outcome of a scientific manipulation and their original predictions (i.e., to determine whether one's predictions were confirmed).

Finally, this dissertation also explores the social context of learning science with peers in the preschool classroom. Due to little prior research in this area, it is currently unclear whether and how preschool children may benefit from working with peers on science activities in the classroom. This work aims to examine preschoolers' collaboration on a science learning activity, as well as the developmental function for such collaborative skills over the preschool years.

ACKNOWLEDGEMENT

TABLE OF CONTENTS

Chapter 3: Preschoolers' Scientific Reasoning: Simple Physics Diagrams of a

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Statement of the Problem.

There have been a number of recent calls (i.e., from schools, government, and industry), for improved and more innovative approaches to science education. This is motivated by the regular demands we face in our increasingly technological society. For example, all citizens need skills for evaluating the array of information they confront each day. Thus, identifying ways to enhance scientific and technological literacy represents a crucial research and educational agenda. The goal is to promote a generation of scientifically literate citizens who understand enough about the disciplinary content and are capable of logical, abstract, and analytical thinking (i.e., Gropen, Clark-Chiarelli, Ehrlich, & Thieu, 2011; Rutherford & Ahlgren, 1989).

A growing number of researchers argue that young children should be introduced to STEM (science, technology, engineering, and mathematics) methods and topics as early as during the preschool years (i.e., Bowman, Donovan, & Burns, 2001; Duschl, Schweingruber, & Shouse, 2006; Early Childhood STEM Working Group, 2017; Gelman & Brenneman, 2012; McClure et al., 2017; NRC, 2015; NSTA, 2014). It is assumed that through early experiences in science, young children learn about pertinent concepts and start to acquire skills they need to think critically and continue learning throughout the school years. As a result, science instruction is being incorporated in increasingly younger age levels, including preschool. Indeed, most states in the U.S. have now outlined learning expectations for preschool science (Brenneman, 2011). These learning expectations involve supporting children's acquisition of conceptual knowledge in scientific domains, as well as scientific reasoning and inquiry skills as children observe and investigate scientific phenomena in the world around them (e.g., NRC, 2007; NRC,

2012). Further, a number of recently developed science-based preschool curricula have been shown to be effective in supporting these goals (i.e., French, 2004; Gelman, Brenneman, Macdonald, & Roman, 2009; Quinn, Taylor, & Taylor, 2004). Overall, there seems to be consensus on the notion that good science programs engage children in a manner that builds on their natural curiosity, fosters their dispositions for active learning, and supports scientific process and inquiry skills (Chalufour & Worth, 2006).

Despite this progress, young children are still often not provided adequate educational experiences in science. Science tends not to be emphasized in the professional learning of educators of young children. Many early childhood teachers experience anxiety and low self-confidence about STEM topics, and as a result, they tend to spend little time in science-related activities (Brenneman, Stevenson-Boyd, & Frede, 2009; Brown, 2005; McClure et al., 2017; Nayfeld, Brenneman, & Gelman, 2011; Tu, 2006). As follows, there is still much to learn about what constitutes appropriate frameworks that blend science education with developmentally appropriate learning environments. Programs that are made up of push-down curricula are likely to fail. Both the child and the teacher are likely to hold alternative conceptions in science (i.e., Anderson & Mitchener, 1994; Driver, 1989; Gilbert & Watts, 1983; Wandersee, Mintzes, & Novak, 1994). Therefore, an important goal for the construction of early science is a better understanding of appropriate learning experiences and expectations for preschool children. The three studies in this dissertation examine some of these concerns by focusing on three dimensions of science learning in the early childhood classroom: (1) the learner, (2) instructional tools and pedagogy, and (3) developmentally-appropriate contexts for supporting science learning (both individually, and with peers).

In terms of the learner, one goal for this dissertation is to examine some dimensions of young children's early scientific reasoning and inquiry skills in the context of potentially relevant, developing general reasoning abilities. As young children undergo rapid cognitive changes during the preschool years, it is important to explore how these cognitive changes may influence scientific thinking. Two features of cognitive functioning have been carefully studied: (1) the demonstration of an epistemic awareness through an emerging theory of mind (i.e., Flavell, 1999; Perner, Leekam, & Wimmer, 1987), and (2) the rapid improvement in executive functioning capacity (i.e., De Luca & Leventer, 2008; Espy, 2004). Both continue to develop through childhood and adolescence, but changes in early childhood are especially striking and have been neglected as regards their potential role in scientific reasoning. This follows from the general stage theory assumption that preschool children are perception bound and unable to learn and think about the abstract concepts involved in science.

There is reason to propose that both executive function and theory of mind relate to children's acquisition of scientific thinking skills. Science learning experiences often require that children engage in processes such as attending to the constraints of a current experiment, manipulating information, and incorporating new findings into existing knowledge structures. It is very likely that these skills put substantial demands on executive functioning capacities, such as inhibition control, attention, working memory and cognitive flexibility (Nayfeld, Fuccillo, & Greenfield, 2013). Further, key theory of mind developments include the ability to distinguish between external reality and mental events, as well as to understand false beliefs (e.g., Baron-Cohen, Leslie, & Frith, 1985; Gopnik & Astington, 1988; Wimmer & Perner, 1983). Such understandings likely relate

to one's capacity for scientific thinking, as such often requires distinguishing between one's predictions (mental events) and the observed outcomes of experimental manipulations (external reality). Further, lacking a false belief understanding is likely associated with difficulty in recognizing when one's initial predictions are incorrect, or not confirmed by scientific manipulations. There is currently theoretical work supporting the notion that theory of mind and executive function skills relate to scientific thinking (as described in further detail below). However, little empirical work directly examines the relationships among these variables with data from classrooms.

If theory of mind and executive functioning developments do play a role in scientific thinking, it is possible that instructional tools which buttress these capacities may aid children's acquisition of early forms of scientific thinking. Accordingly, the research in this dissertation is designed to explicitly examine the effectiveness of recording charts using simple physics diagrams as instructional tools to support preschoolers' external representation of, and reflection upon, their scientific investigations. The charts were designed with the idea that the use of multiple representations of ideas constitutes a critical aspect of the scientific process (see more below). In activities involving scientific predicting and checking, the question is whether such forms of diagrammatic representation aid children in both recalling initial predictions, as well as judging the correctness of those predictions. The rationale is that the charts may support children in attending to the relevant components of their scientific investigations (i.e., their initial predictions, and the actual observed outcomes), as well as in comparing the two, without overloading working memory capacity. Further, use of the diagrams to encourage children's reflection upon their predictions and the observed

outcomes, as well as the direct comparison of the two, may aid children in recognizing when their predictions are incorrect (i.e., especially among those developing false belief understanding). To preview, findings indicate that the diagrams do indeed support such early forms of scientific thinking in children as young as three years old.

Such forms of recording and documentation are important in doing science. For example, the Early Childhood STEM Working Group (2017) highlights the importance of incorporating discussion, visualization and other forms of representation (i.e., writing, drawing, graphing) in early STEM education to encourage learning that leads to generalization of important concepts and practices. Gelman and her colleagues have successfully introduced science notebooks to young children in their preschool science program, *Preschool Pathways to Science*, (Gelman et al., 2009; Brenneman & Lauro, 2008). In this program, children are supported in recording and documenting their scientific investigations, such as by drawing and using numerals to represent them, and having adults write what the child wants to document. Similarly, the Emilio Reggio approach encourages young children to use a variety of forms of media to document and represent ideas as they engage in scientific investigations.

Finally, this dissertation also examines the social context of learning science with peers in the preschool classroom. Science is a process, and it almost always involves collaboration with others. The Next Generation Science Standards (NGSS) are built on the key assumption that "science is fundamentally a social enterprise, and scientific knowledge advances through collaboration and in the context of a social system with well-developed norms" (NGSS Lead States, 2013, p. 27). Thus, learning with peers has been identified as a critical component of science learning and science education, with

science standards highlighting the importance of providing students opportunities for collaboration, discussion and reflection around science.

Social interaction and peer collaboration during science learning opportunities may contribute to children's science and inquiry learning in a number of ways. For example, children may benefit from joint planning of an investigation, raising questions through discussion with peers, describing outcomes and patterns in data, and explaining, sharing, discussing and listening as they work with peers on scientific investigations. It has also been argued that working with peers on science problems exposes children to a variety of forms of thinking and interpersonal interactions that likely benefit children both socially and academically (Johnson-Pynn & Nisbet, 2002). However, very little work explores collaboration in the context of science learning experiences in children younger than school age. As a result, it is currently unclear if preschool children have skills to support their ability to work with peers on science learning activities, as well as whether and how preschool children may benefit from working with peers on science activities in the early childhood classroom.

1.2 Overview of the three studies.

The presentation of the dissertation research is organized as follows:

Study 1: *Simple Physics Diagrams of an Inclined Plane Support Preschoolers' Scientific Predicting and Checking*

Study 2: *Preschoolers' Scientific Reasoning: Simple Physics Diagrams of a Balance Scale Aid their Ability to Predict and Check*

Study 3: *Preschoolers' Collaboration on a Physics Problem Involving Objects' Motion on an Inclined Plane*

In the first study, three- and four-year-old children were asked to work individually on a science activity involving predicting and checking about objects traversing an inclined plane. Children were provided a number of items, one at a time, and were asked to predict whether each would roll or slide down an inclined plane. Children then tested the items on the ramp, watched what happened, and were asked to reflect on the outcome and judge the accuracy of their initial predictions. The study aimed to: (1). explore children's ability to predict and check in this context (and to determine whether the context is developmentally-appropriate for investigating children's scientific thinking); (2). examine the relationships among children's developing theory of mind, executive functioning skills, and their early scientific thinking; and (3). verify whether a recording chart using simple physics diagrams of an inclined plane serves as an effective instructional tool for facilitating children's scientific thinking in this context. To preview, findings indicate that preschool children are indeed capable of accurately predicting and checking in this context, and further, use of the recording chart and physics diagrams does in fact support young children's scientific predicting and checking in this context. Finally, children's theory of mind and executive functioning scores were both found to positively correlate with scientific thinking scores.

The second study aimed to determine whether the results of the first study replicate in a novel science learning context, and with a new set of materials. Preschool children (ages three to five years) were asked to work individually on a science activity involving predicting and checking about the relative weights of objects using a balance scale. Children were provided pairs of familiar items and were asked to predict whether placing the items on the scale would result in the scale balancing. Children then tested

the items on the scale, watched what happened, and were asked to reflect on the outcome and judge the accuracy of their initial predictions.

To preview, findings illustrate that preschool children are indeed capable of accurately predicting and checking in this context for scientific investigation. Further, the result regarding the effectiveness of the recording chart and physics diagrams also replicated. Specifically, the option to use recording charts and relevant physics diagrams effectively supported children's predicting and checking skills. Finally, the second study also again found positive correlations among children's theory of mind and executive functioning scores, and their scientific thinking scores.

Taken together, these two studies provide evidence that, from a young age, preschool children are indeed capable of engaging in basic forms of scientific thinking in the early childhood classroom. Activities that ask children to make and check predictions in scientific contexts in which they already have some relevant knowledge are developmentally appropriate for supporting such abilities. The chosen contexts also constitute appropriate sets of conversational options, objects and diagrams that can support science learning in the early childhood classroom. Since the materials are neither too easy nor too difficult, the teacher has room to help children actively construct understanding on what they already know something about.

Finally, the third study in this dissertation examines preschool children's collaborative and communicative behaviors as they worked in pairs to solve simple problems in physical science. Children were asked to collaborate with a classmate as they explored problems focused on objects traversing an inclined plane together. Of interest was how children work with peers on problems that draw upon their knowledge

of physics within a learning setting that is age appropriate, as well as the developmental function for relevant collaborative and communicative skills over the preschool years.

To preview, results revealed that all dyads engaged in a number of collaborative and communicative behaviors as they worked together to solve the problems. For example, children explained, instructed and modeled to peers how to do things as they worked on the task. They asked one another questions and/or for assistance, and responded to such requests from peers. Children also imitated one another, as well as agreed with peers' ideas or suggestions. Further, there was little difference between the older and younger preschoolers (i.e., 3 ½ - and 4-year-olds, versus 4 ½- and 5-year-olds) in terms of the patterns of their collaborative and communicative behaviors. This indicates that children as young as 3 ½ years are capable of working together on science learning activities in the preschool classroom.

Taken together, this work indicates that it is appropriate to have young children work together on science learning activities in the early childhood classroom. Further, the chosen context is suitable for fostering children's acquisition of both scientific content knowledge and inquiry skills, as well as early collaborative and communicative abilities. In such context, children can support one another as they work together to actively construct understanding on what they already know something about. As learning with peers has been identified as a critical component of science learning and science education (i.e., science standards highlight the importance of providing students opportunities for collaboration, discussion and reflection around science), in doing so, educators may better prepare young children for learning in the later school years.

Chapter 2: Study 1

Simple Physics Diagrams of an Inclined Plane Support Preschoolers' Scientific

Predicting and Checking

2.1 Introduction.

The past twenty years have witnessed an increased interest in early science education, with educators and policymakers alike showing great concern for improving both mathematics and science learning (i.e., Brenneman, Stevenson-Boyd, & Frede, 2009). Science and mathematics often are not emphasized in the early childhood classroom (McClure et al., 2017; NRC, 2015; Ryan, Whitebook, & Cassidy, 2014). However, we now have evidence that early experiences in science are important for supporting not only gains in science, but also gains in other disciplines as well, such as literacy and language development, and mathematics learning (i.e., Brenneman, 2014; French, 2004; Greenfield et al., 2009). Thus, science is being identified as a key element in school readiness (Brenneman, 2011; Brenneman, Stevenson-Boyd, & Frede, 2009). Goals for early science include providing young children with opportunities to investigate the world around them, as well as confront an array of concepts in areas such as science, technology, engineering and mathematics. Further, some comprehensive early childhood curricula have adapted science specific components in the name of potential best practices (i.e., French, Conezio, & Boynton, 2000; Gelman, Brenneman, Macdonald, & Roman, 2009; Howitt, Upson, & Lewis, 2011).

Young children are like little scientists. They are eager and motivated to question and explore scientific concepts in their everyday world. Further, from a young age, they start to acquire an array of knowledge about the natural world. They are capable of thinking about and understanding concepts that relate to a number of scientific disciplines, such as biology, physics, psychology and chemistry (i.e., Au, 1994; Hatano, & Inagaki, 1994; NRC, 2007; Spelke, Katz, Purcell, Ehrlich, & Breinlinger, 1994;

Wellman & Gelman, 1998). Young children also have emerging reasoning and thinking

skills that they can apply to scientific content, and that might support the development of

more sophisticated scientific thinking in the later school years. For example, young

children ask questions about, and generate explanations for, phenomena that they observe

in the world around them (i.e., Callanan & Oakes, 1992). They also are capable of

engaging in basic scientific inquiry (Ashbrook, 2016; Klar & Chen, 2003), as well as

basic forms of logical thinking, such as transitivity (i.e., Bryant & Trabasso, 1971;

Nayfeld et al., 2011), and they can predict and check what will happen as a result of

relevant and irrelevant causal variables (i.e., Duschl, Schweingruber, & Shouse, 2007;

Gelman & Brenneman, 2004, 2012; Gopnik & Sobel, 2000; Schulz & Gopnik, 2004).

In spite of children's readiness and enthusiasm for engaging with scientific

concepts and practices, little science teaching occurs in early childhood classrooms

(Brenneman, 2014; NRC, 2015). Research suggests that teachers devote relatively little

time engaged in science-relevant learning activities, and they seldom spend time in the

science area during children's free-choice time (Bowman, 2006). Further, a lack of

empirical research focuses on the effectiveness of classroom practices in science, and

therefore, still relatively little is known about effective teaching of science in preschool

(i.e., Brenneman, 2014; Brenneman, Stevenson-Boyd, & Frede, 2009; Greenfield et al.,

2009). It therefore is no surprise that little is also known about the effectiveness of the

pedagogy and instructional tools that educators can use to support science learning in the

preschool classroom.

Although an increasing number of resources have been developed for teaching

science to young children, few of these are designed to support educators both

theoretically and practically (i.e., to provide educators with practical ideas for use in the classroom, as well as the knowledge of how to use them effectively). Finally, there remains work to identify elements of young children's thinking that can add to the foundation for developing scientific reasoning skills in the early years. It is important for researchers to pursue these relatively underexplored areas.

As described in further detail below, this study explores young children's capacity to predict and check in scientific investigations of objects traversing an inclined plane. It also considers the effectiveness of a pedagogical tool to support preschool children's scientific thinking. Finally, the study addresses whether it is reasonable to assume that there is a relationship between children's scientific thinking and their developing theory of mind and executive function skills.

2.2 Research Goals.

1. Basic scientific practices, such as making and checking predictions within scientific domains, have been deemed developmentally-appropriate for children at the preschool level (i.e., Ashbrook, 2016; Gelman, Brenneman, Macdonald, & Roman, 2009). Accordingly, one goal for this study was to explore young children's ability to predict and check in the context of scientific investigations in physical science, as well as the developmental trajectory of such skills over the preschool years. Preschool children were provided complex objects constructed from TinkerToy™ pieces (See Figure 2.3 below) and were asked to predict whether each would roll or slide down an inclined plane, then think about the outcome and judge the accuracy of their initial predictions.

The aim was to check our assumption that the task would be suitable for fostering scientific thinking in the early childhood classroom.

Investigating the motion of various objects on an inclined plane, or "the simple mechanics of solid bounded objects" (Michaels, Shouse, & Schweingruber, 2008, p. 38), exposes children to a number of concepts within the domain of physical science. Such concepts about motion are also routinely included in standards for early science education. For example, part of the Next Generation Science Standards kindergarten performance expectations for *Motion and Stability: Forces and Interactions* (NGSS Lead States 2013) encourage teachers: (a) to have children investigate the outcomes of pushes and pulls on the motion of an object; and (b) analyze whether their solutions changed the speed and/or the direction of an object as they expected. Further, when children investigate the motion of various objects down an inclined plane, they are introduced to physical science concepts associated with the motion and position of objects. This is one way to demonstrate the crosscutting concept, *Structure and Function* (NRC, 2012). This type of scientific investigation also meets the criteria for an age-appropriate science activity as defined by Worth (2010). This is because it involves: (a) concepts which are important to science; (b) phenomena drawn from the environment which are available for direct exploration; (c) concepts which can be explored in-depth and from multiple perspectives over time; and (d) phenomena and concepts which are interesting and engaging to young children and their teachers.

Exploring the motion of objects traversing an incline is a particularly interesting choice because it provides the opportunity to offer young children intuitions about motion and its laws, without any use of mathematics or advanced language. However, it still

allows for the introduction of new, relevant vocabulary. For example, such context exposes children to the physics principle that as the angle of the inclined plane is increased, the acceleration of the object also increases. Another physics principle involved relates to objects *sliding* down a frictionless ramp. In such cases, two forces act on the object: (a) its weight pulls straight down toward the center of the earth; and (b) the ramp exerts upward force perpendicular to the surface of the ramp (i.e., the "normal" force). The result is that mass cancels out of Newton's Second Law, meaning that any object, regardless of size or mass, will slide down the frictionless incline with the same acceleration, or at the same rate (and again, the rate is dependent upon the angle of the incline).

Finally, a number of physics principles are also involved in considering the motion of objects that *roll* down an inclined plane, such as the influence of mass and radius on the object's linear and rotational acceleration. For example, physics principles tell us that the rotational acceleration of an object does not depend on the object's mass, although it does depend on its radius. Further, the linear acceleration of the object is really what is of interest, and it depends not on the object's radius or mass, but rather on how the mass is distributed. This means that all hoops for example, regardless of mass or size, roll at the same rate down the incline. The same is true for empty cans (all roll down the incline at the same rate, regardless of mass or size), but an empty can will always roll faster than a hoop. Further, all solid spheres will roll down the incline with the same acceleration, but every solid sphere, regardless of mass or size, will roll faster than any solid cylinder.

Although these physics concepts and principles are discussed in greater depth than a preschool child would be expected to understand, providing young children with opportunities to explore such concepts in an engaging context may lay the foundation to support deeper learning of such concepts in the later school years.

2. It is also important for the field of Early Childhood Education to identify elements of young children's thinking that can buttress the foundation for developing scientific reasoning in the early years. Thus, another goal of the research was to examine the relationships among some of young children's emerging cognitive skills, namely their theory of mind and executive function capacities, and early forms of scientific thinking. In particular, the study explores whether executive function and theory of mind skills relate to children's capacity to accurately predict and check their predictions in this context.

Predicting and checking requires the ability to hold predictions and observed outcomes in mind, as well as judge the accuracy of one's initial predictions after engaging in scientific manipulations. As described in further detail below, I hypothesized that children with a more advanced theory of mind and executive functioning skills would perform better on the science reasoning activity than those with less advanced skills in these areas. This is because such early developing cognitive skills may relate to children's ability to recall initial predictions after testing items on the ramp, as well as determine whether their predictions were confirmed. Thus, children's theory of mind and executive function scores would likely correlate positively with their science scores.

Executive function skills are a set of domain-general cognitive capacities that include the ability to inhibit and override dominant responses (i.e., Korkman, 2000), to

shift attention flexibly among multiple pieces of information (i.e., Davidson, Amso, Anderson, & Diamond, 2006), and to retain and manipulate multiple pieces of information in working memory at the same time (Hughes & Graham, 2002). These skills develop markedly during infancy and the preschool years, and continue to develop during the school years (i.e., Carlson, 2005; DeLuca & Leventer, 2008). Further, executive function capacity plays an important role in academic success and is foundational to learning across numerous academic domains, including during the preschool years. For example, in a sample of Head Start preschoolers moving into kindergarten, Blair and Razza (2007) report that executive function measures, such as inhibition and cognitive flexibility, predict academic outcomes in kindergarten. Similarly, Vitiello (2009) found that executive function capacity predicts gains within the preschool year in vocabulary, mathematics, and listening comprehension in Head Start preschoolers.

Although numerous researchers suggest that executive functions play a key role in learning during the preschool years and relate to academic outcomes in literacy and mathematics, few have directly examined such relationships with children's learning of science. An exception is found in work by Nayfeld, Fuccillo, and Greenfield (2013), who explore the relationship between children's executive functioning skills and their scientific thinking. They report that executive functioning scores predict gains in science, as well as math and literacy outcomes, in a low-income sample of preschool children. The team also found that executive function scores significantly predicted gains in all domains, but the gains for science were significantly larger than they were for math and literacy. Similarly, Zaitchik, Iqbal, and Carey (2014) examined the relationship between

young children's executive function capacities and their early biological reasoning. They also find that executive function scores significantly predict children's biology scores, even when controlling for verbal IQ and age. Taken together, these findings provide encouraging support for my hypothesis that children's science learning may benefit, at least to some degree, from their executive function skills.

Thus, in the current study it was hypothesized that children with stronger executive function skills would perform better on the predicting and checking science activity. To first predict and then check the outcome requires one to recognize similarities or discrepancies between an original prediction and an observed outcome. One must keep predictions and observed outcomes in mind, as well as evaluate the two in order to judge the accuracy of one's initial predictions. In the science reasoning activity used in this study, children were first asked to predict whether different items would roll or slide down the inclined plane, and after testing items on the ramp, to determine whether their initial predictions were confirmed. To be successful, children must make an appropriate prediction, hold that prediction in mind during the testing with the items on the inclined plane, and correctly observe and hold the outcome in mind. Then they must compare the outcome to the initial prediction, and finally, compare the two to determine whether the prediction was confirmed. Executive function skills are likely involved in this process, as such skills relate to the ability to retain and manipulate multiples pieces of information in working memory at the same time, as well as to shift attention flexibly among multiple pieces of information.

Finally, executive function skills also involve the ability to inhibit and override dominant responses. The need to inhibit may arise during scientific predicting and

checking, especially when there is a discrepancy between one's prediction and the observed outcome. For example, when asked to recall a prediction that was not confirmed, a child must recognize that the outcome she just witnessed is correct while inhibiting her expectation. Thus, it is necessary to inhibit this response in order avoid stating the incorrect prediction. It follows that children with stronger executive function skills would be better able to inhibit such responses compared to children with weaker executive function skills. Therefore, it was hypothesized that there would be a positive relationship between children's executive function scores and their science reasoning scores.

It was also hypothesized that children's theory of mind scores would predict their science reasoning scores. Theory of mind involves the capacity to attribute mental states, such as beliefs, knowledge, desires and intents to oneself and others, and to understand that others have beliefs, knowledge, desires, etc. that differ from one's own (i.e., Premack & Woodruff, 1978; Wellman & Liu, 2004). Theory of mind also involves the understanding that one's mental states have causal power. In other words, having a theory of mind allows one to understand that mental states can be the source of, and thus be used to predict and explain the actions of others (Premack & Woodruff, 1978).

Some argue that such understandings serve as a critical resource underlying scientific thinking, because early cognitive capacities associated with theory of mind may lead to later development of higher-order thinking skills required for scientific thinking (e.g., Duschl, Schweingruber, & Shouse, 2007; Kuhn, 2000). For example, being able to distinguish between external reality and mental events is an important factor in the development of a theory of mind (e.g., Wimmer & Perner, 1983). Understanding this

distinction likely relates to one's capacity for scientific thinking as scientific processes often require the ability to distinguish between one's predictions (mental events) and the observed outcomes of experimental manipulations (external reality). Further, key to the scientific domain is understanding evidence as separate from and influencing theories, and this understanding requires the theory of mind development of an awareness of the source of one's knowledge.

Another critical milestone in the development of a theory of mind is the attainment of the ability to detect and understand false beliefs, that is, the realization that: (a) others can have diverging beliefs about the same world; and (b) that one's own beliefs, like the beliefs of others, may be false. This milestone in development involves multiple understandings, such as that of how knowledge is formed, that individual's beliefs stem from their knowledge, that mental states can differ from reality, and that other's actions can be predicted by their mental states (Wimmer & Perner, 1983). A further argument is that the capacity to think about beliefs as distinct from reality is a "milestone of foundational status in the development of scientific thinking" (Kuhn & Pearsall, 2000, p. 119), because it underlies the core scientific ability to use evidence to falsify hypotheses. The notion of false belief is a critical component of scientific thinking because the inquiry process involves viewing ideas other than one's own as plausible and being open to testing the alternatives (Duschl, Schweingruber, & Shouse, 2007).

Thus, I hypothesized that children's theory of mind scores would predict their science reasoning scores. Again, this is because the development of a theory of mind involves the ability to distinguish between external reality and mental events, and such

skills may relate to children's capacity to distinguish between their predictions (mental events) and the observed outcomes of experimental manipulations (external reality). Further, the theory of mind ability to understand that mental states can differ from reality may relate to children's capacity to accurately recall, and judge the correctness of, initial predictions, since recalling and judging predictions (especially incorrect predictions) involves the understanding that predictions may be incongruent with reality. Theory of mind also involves the ability to recognize false beliefs. This skill may be associated with children's ability to judge the accuracy of their own predictions, especially when those predictions are wrong, or not confirmed by testing items on the inclined plane. This is because recognizing that an initial prediction is incorrect involves the realization that a prior belief was false. Therefore, it is likely that children with stronger theory of mind skills will also perform better in recalling and judging the accuracy of initial predictions, especially when those predictions are wrong, or false.

3. Finally, there is little work about the pedagogical methods and tools that educators can use to support effective science learning in the early childhood classroom (Brenneman, Stevenson-Boyd, & Frede, 2009). As described in further detail below, an overarching goal for this study was to explore the effectiveness of pedagogical tools for use in the preschool classroom, ones that can support young children's external representation of, and reflection upon, key components of their scientific investigations. There are representational tools that are suitable for communicating and recording scientific methods and ideas. These include pictures, diagrams, figures, etc. to convey complex ideas, trends, and explanations of phenomena in more accessible and condensed formats. Further, goals for science education include providing students with

opportunities to represent scientific ideas in various formats (as well as support in doing so), and opportunities for reflection, as reflection helps students to monitor their understanding and track the progress of their scientific investigations (Brown, 1997; Duschl, Schweingruber, & Shouse, 2007; Larkin & Simon, 1987; NSTA, 2014).

Prior research supports the effectiveness of affording students with opportunities to reflect during science learning activities. For example, Kloos and Somerville (2011) argue that providing children with opportunities to reflect on science learning experiences reinforces their abilities to revise initially mistaken beliefs. Likewise, Gropen, Clark-Ciarelli, Hoisington, and Ehrlich (2011) suggest that providing opportunities to reflect during scientific investigations helps children to revise originally mistaken scientific beliefs, because reflection helps children to differentiate between what happened in a scientific manipulation and what they originally thought would occur.

Gelman and her colleagues have successfully introduced such ideas to children in their preschool science program, *Preschool Pathways to Science* (PrePS), with use of science journals (Gelman et al., 2009; Brenneman & Lauro, 2008). In this program, children are supported in recording and documenting their scientific investigations in science journals, such as by drawing and using numerals to represent them, date stamping, and having adults write what the children want to document. This is similar to the Emilio Reggio approach, which encourages children to use a variety of media to document and reflect upon their investigations (i.e., Katz, 1998). Finally, research suggests that classroom practices that explicitly support children's reflection on their own knowledge, such as what they already know, what they need to find out, and how they might find out, foster children's development as effective learners and problem solvers

(NRC, 2001). However, questions remain about how to best support young children's representation of, and reflection upon scientific investigations.

In the current study, simple diagrams of an inclined plane were developed to support children in externally representing, and reflecting upon, their predictions and observations as they engaged in the science reasoning activity. Specifically, the diagrams were designed to represent physical science concepts concerning the mechanisms for rolling and sliding down the inclined plane, and they may serve as design tools for use by teachers in the early childhood classroom, both to keep track of children's knowledge and to offer support of their thinking about science activities. (See inclined plane diagrams in Figures 2.1 and 2.2, below).
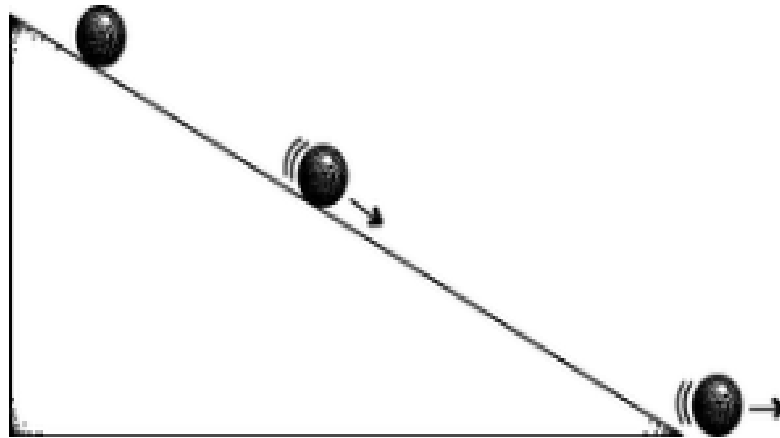
Figure 2.1. *Physics diagram of an inclined plane, representing the mechanism of rolling.*
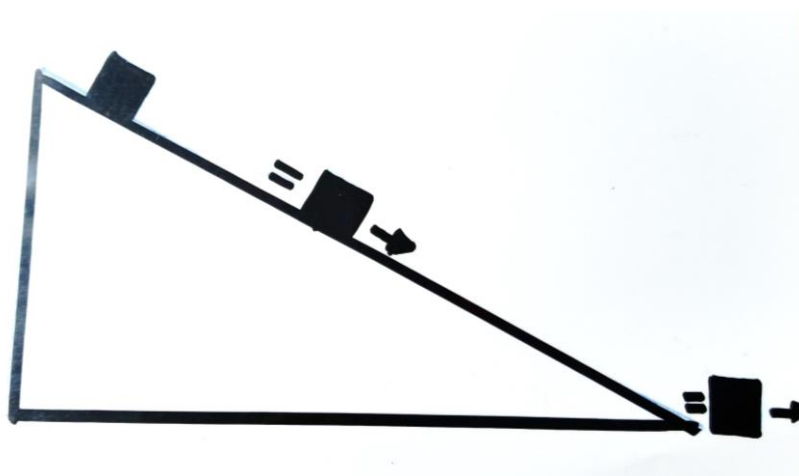
Figure 2.2. *Physics diagram of an inclined plane, representing the mechanism of sliding.*

Children used the diagrams to externally represent and record their predictions (i.e., whether items would roll or slide down the ramp), as well as the observed outcomes of testing items on the ramp. Further, when asked to recall, as well as judge the accuracy of, initial predictions, children used the diagrams to reflect upon their scientific investigations. The goal was to determine whether children are capable of understanding the diagrammatic representation characterized by the stimuli, as well as whether use of the diagrams supports children's predicting and checking abilities in this context.

I hypothesized that use of the diagrams would extend and support children's developing cognitive capacities (i.e., executive function and theory of mind) by helping children to keep their predictions in mind, as well as compare the initial predictions to the observed outcomes in order to judge the accuracy of those predictions. The rationale is that the charts may aid children in attending to the relevant components of their scientific investigations, without overloading working memory capacity. Further, encouraging children's reflection upon such components of the scientific investigations may support children in recognizing when their predictions are incorrect (especially among those

developing false belief understanding). Thus, use of the inclined plane diagrams will likely support children's predicting and checking abilities by supporting their developing executive function and theory of mind capacities.

2.3 Research Questions.

To summarize, the research questions guiding this work include:

1. *How accurately do preschool children make and check predictions about objects' motion on an inclined plane? Further, what is the developmental trajectory of such skills over the preschool years?*
2. *Do preschool children's theory of mind and executive functioning capacities relate to their abilities to make and check predictions in this context?*
3. *Do simple diagrams of an inclined plane support preschool children's ability to recall, as well as judge the accuracy of, their initial predictions?*

2.4 Experimental Details.

2.4a Participants. The children for the study were drawn from two Central New Jersey preschools. The first school is a private preschool situated on the campus of a prestigious private university, and it serves both university-affiliated families as well as community families unassociated with the university. The school has six classrooms grouped by age, for children ranging in age from 2 ½ to 5 years. The school is licensed by the State of New Jersey as a Child Care Center for children over the of age 2 ½ years, and it is accredited by the National Association for the Education of Young Children (NAEYC). The majority of children were Caucasian and from families of middle-class socio-economic backgrounds.

The second school is a preschool program located within a public, Central New Jersey high school. The program is part of the high school's Child Development program. High school students have the option to enroll in the course as an elective, and the course involves the high school students running the preschool program. For example, under the direction of the classroom instructor, high school students assist in the design and implementation of instruction. The program comprises half day sessions which are open to 3- and 4-year-old children in the community. Preschool children attend free of charge, on a first-come, first-served basis. The school does not follow a specific curriculum; however, each week includes learning experiences within the areas of literacy, mathematics, science, social studies, and the arts. Children attending the school were predominantly Caucasian and from families of middle-class socio-economic backgrounds.

Once the work received IRB approval, parents of all 3- and 4-year-old children at both schools were provided information about the study and offered consent forms. Parents returned signed consent forms for sixty-one children. One child did not want to participate, resulting in a final sample of 60 children (40 girls, 20 boys). The mean age was 49 months, with a range of 34 to 61 months. All children spoke English (although for a few, as a second language).

2.4b Materials. Items for the experiment consisted of a wooden inclined plane, six objects constructed from TinkerToy™ pieces, and a felt board with physics diagrams of an inclined plane to represent and record children's predictions and observations during the science activity. See Figures 2.3 through 2.5 below.
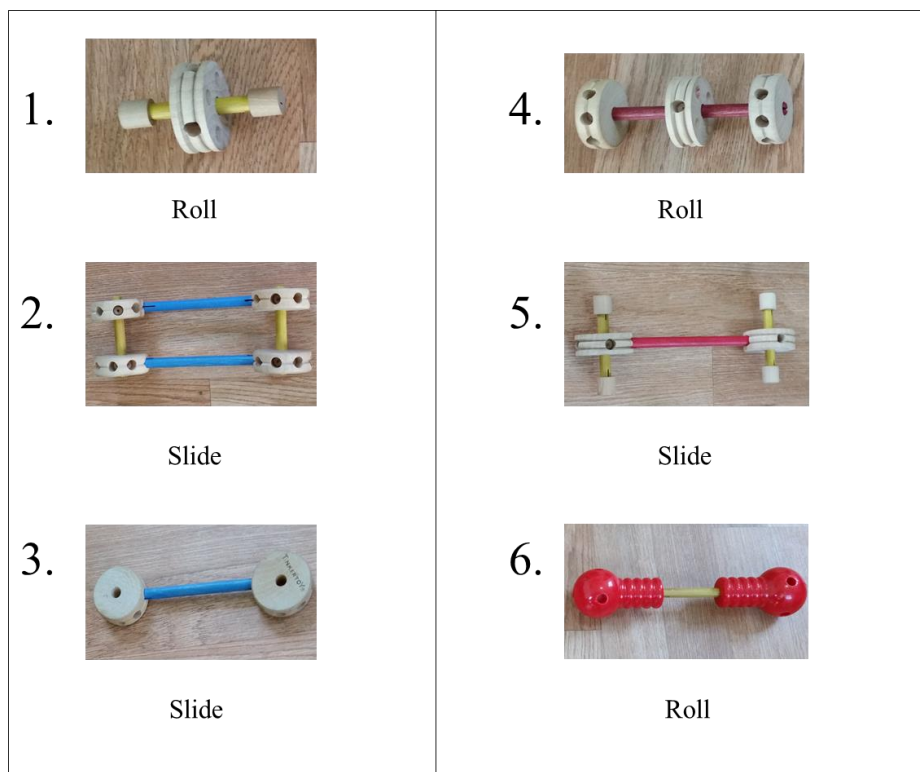
Figure 2.3. *Test items, and their mechanisms of motion down the inclined plane.*

Figure 2.4. *Inclined Plane.*
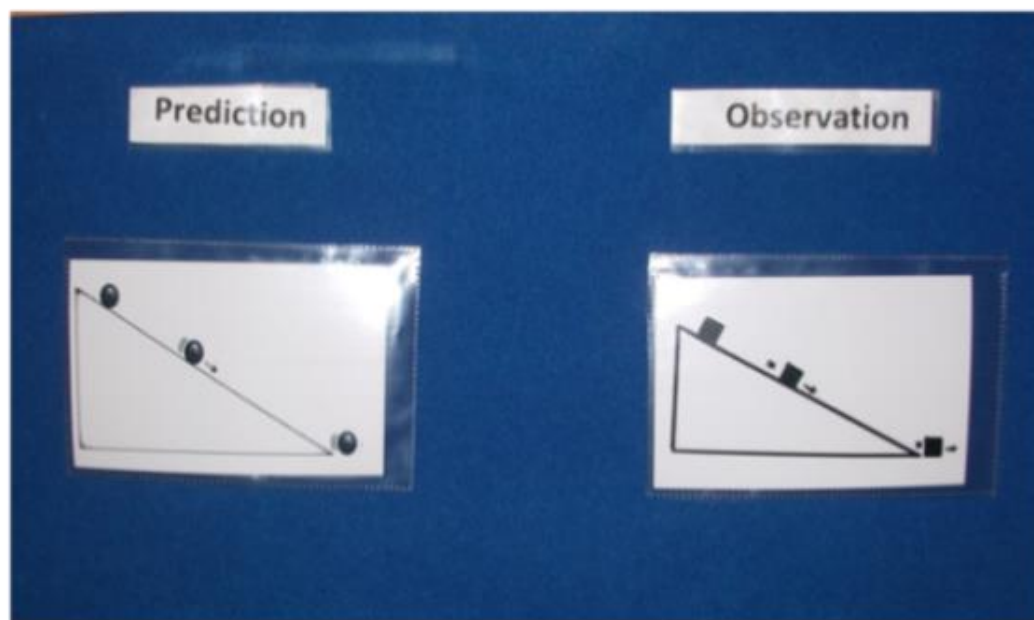
Figure 2.5. *Felt board and inclined plane diagrams for representing and recording predictions and observations. (Note: In the above example, the child incorrectly predicted that an object would roll down the ramp, and then observed that the object actually slid down).*

2.4c Methods. The experimenter first spent a few days in the classroom for children to become familiarized and comfortable with her. Next, two pre-tests were administered to obtain initial measures of children's executive function and theory of mind capacities: (1) a standard Dimensional Change Card Sort task (DCCS), and (2) a standard theory of mind task. The DCCS (Frye, Zelazo, & Palfai, 1995) is a commonly used, developmentally-sensitive measure of executive function for young children. This rule-based, card sorting task involves a combination of working memory and inhibition. Children are first shown two boxes with target cards affixed to the boxes (e.g., a blue boat and a red rabbit). Children are presented a series of cards (e.g., red and blue rabbits and boats) and are first asked to sort by shape, and then by color (in counterbalanced order). The final phase asks children to switch rules on each trial. Previous work shows that three-year-olds rarely pass the task, whereas four-year-olds are just above chance

levels, and five-year-olds are near ceiling (i.e., Carlson, 2005). See Appendix A for full scripts and procedure for the Dimensional Change Card sort pre-test.

A standard, well-established theory of mind task was also administered (i.e., Gopnik & Astington, 1988). Children were presented with a deceptive item (i.e., a crayon box containing Band-Aids™). The item was first presented in its deceptive state, and it was assumed that children would represent the item incorrectly. The experimenter then revealed the item's true state by opening the box and showing the child what was inside. The item was then returned to its deceptive state and taken out of the child's reach. The false belief question asked children what a naïve observer (e.g., a classmate who had not yet had a turn to play the game) would think the item contained. The representational change question asked children what they thought the box contained before its true state was revealed. Finally, the appearance-reality task involved a reality question ("*What's really and truly in the box, Band-Aids™ or crayons?*"), and an appearance question ("*What does this look like it has in it, Band-Aids™ or crayons?*"). (See Appendix B for full scripts and procedure for the Theory of Mind pre-test).

Next (and on a different day), children engaged in a science reasoning activity. This involved interviewing children individually in either a quiet area of their classroom or an empty room adjacent to their classroom, and asking them to reason about objects' motion down an inclined plane. Children were shown each item (see Figure 2.3 above), one at a time, and were asked to predict whether each would roll or slide down the incline. After making a prediction, children were asked to test the items on the ramp, watch what happened, and then revisit their original prediction to judge its accuracy.

Controlling for age, children were randomly assigned to either the diagram or control condition. During the science activity, children in the diagram condition used physics diagrams of an inclined plane to externally represent and reflect upon their predictions and observations. Children used the diagrams to record each prediction on a felt board, as well as the outcome of testing each item on the ramp. (See Figure 2.5 above). They were able to use the diagrams to support their recall of initial predictions, as well as their judgment regarding the correctness of those predictions. In contrast, children in the control condition did not have access to the physics diagrams. Instead, they made their predictions, observations, and judgments verbally. See questions and prompts used for the ramp activity in Table 2.1 below. (Also see Appendix C for complete scripts and procedures for the ramp activity).

Table 2.1. *Prompts/Questions used for the Science Activity, by Condition.*

Diagram Condition

Make prediction prompt- "First, predict, or guess. (Holding item at top of ramp) If I let go, will it roll down, or slide down (counterbalanced order)? Okay. Let's put your prediction on the chart." (Experimenter helps child place prediction diagram card on the felt board).

Test and observe outcome prompt- "Now it's time to try it on the ramp. Did it roll down, or slide (counterbalanced)? Let's put it on the chart." (Experimenter helps child place observation diagram card on the felt board).

Prediction recall prompt- "Before we tried it on the ramp, what was your guess, or prediction? Did you guess roll down, or slide down (counterbalanced)? You can use the chart to help you remember."

Judgment prompt- "Was your prediction right, or do you need to change it (counterbalanced)? You can use the chart to help you remember."

Control Condition

<u>Make prediction prompt</u>- "First, predict, or guess. (Holding item at top of ramp) If I let
	go, will it roll down, or slide down (counterbalanced)? Okay."

<u>Test and observe outcome prompt</u>- "Now it's time to try it on the ramp. Did it roll down,
	or slide (counterbalanced)?"

<u>Prediction recall prompt</u>- "Before we tried it on the ramp, what was your guess, or
	prediction? Did you think it would roll down, or slide down (counterbalanced)?"

<u>Judgment prompt</u>- "Was your prediction right, or do you need to change it?

2.5 Analyses.

Scores on the Dimensional Change Card Sort (DCCS) pre-test were assigned

following the procedures described in Frye, Zelazo, and Palfai (1995). Children were

first taught to sort cards according to one dimension (e.g., color) until they correctly

sorted at least four out of five cards according to that dimension. Next, children were

taught to sort according to the other dimension (e.g., shape). Finally, children were asked

to sort four cards, switching rules on each trial. Children received one point for each card

correctly sorted, resulting in a total score between 0 and 4 points. A passing score was

considered scoring a 4 out of 4 possible points.

Scoring for the theory of mind pre-test followed the procedures described by

Gopnik & Astington (1988). Children were scored as having passed the representational

change question if they correctly reported their initial representation of the object (i.e., "I

thought it had crayons"). Children passed the false-belief question if they said that a

classmate who had not seen the box would think it had crayons in it. To pass the

appearance-reality question, children had to answer both parts correctly; they had to

answer that: (1). the box looks like it has crayons in it, but (2). it actually contains Band-

Aids™. Thus, each child received a score between 0 and 3 points. A passing score was considered scoring 3 out of a possible 3 points.

The science activity involved reasoning about whether each of six objects constructed from TinkerToy™ pieces rolls or slides down the ramp. Further, children reasoned about each item two times consecutively using the same procedure (i.e., resulting in 12 total trials). Children were scored on the accuracy of their predictions, as well as their ability to recall, and judge the correctness of, their predictions after testing items on the inclined plane. Thus, each child received three separate scores: (1). a prediction accuracy score, based on the total number of correct predictions made over the course of the session; (2). a prediction recall score, based on the total number of initial predictions accurately recalled after testing items on the ramp; and (3). a judgment score, based on the total number of predictions correctly judged as right (i.e., those which were confirmed by testing items on the ramp), or wrong (i.e., those which were disconfirmed by testing items on the ramp). Raw scores were then converted to proportion correct.

Analyses examined children's overall mean scores, as well as group differences in children's performance based on age, condition and gender. Patterns and consistencies in children's responses were also identified. Finally, analyses also explored the relationships among children's scientific thinking skills and their theory of mind and executive functioning capacities, specifically, by examining correlations among children's passing status on the pre-tests and their ability to recall predictions after testing items on the ramp, as well as to judge the accuracy of those predictions. Analyses also examined the effect of the inclined plane diagrams on children's prediction recall and judgment scores, based on children's passing status on the pre-tests. The goal was to

determine whether use of the diagrams showed similar effects in children of differing

levels of development concerning executive function and theory of mind skills.

2.6 Results

2.6a Accuracy of Children's Predictions. Analyses examined children's ability to

accurately predict whether various items roll or slide down the inclined plane. Each child

engaged in 12 trials of predicting and checking. Accordingly, children were assigned a

prediction accuracy score out of a possible 12 points. See distributions of prediction

accuracy scores in Figures 2.6 through 2.8 below.
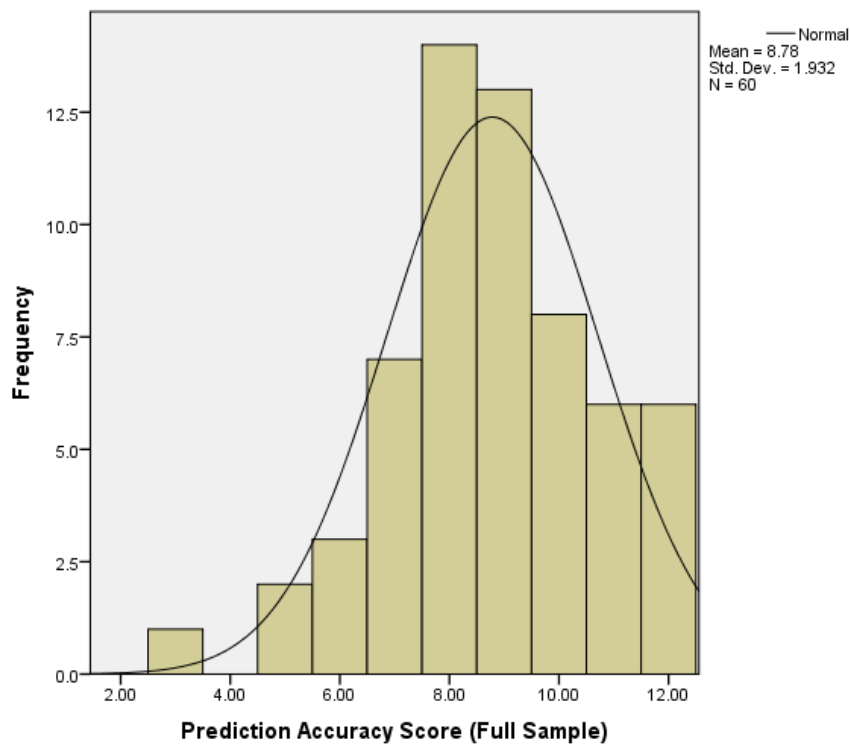
Figure 2.6. *Distribution of prediction accuracy scores for the full sample (n = 60).*
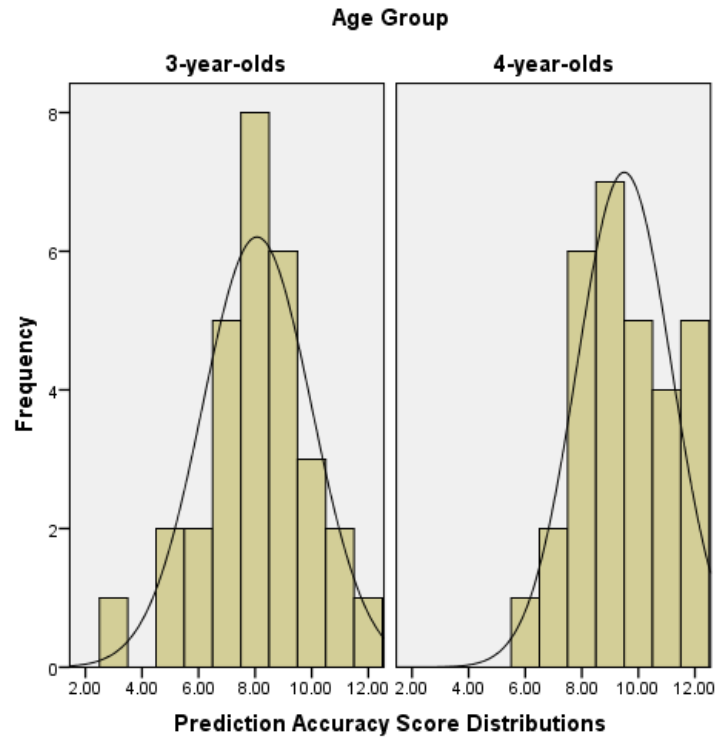
Figure 2.7. *Distribution of prediction accuracy scores, by age group.*
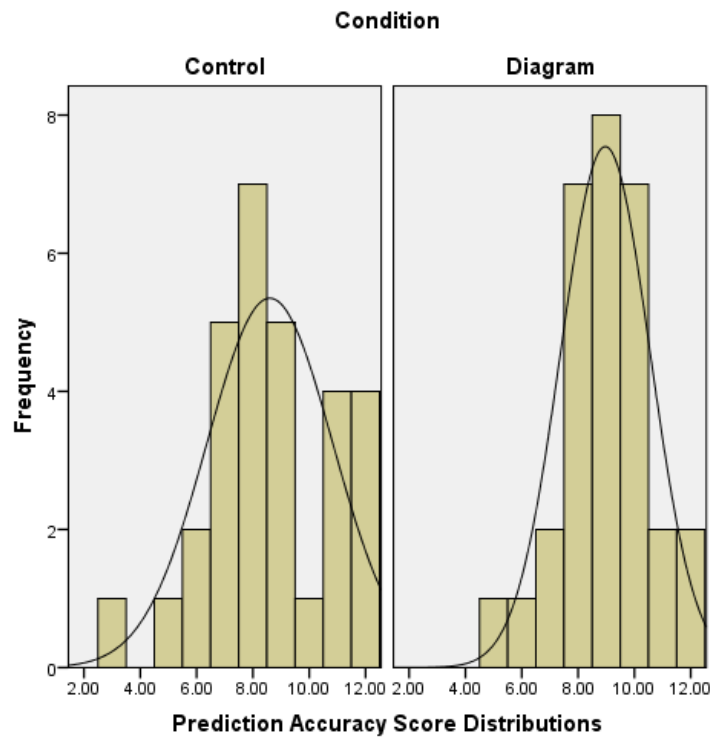
Figure 2.8. *Distribution of prediction accuracy scores, by condition.*

Scores were then converted to proportion correct. Overall, the mean prediction accuracy score (mean proportion correct) was 0.73 ($SD = 0.16$). One sample t-tests (2-tailed) against chance found that children's predictions were above chance levels for both age groups. The mean for the three-year-olds was 0.67 ($SD = 0.16$), whereas the mean for the four-year-olds was 0.79 ($SD = 0.14$); $t(29) = 5.87$, $p < .001$, and $t(29) = 11.44$, $p < .001$, respectively.

Analyses also examined the effects of condition, age and gender on children's prediction accuracy scores. A Mann-Whitney U test (2-tailed)[1] indicated that prediction accuracy scores were higher for the 4-year-olds ($Mdn = 0.75$, mean rank $= 36.73$) compared to the 3-year-olds ($Mdn = 0.67$, mean rank $= 24.27$), $U = 263.00$, $p = .005$. No condition nor gender effects were found, $U = 392.50$, $p = .39$ (2-tailed), and $U = 389.50$, $p = .87$ (2-tailed), respectively. (Note: A significant effect for condition was not expected, as the inclined plane diagrams are not intended to support children in making correct predictions, but rather to aid children in keeping predictions and outcomes in mind, as well as comparing them). See Table 2.2 below for descriptive statistics. Also see Figure 2.9 below.

Table 2.2. *Descriptive Statistics for Prediction Accuracy Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample ($n = 60$)	0.73 (0.16)
3 YOs ($n = 30$)	0.67 (0.16)
4 YOs ($n = 30$)	0.79 (0.14)
Diagram Cond. ($n = 30$)	0.74 (0.13)
Control Cond. ($n = 30$)	0.72 (0.19)
Boys ($n = 20$)	0.74 (0.15)
Girls ($n = 40$)	0.73 (0.17)
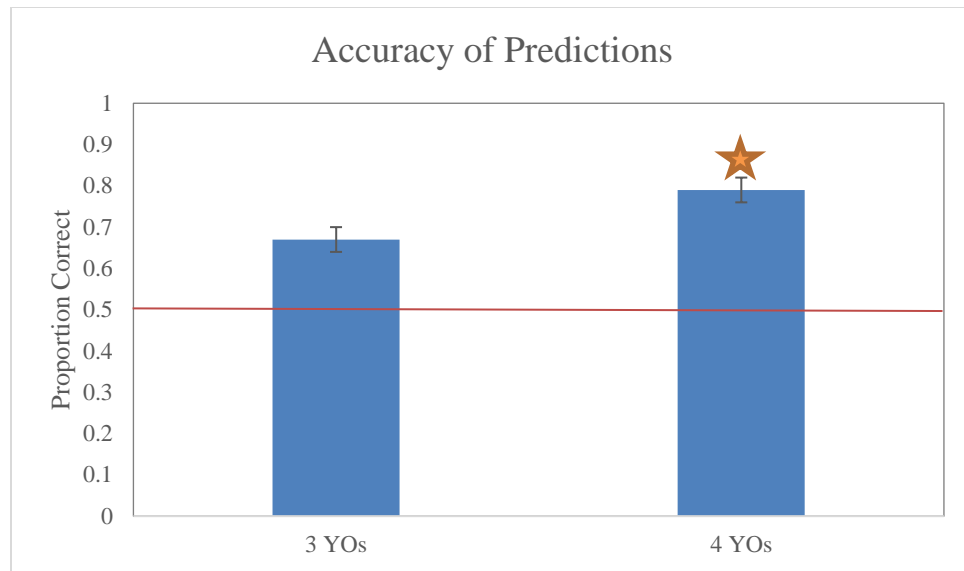
Figure 2.9. *Accuracy of children's predictions (mean proportion correct, by age).* Note: Error bars on all graphs represent one standard error of the mean[2].

In summary, from a young age, children were above chance levels, but not at ceiling, in terms of accurately predicting about objects' motion down an inclined plane. Further, this skill increased with age, as the older children reliably outperformed the younger age group.

2.6b Individual Item Analysis. Analyses also examined prediction scores on each trial individually to assess whether children's predictions were correct above chance levels for each test item, as well as whether there were any item effects (i.e., if some items were more difficult for children than others). As children predicted and checked two times sequentially for each test item, the prediction scores were averaged for each item.

Within each age group, one sample t-tests (2-tailed) against chance were conducted for each item. The three-year-olds performed better than chance (0.50) on item 1 ($t(29)$= 3.50, $p = .002$, $d = 1.30$), item 2 ($t(29)$= 2.41, $p = .023$, $d = .89$), item 3

($t(29)$= 4.07, $p < .001$, $d = 1.51$), item 4 ($t(29)$= 2.41, $p = .023$, $d = .89$), and item 6

($t(29)$= 2.80, $p = .009$, $d = 1.04$). For each of these analyses, the effect size was found to

exceed Cohen's (1988) convention for a large effect ($d = .80$). It was only for item 5 that

the three-year-olds were at chance levels, ($t(29)$= 0.94, $p = .35$). Cohen's (1988) effect

size value suggested low practical significance ($d = .35$).

The four-year-olds performed better than chance on all six test items. For item 1

($t(29)$= 7.17, $p < .001$, $d = 2.66$), item 2 ($t(29)$= 3.53, $p = .001$, $d = 1.31$), item 3 ($t(29)$=

5.29, $p < .001$, $d = 1.96$), item 4 ($t(29)$= 4.96, $p < .001$, $d = 1.84$), item 5 ($t(29)$= 2.57, p

$= .016$, $d = .95$), and item 6 ($t(29)$= 16.16, $p < .001$, $d = 6.00$). Again, the effect size for

each of these analyses was found to exceed Cohen's convention for a large effect ($d =$

$.80$).

Thus, overall, children were capable of accurately predicting whether the various

test items roll or slide down the inclined plane. Four-year-olds met statistical

significance on all 6 items, and three- year-olds were above chance on all but item 5. See

Table 2.3 below for overall means and standard deviations for each item, by age group.

Table 2.3. *Mean and SD for Children's Performance on each Item (Trial 1 and 2), by Age.*

Item	1a	1b	2a	2b	3a	3b	4a	4b	5a	5b	6a	6b
3's	0.77(.43)	0.67(.48)	0.60(.50)	0.73(.45)	0.70(.47)	0.77(.43)	0.60(.50)	0.73(.45)	0.47(.51)	0.67(.48)	0.70(.47)	0.67(.48)
4's	0.83(.38)	0.87(.35)	0.70(.47)	0.70(.47)	0.90(.31)	0.70(.47)	0.73(.45)	0.83(.38)	0.57(.50)	0.77(.43)	0.93(.25)	0.97(.18)
Tot.	0.80(.40)	0.77(.43)	0.65(.48)	0.72(.45)	0.80(.40)	0.73(.45)	0.67(.48)	0.78(.42)	0.52(.50)	0.72(.45)	0.82(.39)	0.82(.39)

2.6c Recalling Initial Predictions. Analyses also explored children's ability to

recall their initial predictions after testing items on the ramp to determine whether they

roll or slide down. Each child was assigned a prediction recall score (out of a possible 12 points), based on the total number of predictions correctly recalled, regardless of whether those predictions were confirmed or disconfirmed by testing the items on the ramp. See distributions of prediction recall scores in Figures 2.10 through 2.12 below.
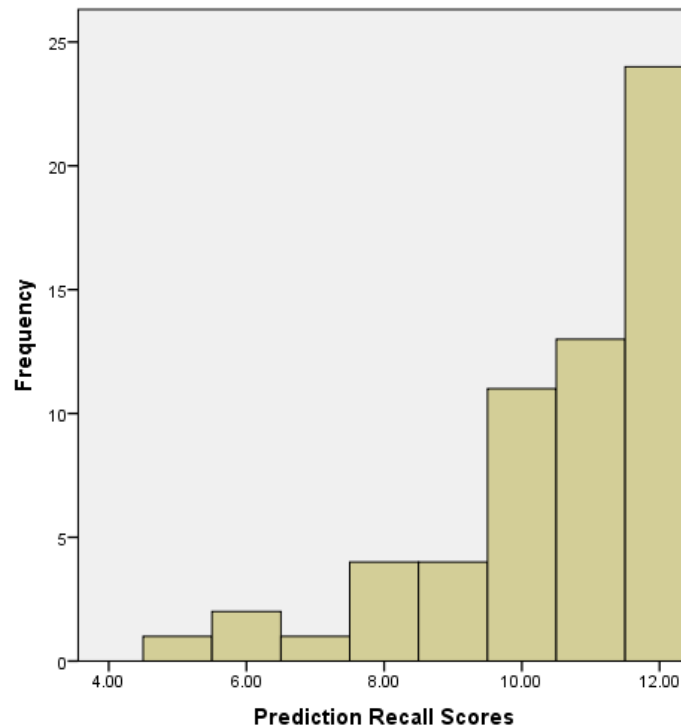
Figure 2.10. *Distribution of prediction recall scores for the full sample (n = 60).*
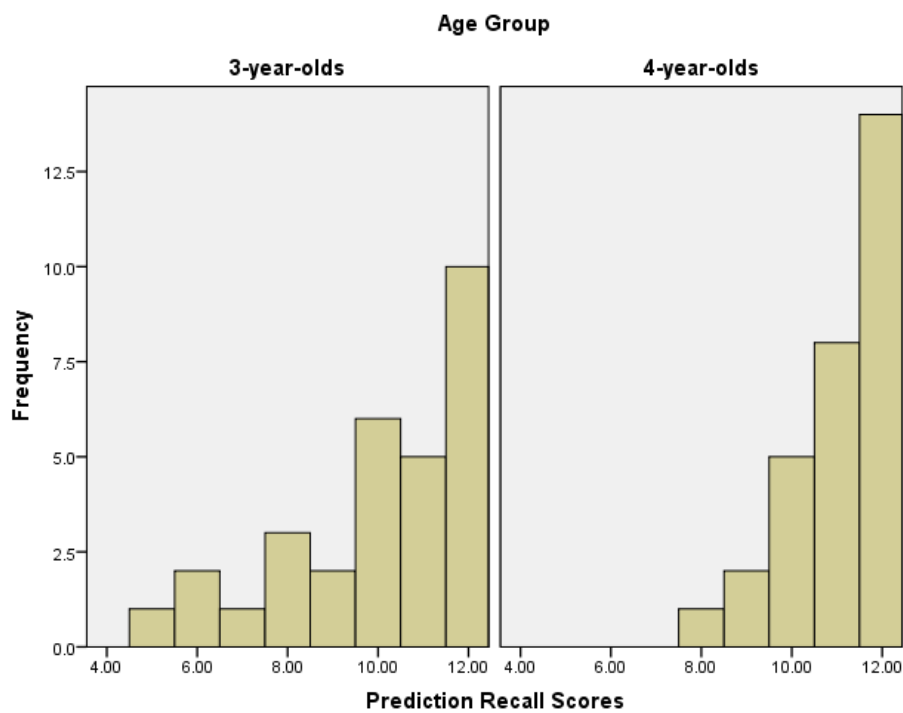
Figure 2.11. *Distribution of prediction recall scores, by age group.*
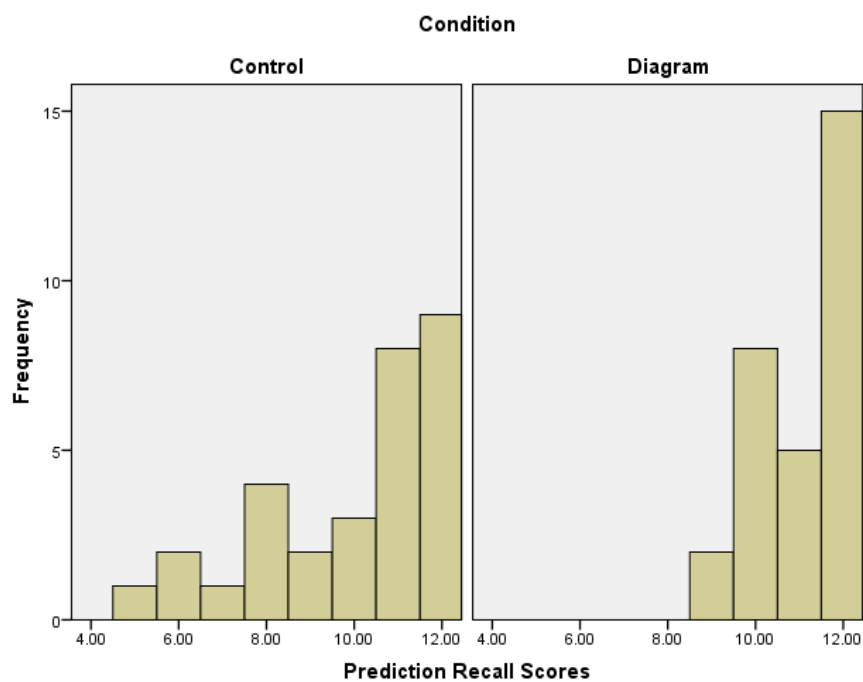
Figure 2.12. *Distribution of prediction recall scores, by condition.*

Raw scores were then converted to proportion correct. Overall, the mean prediction recall score (mean proportion correct) was 0.88 ($SD= 0.14$). One-sample t-tests (2-tailed) against chance found that, for both age groups, children's prediction recall scores were reliably above chance levels. The mean proportion correct for the three-year-olds was 0.84 ($SD = 0.17$), $t(29) = 10.64$, $p < .001$. For the four-year-olds, the mean was 0.92 ($SD = 0.09$), $t(29) = 24.95$, $p < .001$. Further, the prediction recall scores for 48 children (80%) were above chance levels according to the strict binomial criterion of correctly recalling at least 10 out of 12 predictions over the course of the activity (binomial sign test, $p = .02$). By age, 70% of three-year-olds, and 90% of four-year-olds performed better than chance according to this criterion.

Analyses also examined the effects of condition, age and gender on children's prediction recall scores. A Mann-Whitney U test (2-tailed) indicated that prediction recall scores were significantly higher for the diagram condition ($Mdn = 0.96$, mean rank = 34.75) compared to the control condition ($Mdn = 0.92$, mean rank = 26.25), $U = 322.50$, $p = .05$. Further, a Mann-Whitney U test (2-tailed) exploring the effect of age on prediction recall scores reached marginal significance. The prediction recall scores for the 4-year-olds ($Mdn = 0.92$, mean rank= 34.55) were higher than those of the 3-year-olds ($Mdn = 0.88$, mean rank = 26.45), $U = 328.50$, $p = .06$. Finally, no gender difference was found, $U = 294.50$, $p = .08$. See descriptive statistics in Table 2.4 below. Also see age and condition effects in Figure 2.13 below.

Table 2.4. *Descriptive Statistics for Prediction Recall Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample (*n* = 60)	0.88 (0.14)
3 YOs (*n* = 30)	0.84 (0.17)
4 YOs (*n* = 30)	0.92 (0.09)
Diagram Cond. (*n* = 30)	0.93 (0.09)
Control Cond. (*n* = 30)	0.83 (0.18)
Boys (*n* = 20)	0.85 (0.14)
Girls (*n* = 40)	0.90 (0.14)
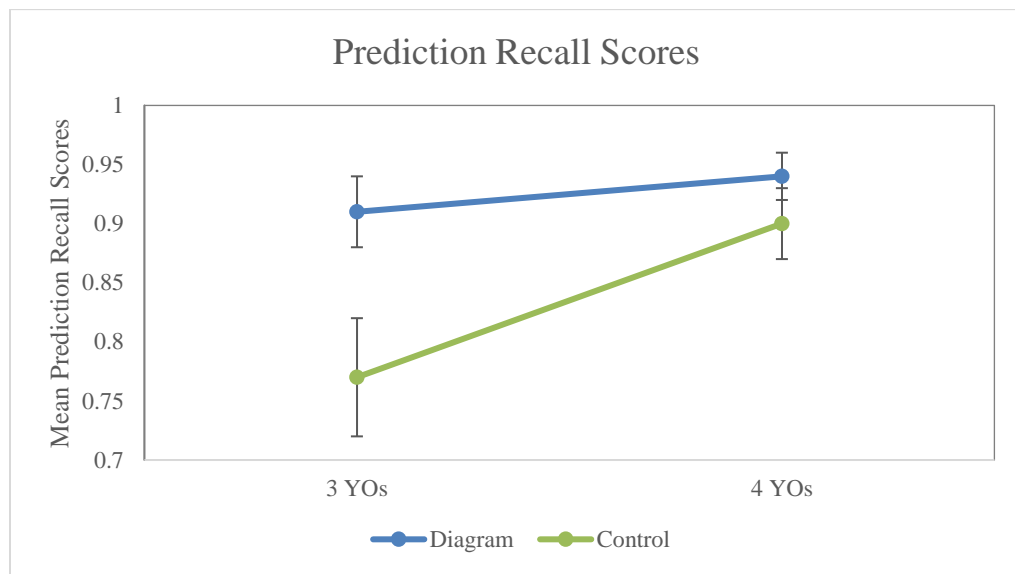
Figure 2.13. *Condition and age effects on children's mean prediction recall scores.*

In summary, from a young age, children were skilled in accurately recalling their predictions after testing items on the inclined plane. This skill also increased over the preschool years. Importantly, use of the inclined plane diagrams effectively supported children's ability to accurately recall their initial predictions after testing items on the inclined plane. The effect was especially salient for the younger group.

2.6d Accuracy of Children's Judgments. Analyses also more closely examined children's ability to judge whether or not their predictions were confirmed by testing the items on the ramp. As the science activity involved 12 trials of predicting and checking, each child received 1 point for each correct answer concerning the accuracy of their predictions. In other words, 1 point was awarded each time the child agreed that his/her correct prediction was in fact right, as well as each time the child agreed that his/her incorrect prediction was indeed wrong. Thus, children could receive a judgment score between 0 and 12 points. See distributions of judgment scores in Figures 2.14 through 2.16 below.
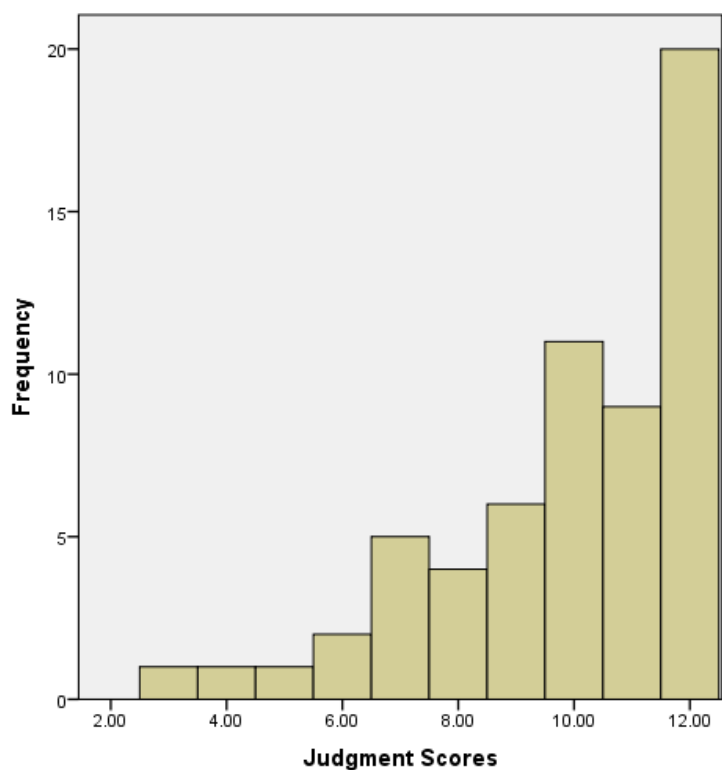
Figure 2.14. *Distribution of judgment scores for the full sample (n = 60).*
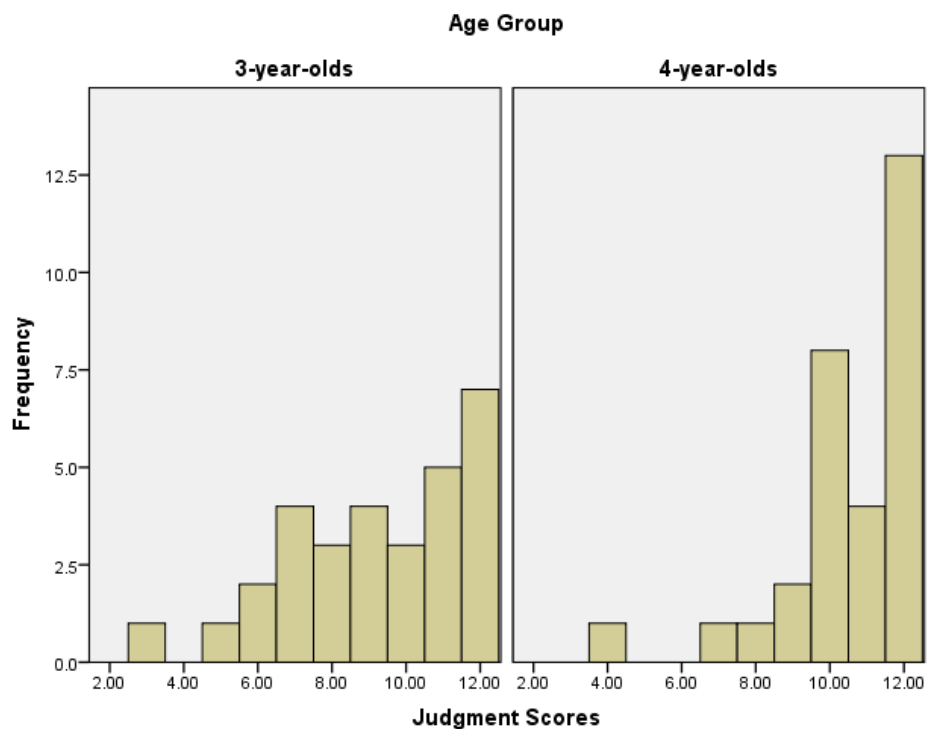
Figure 2.15. *Distribution of judgment scores, by age group.*
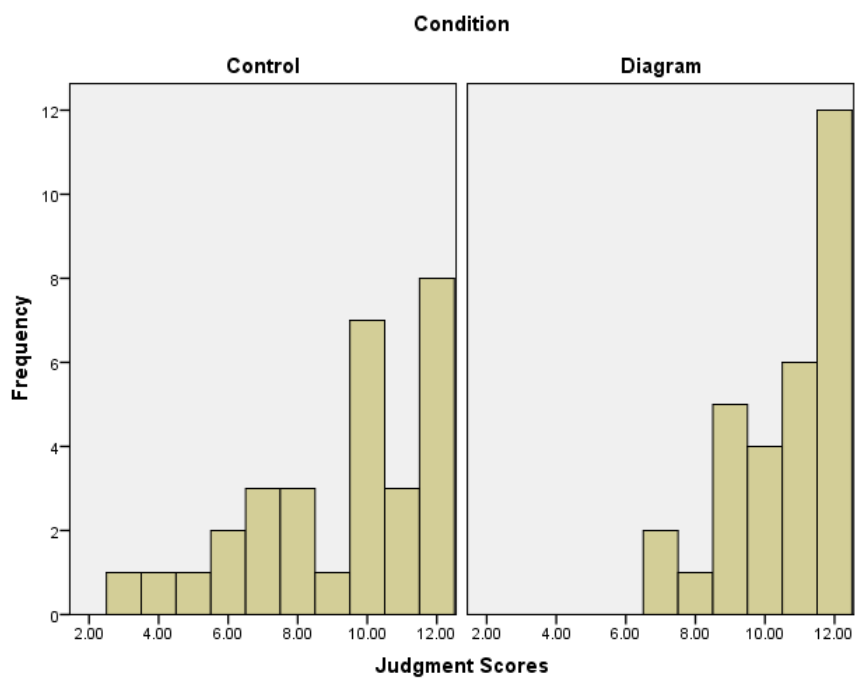
Figure 2.16. *Distribution of judgment scores, by condition.*

Scores were then converted to proportion correct. Overall, the mean judgment score (mean proportion correct) was 0. 83 (SD = 0.19). One-sample t-tests (2-tailed) against chance indicated that, for both age groups, children's judgment scores were reliably above chance levels. The mean was 0.77 (SD = 0.20) for the three-year-olds, and 0.88 (SD = 0.15) for the four-year-olds, ($t(29)$ = 7.24, p < .001; and $t(29)$ = 13.65, p < .001, respectively). Further, 40 children (67%) had passing judgment scores according to the strict binomial criterion of making 10 or more correct judgments out of 12 (binomial sign test, p = .02). By age, 50% of the three-year-olds, and 83% of the four-year-olds performed better than chance according to this criterion.

Analyses also examined the effects of condition, age and gender on children's judgment scores. A Mann-Whitney U test (2-tailed) indicated that judgment scores were reliably higher for the diagram condition (Mdn = 0.92, mean rank = 34.73) compared to the control condition (Mdn = 0.83, mean rank = 26.27), U = 323.00, p = .05. There was also a significant age effect. Scores for the four-year-olds (Mdn = 0.92, mean rank = 35.53) were significantly higher than those of the three-year-olds (Mdn = 0.79, mean rank = 25.47), U = 299.00, p = .022. Finally, no gender difference was found, U = 334.50, p = .29. See descriptive statistics in Table 2.5 below. Also see age and condition effects in Figure 2.17 below.

Table 2.5. *Descriptive Statistics for Judgment Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample (n = 60)	0.83 (0.19)
3 YOs (n = 30)	0.77 (0.20)
4 YOs (n = 30)	0.88 (0.15)
Diagram Cond. (n = 30)	0.88 (0.13)
Control Cond. (n = 30)	0.77 (0.22)
Boys (n = 20)	0.78 (0.22)
Girls (n = 40)	0.85 (0.17)
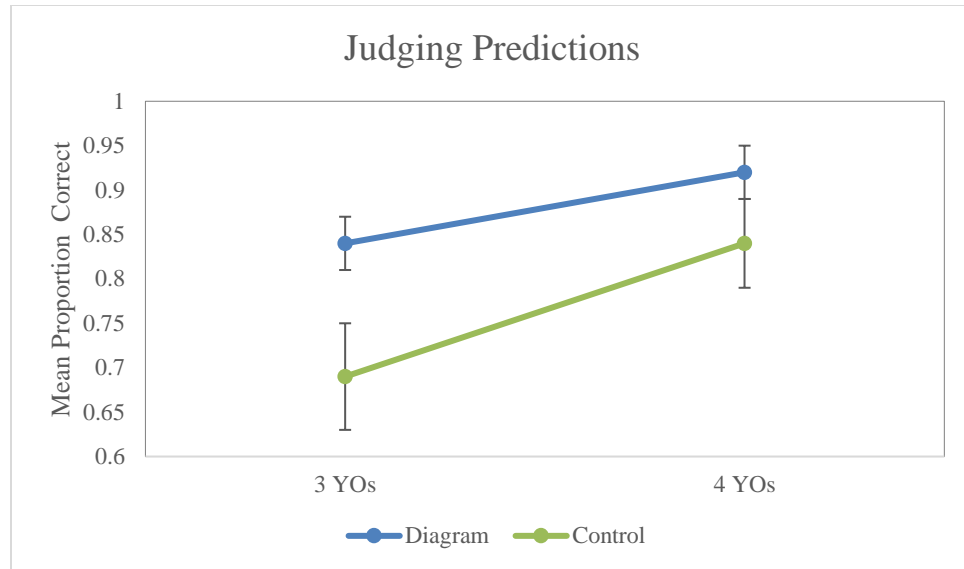
Figure 2.17. *Condition and age effects on children's mean judgment scores.*

In summary, preschool children in this study were able to judge the accuracy of their predictions about various objects' motion on the inclined plane, with this skill increasing over the preschool years. Further, use of the inclined plane diagrams reliably supported children's ability to judge the correctness of initial predictions after testing items on the ramp. The effect was again particularly salient among the three-year-olds.

We also examined the accuracy of children's judgements based on the correctness of their predictions, that is, how well children judged correct versus incorrect predictions (i.e., whether children were more skilled in judging one versus the other). Considering trials with correct initial predictions and those with incorrect initial predictions as separate groups, the total number of times those predictions were either judged correctly or incorrectly was tallied. The proportion of correct and incorrect predictions judged accurately was then calculated.

Table 2.6 below illustrates the proportions of children's predictions that were judged accurately, based on whether or not those predictions were confirmed by testing items on the inclined plane. More specifically, the table includes the proportion of trials in which children both made correct initial predictions, as well as accurately judged those predictions (i.e., agreed that they were in fact right), out of the total number of trials in which children made correct initial predictions. The table also includes the proportion of trials in which children both made incorrect initial predictions, as well as accurately judged those predictions (i.e., agreed that those predictions were indeed wrong), out of the total number of trials in which children made incorrect initial predictions.

Table 2.6. *Mean Proportion of Correct Judgments, for both Correct and Incorrect Initial Predictions, by Age.*

	Diagram		Control	
	Correct P	**Incorrect P**	**Correct P**	**Incorrect P**
3 YOS	0.85(0.24)	0.65(0.48)	0.72(0.38)	0.60(0.44)
4 YOs	0.98(0.05)	0.73(0.29)	0.91(0.17)	0.59(0.40)
Full Sample	0.91(0.18)	0.69(0.40)	0.82(0.30)	0.59(0.41)

First, in looking only at those trials in which children made correct initial predictions, one sample t-tests against chance (2-tailed) found that, for both age groups, when children's predictions were correct, children were also above chance levels ($M = 0.78$, and $M = 0.95$, respectively) in judging the accuracy of those predictions, $t(29) = 4.95$ $p < .001$; and $t(29) = 18.72$, $p < .001$.

Analyses also examined the influence of condition, age and gender on the proportion of correct predictions judged accurately. A Mann-Whitney U test (2-tailed) found that the older group outperformed the younger group. The mean rank was 26.78 for the three-year-olds, compared to 34.22 for the four-year-olds, $U = 338.50$, $p = .049$. No difference was found between the diagram and control conditions, $U = 380.00$, $p = .22$ (2-tailed). Neither was there a difference between boys and girls, $U = 330.00$, $p = .19$ (2-tailed). See Figure 2.18 below.
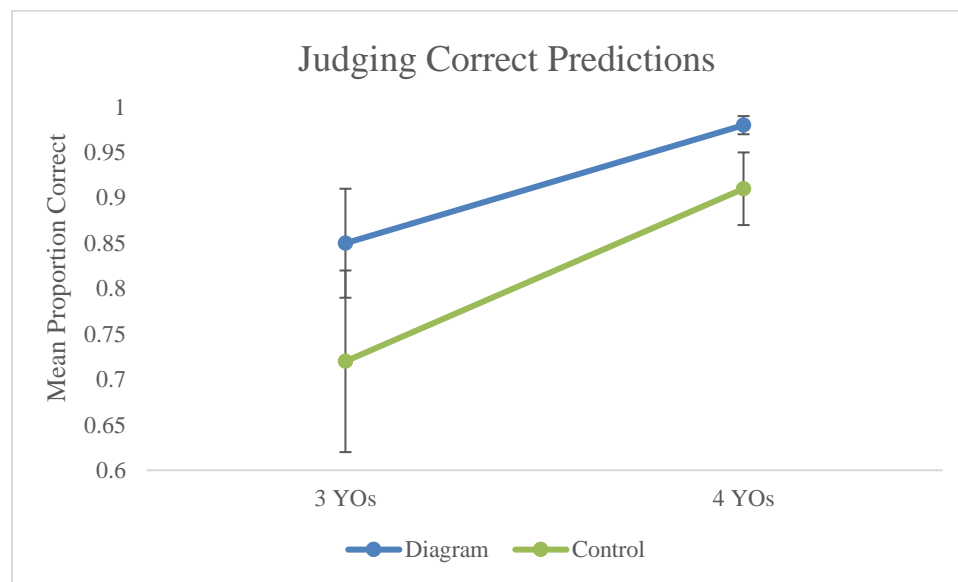
Figure 2.18. *Condition and age effects on children's judgments of their correct predictions.*

In summary, children were above chance levels in judging the accuracy of their correct predictions, and this was a skill that increased with age.

Next, analyses examined the accuracy of children's judgments for only those trials in which children made incorrect initial predictions. (Note: 6 children, 1 three-year-

old and 5 four-year-olds, made only correct initial predictions throughout the task; thus, these six children were not included in the following analyses).

One sample t-tests against chance (2-tailed) found that the four-year-olds, but not the three-year-olds, were above chance levels in judging the accuracy of their incorrect predictions (i.e., in agreeing that those predictions were in fact wrong). The mean proportion of incorrect predictions accurately judged was 0.62 ($SD = 0.45$) for the three-year-olds, and 0.66 ($SD = 0.34$) for the four-year-olds, ($t(28) = 1.47$, $p = .15$; and $t(24) = 2.38$, $p = .026$, respectively).

Analyses also examined the influence of condition, age and gender on the proportion of incorrect predictions accurately judged. Mann-Whitney U tests (2-tailed) found no group differences. For condition, $U = 317.50$, $p = .39$. For age, $U = 356.00$, $p = .90$. For gender differences, $U = 285.50$, $p = .44$. See Figure 2.19 below.
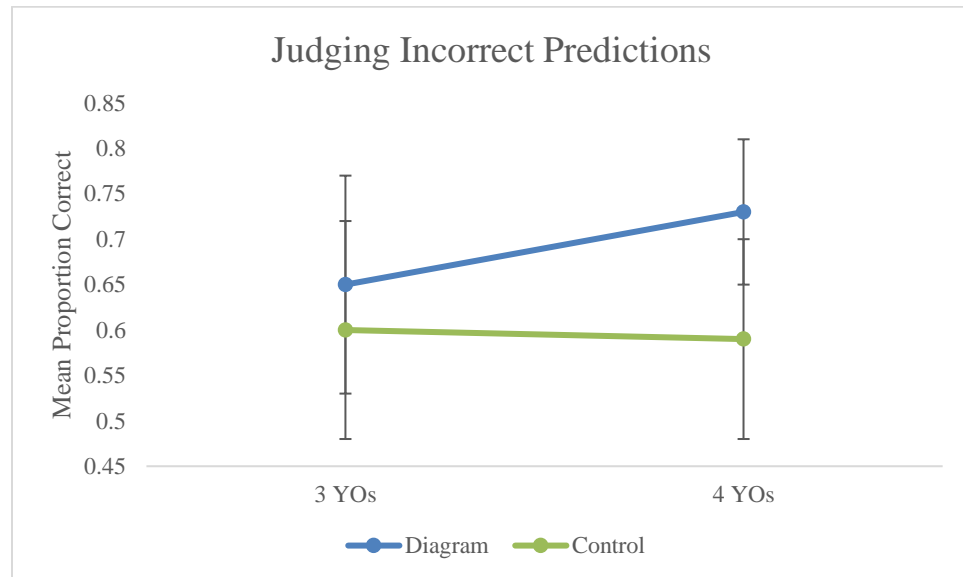
Figure 2.19. *Age and condition effects on children's judgments of their incorrect predictions.*

In summary, children showed difficulty in judging the accuracy of their incorrect predictions (i.e., in agreeing that those predictions were in fact wrong). Although the four-year-olds' judgments were above chance, their scores were not near ceiling levels. Further, though there was a trend for children in the diagram condition to outperform the control, this difference did not reach significance. This suggests that even with use of the inclined plane diagrams, young children have difficulty in judging the accuracy of their incorrect predictions.

Finally, analyses also explored whether children from each age group showed a difference in their ability to judge the accuracy of their correct versus incorrect initial predictions (i.e., whether the mean proportion of correct predictions accurately judged was higher than the mean proportion of incorrect predictions accurately judged). Among the three-year-olds, the mean proportion of correct predictions accurately judged was 0.78 ($SD = 0.31$), whereas the mean proportion of incorrect predictions accurately judged was 0.62 ($SD = 0.45$). A Wilcoxon Signed Ranks test found that this difference was not significant, $Z = 1.43$, $p = .15$ (2-tailed). Among the four-year-olds, the mean proportion of correct predictions accurately judged was 0.95 ($SD = 0.13$), compared to a mean proportion of incorrect predictions accurately judged of 0.66 ($SD = 0.34$). This difference was, however, significant, $Z = 3.30$, $p < .001$ (2-tailed). Thus, three-year-old children showed no difference in their ability to judge the accuracy of their correct and incorrect predictions. However, the four-year-olds were reliably more skilled in judging the accuracy of their correct predictions, compared to those which were not confirmed by testing items on the inclined plane.

2.6e Performance on Repeated Trials. Recall that children engaged in the process of predicting and checking two times sequentially for each of the six test items. Analyses also explored whether children performed better on the second attempt at predicting and checking for each item compared to the first attempt. The number of times that children predicted correctly on both the first and second attempts was tallied, as well as the number of times children predicted correctly on the first attempt but incorrectly on the second attempt. The number of times children predicted incorrectly on the first attempt and correctly on the second attempt was also tallied, as well as the number of times children predicted incorrectly on both the first and second attempts. Figure 2.20 below illustrates the proportions for each possible response pattern.
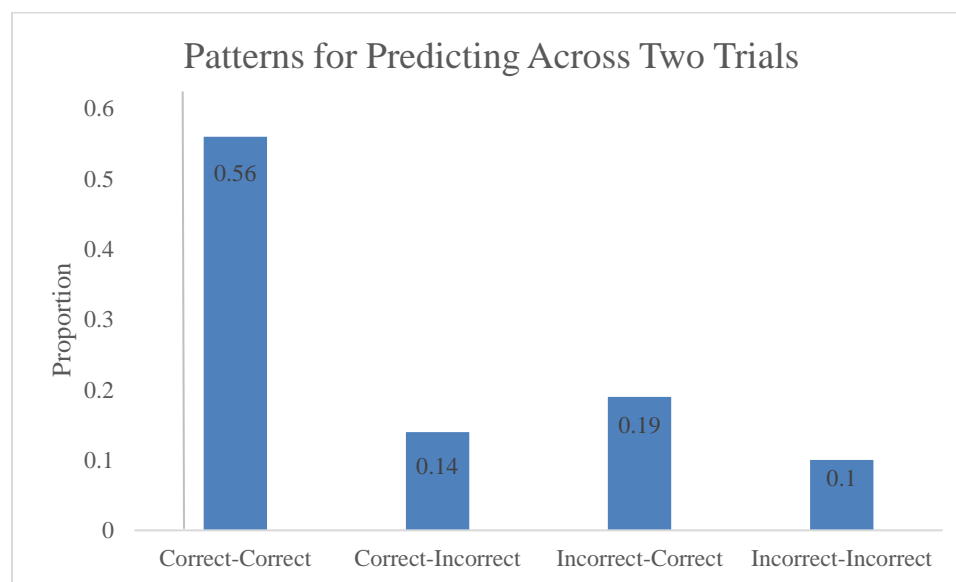
Figure 2.20. *Patterns for predicting across two trials (proportion for each pattern).*

For the trials in which children predicted correctly on the first attempt, it was possible for them to either predict correctly again on the second attempt, or to predict incorrectly on the second attempt. If children were more likely to predict correctly the

second time after making a correct initial prediction, it would suggest that children were taking prior knowledge and experiences into account when making predictions, rather than predicting at random. The mean number of trials for which children demonstrated the correct-correct pattern was 3.40/6.00 ($SD = 1.43$), whereas the mean number of occurrences for the correct-incorrect pattern was 0.90 ($SD = 1.05$). A Wilcoxon Signed Ranks test (2-tailed) found that this difference was significant, $Z = 5.76$, $p < .001$. Thus, children were significantly more likely to predict correctly on both trials for each item, than they were to predict correctly on the first trial and then incorrectly on the second.

When looking at those trials in which children predicted incorrectly on the first attempt, it was possible for them to predict either incorrectly again, or correctly on the second trial. The mean number of occurrences for the incorrect-correct pattern was significantly higher than was the mean for the incorrect-incorrect pattern; the mean for the incorrect-correct pattern was 1.12/6.00 ($SD = 0.94$), whereas the mean for the incorrect-incorrect pattern was 0.60 ($SD = 0.85$), $Z = 2.72$, $p = .007$ (2-tailed). Thus, after making an incorrect initial prediction on the first attempt with an item, children were significantly more likely to recognize their error and predict correctly on the next attempt, than they were to make the same mistake again. See Figure 2.21 below.
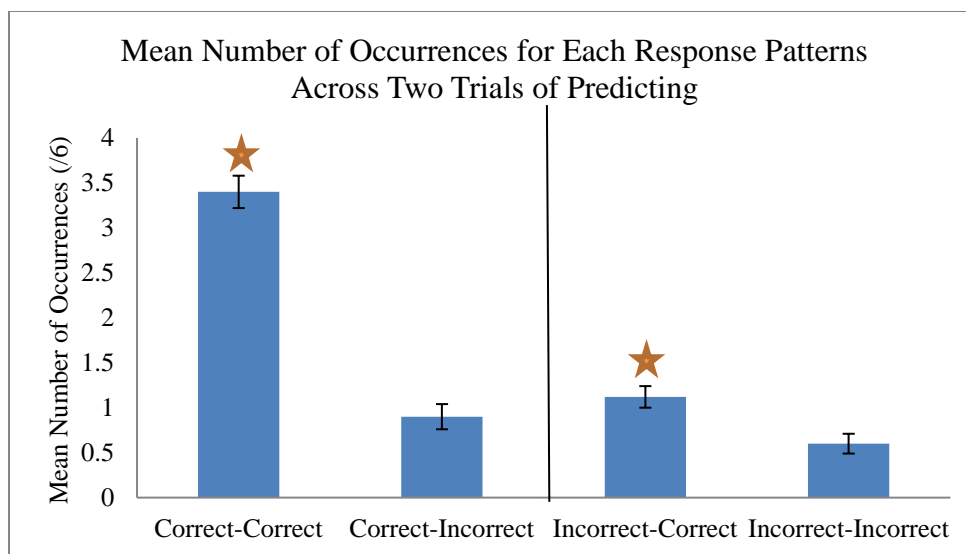
Figure 2.21. *Mean number of occurrences for each possible response pattern.*

2.6f Patterns in Responses. Analyses also explored patterns and/or consistencies in children's responses on the science reasoning activity. Again, the activity asked children to: (1). predict whether various items roll or slide down the inclined plane, (2). recall predictions after testing items on the ramp, and (3). judge the accuracy of those predictions. Note, as it was possible for children to demonstrate certain patterns and/or consistencies in any of these areas, the response patterns described below are not mutually exclusive.

Nineteen children (32%) achieved perfect prediction recall and judgment scores. Of these, 3 children (all from the older half of the sample), correctly predicted whether each object rolls or slides down the inclined plane, as well as correctly responded to all other questions. Thus, these children never needed to revise incorrect initial predictions (as their predictions were always confirmed by testing the items on the ramp); however, they still achieved perfect recall and judgment scores. The other 16 children also

received perfect recall and judgment scores, but for at least one item provided an incorrect prediction which needed to be revised. Of these, 6 were three-year-olds and 10 were four-year-olds. Thus, together, these 16 children always recognized when their prediction(s) was not confirmed and needed to be revised. Eleven other children (18%) made only one or two errors on the activity, either in recalling or judging predictions, suggesting that they understood the task.

When asked to recall their initial prediction immediately after testing an item on the inclined plane, another 6 children (10%) always stated the true outcome of testing the item on the ramp. For example, in cases in which these children predicted that an item rolls down the ramp and then observed that it slides, they always stated that they correctly predicted that the object would slide down (although they actually predicted that it would roll). Instead of providing their initial prediction, these children always stated what actually happened. Further, when asked to state the outcome of testing items on the inclined plane, another 3 children (5%) always stated that their prediction occurred.

Twelve children (20%) showed response biases when asked to judge the accuracy of their predictions. Ten children always said that their initial predictions were correct, even in cases in which they were not. Of these, 8 were three-year-olds, and 2 were four-year-olds. Another two children, both from the younger half of the sample, always stated that their initial predictions needed to be changed or revised, even when the initial predictions were correct. The remaining 16 children's (27%) errors revealed no systematic patterns or biases. See Table 2.7 below for a summary.

Table 2.7. *Patterns and Biases in Children's Responses on the Science Activity.*

Response	Number of Cases
All correct recall and judgments	19 (32%)
Only 1 or 2 errors, either in recalling or judging predictions	11 (18%)
Bias in Recalling Predictions: Always stated true outcome	6 (10%)
Bias in Stating Outcome: Always said prediction occurred	3 (5%)
Bias in Judging Predictions: Said all were right, or all were wrong	12 (20%)
Random errors; no systematic patterns or biases	16 (27%)

Relationships between Pre-test and Science Scores.

2.6g Pre-test Scores. Twenty-two children (37%) passed the theory of mind

(ToM) pre-test under the strict criterion of scoring 3 out of a possible of 3 points, whereas

38 (63%) did not pass. By age, 7 three-year-olds (23%), and 15 four-year-olds (50%)

passed the task. Further, 19 children (32%) passed the Dimensional Change Card Sort

(DCCS) pre-test under the strict criterion of scoring 4 out of a possible of 4 points,

whereas 41 (68%) did not pass. By age, 6 three-year-olds (20%), and 13 four-year-olds

(22%) passed the task. Finally, 16 children (27%) passed both pre-tests (5 three-year-

olds, and 11 four-year-olds). Mann-Whitney U tests (2-tailed) found no difference

between the diagram and control conditions in terms of the number of children in each

group who passed the pre-tests; for the theory of mind pre-test, $U = 435.00$, $p = .78$, and

for the DCCS pre-test, $U = 390.00$, $p = .29$.

2.6h Correlations between Pre-test Scores and Science Scores. Spearman's rho

correlations were used to explore the relationships between children's passing status on

the pre-tests, and their prediction recall and judgment scores. Results found that children

who passed the theory of mind pre-test tended to have higher prediction recall scores $r_s =$

$.40$, $p = .001$, as well as higher judgment scores $r_s = .33$, $p = .01$. Similarly, children

who passed the DCCS pre-test also tended to have higher prediction recall scores $r_s =$

.37, *p* = .004, as well as higher judgment scores r_s = .29, *p* = .02. Thus, children who passed the theory of mind and DCCS pre-tests tended to score higher on prediction recall, as well as judging predictions, compared to those children who did not pass the pre-tests. See Table 2.8 below for descriptive statistics.

Table 2.8. *Descriptive Statistics for Science Scores, based on Passing Status on the Pre-tests.*

Group	Mean Prediction Recall Score (SD)	Mean Judgment Score (SD)
Passed ToM (*n* = 22)	0.95 (0.08)	0.90 (0.15)
Failed ToM (*n* = 38)	0.84 (0.16)	0.78 (0.19)
Passed DCCS (*n* = 19)	0.96 (0.06)	0.87 (0.22)
Failed DCCS *(n* = 41)	0.84 (0.16)	0.80 (0.17)

2.6i Effect of Diagrams on Children's Science Scores, Based on their Passing Status on the Pre-tests. Analyses also examined the effect of the inclined plane diagrams on children's prediction recall and judgment scores, based on children's passing status on the pre-tests. First, among children who passed the theory of mind pre-test, the mean *prediction recall score* was 0.97 (*SD* = 0.05) for those in the diagram condition (*n* = 9), compared to a mean of 0.94 (*SD* = 0.08) for those in the control condition (*n* = 13). A Mann-Whitney U test (2-tailed) found that this difference was not significant, *U* = 44.00, *p* = .26. Further, in comparing *judgment scores* among the groups, the mean was 0.94 (*SD* = 0.09) for those in the diagram condition, compared to a mean of 0.87 (*SD* = 0.17) for those in the control. This difference also was not significant, *U* = 43.50, *p* = .27.

However, among children who did not pass the theory of mind pre-test (*n* = 38), those in the diagram condition reliably outperformed those in the control condition in both recalling and judging predictions. The mean *prediction recall score* for those in the diagram condition was 0.90 (*SD* = 0.09, *n* = 21), compared to a mean score of 0.75 (*SD* = 0.19) for those in the control condition (*n* = 17). A Mann-Whitney U test (2-tailed)

confirmed that the difference between the groups was significant. The diagram condition (*Mdn* = 0.92, mean rank = 23.45) reliably outperformed the control (*Mdn* = 0.75, mean rank = 14.62), *U* = 95.50, *p* = .01. For *judgment scores,* the mean for children in the diagram condition was 0.85 (*SD* = 0.14), compared to a mean of 0.70 *(SD* = 0.23) for the control condition. A Mann-Whitney U test (2-tailed) again found this difference to be significant, with the diagram condition (*Mdn* = 0.92, mean rank = 23.07) again reliably outperforming the control (*Mdn* = 0.67, mean rank = 15.09), *U* = 103.50, *p* = .03. See Figures 2.22 and 2.23 below.
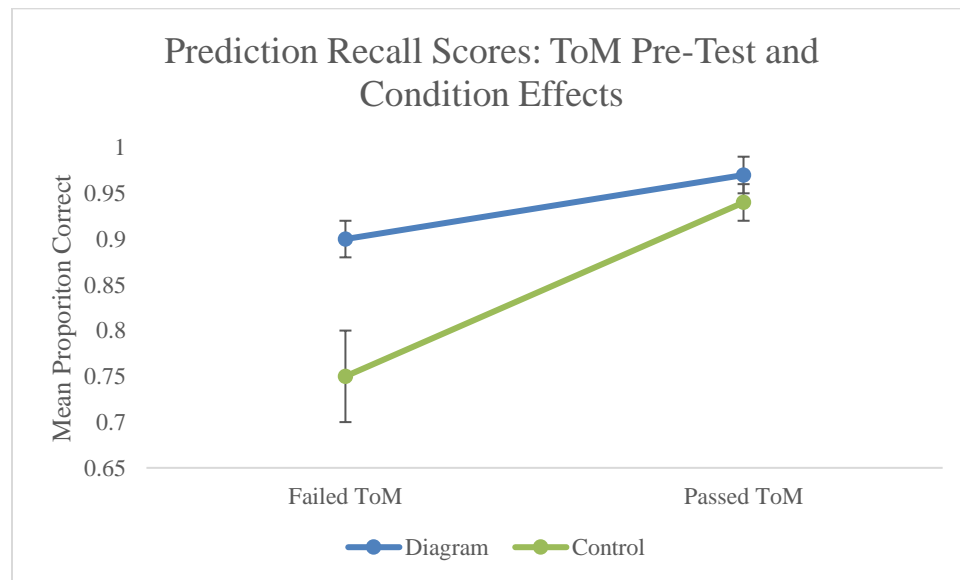
Figure 2.22. *Condition and ToM pre-test effects on prediction recall scores*.
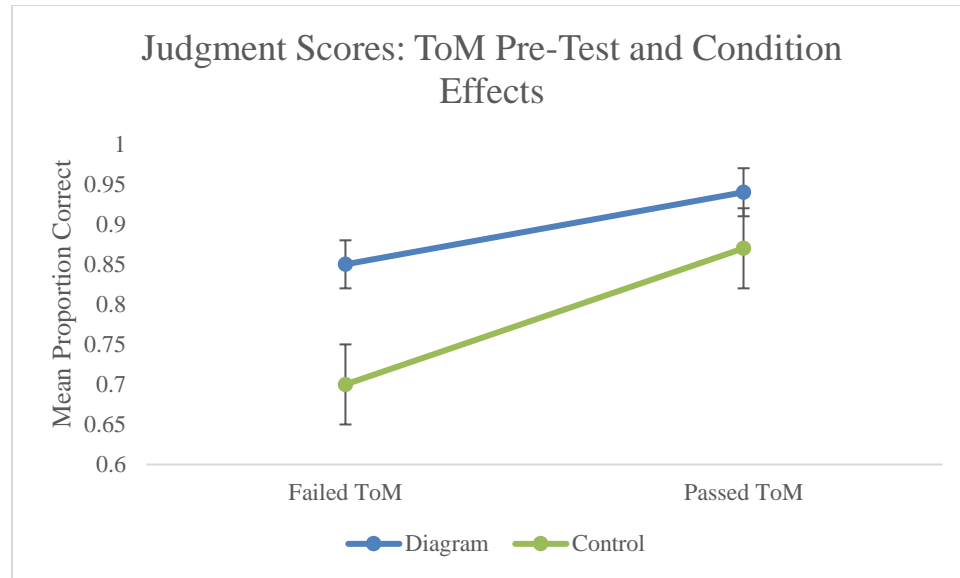
Figure 2.23. *Condition and ToM pre-test effects on judgment scores.*

Analyses also explored the effect of the inclined plane diagrams on children's prediction recall and judgment scores, based on children's passing status on the Dimensional Change Card Sort pre-test. Among the children who passed the DCCS pre-test, the mean *prediction recall score* was 0.97 (*SD* = 0.04) for those in the diagram condition *(n* = 9), compared to a mean of 0.94 (*SD* = 0.07) for those in the control (*n* = 10). A Mann-Whitney U test (2-tailed) found that the difference between the groups was not significant, *U* = 34.50, *p* = .33. Similarly, when looking at *judgment scores* in children who passed the DCCS pre-test*,* the mean was 0.94 (*SD* = 0.14) for those in the diagram condition, compared to a mean of 0.81 (*SD* = 0.26) for the control condition. This difference also was not significant, *U* = 26.50, *p* = .09.

However, among those who did not pass the DCCS pre-test (*n* = 41), those in the diagram condition reliably outperformed those in the control condition in accurately recalling, but not in judging, predictions. The mean *prediction recall score* for those in the diagram condition was 0.90 (*SD* = 0.09, *n* = 21), compared to a mean of 0.78 (*SD* =

0.19) for the control (*n* = 20). A Mann-Whitney U test (2-tailed) confirmed that the

difference between the two groups was significant, with the diagram condition *(Mdn =*

0.92, mean rank = 24.86) reliably outperforming the control (*Mdn* = 0.79, mean rank =

16.95), *U* = 129.00*, p* = .03. Further, the mean *judgment score* for those in the diagram

condition was 0.85 (*SD* = 0.12), compared to a mean of 0.75 (*SD* = 0.20) for those in the

control. The difference between the groups was not significant, *U* = 147.50*, p* = .10*,* 2-

tailed. See Figures 2.24 and 2.25 below.
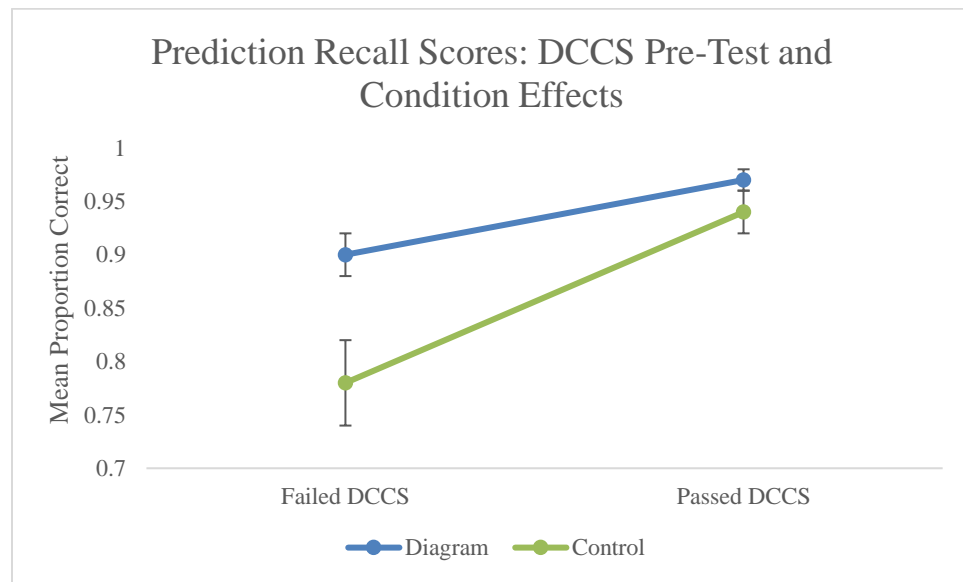
Figure 2.24. *Condition and DCCS pre-test effects on prediction recall scores.*
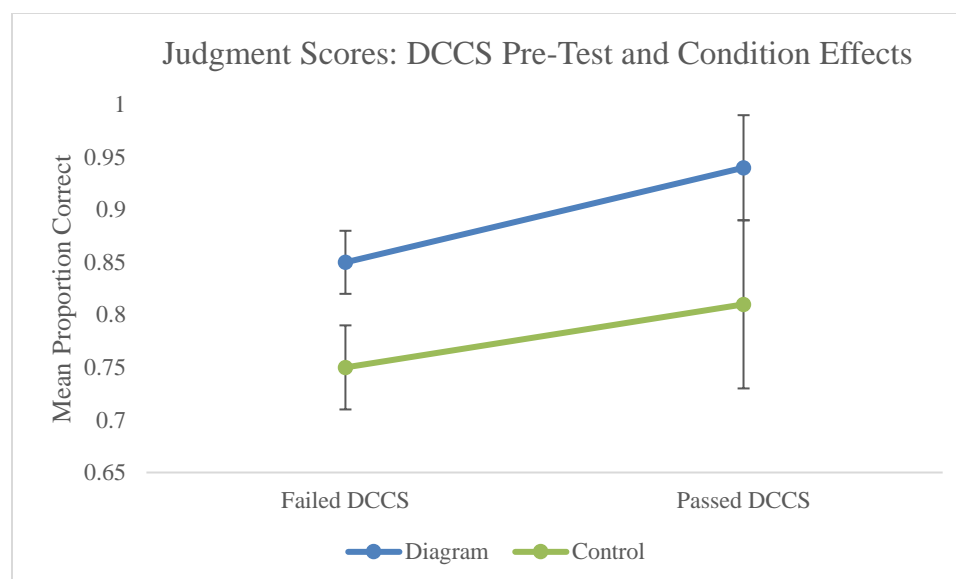
Figure 2.25. *Condition and DCCS pre-test effects on mean judgment scores.*

In summary (though with small samples), results found that, among children with stronger theory of mind skills, condition had no effect on children's prediction recall and judgment scores. However, use of the inclined plane diagrams was associated with reliably higher prediction recall and judgment scores among children with less well-developed theory of mind skills. Similarly, among children with stronger executive functioning skills, condition had no effect on children's prediction recall or judgment scores. However, among children with less-well developed executive functioning skills, use of the inclined plane diagrams was associated with significantly higher prediction recall, but not judgment scores. These results indicate that such diagrams may serve as effective pedagogical tools for supporting early forms of scientific thinking in the early childhood classroom, especially among children with less well-developed theory of mind and executive functioning skills.

2.7 Conclusions and Implications for Early Childhood Education

The research results presented here show that preschool children ages three and four years can use the scientific practices of predicting and checking. When asked to predict the movement of objects traversing an inclined plane, even three-year-old children's predictions were above chance levels. The children in the study also accurately recalled their initial predictions after testing items on the inclined plane, with this skill increasing over the preschool years. Children were also capable of judging the correctness of their initial predictions, with this skill too increasing over the preschool years. Of special interest is the finding that the children benefited from the option to use a relevant physics diagram during the science investigation.

These findings indicate that we used a suitable context for investigating the ability of preschool-aged children to engage in scientific investigations. The context also constitutes a developmentally-appropriate set of conversational options, objects and diagrams that can support science learning in the early childhood classroom. Since the materials are neither too easy nor too difficult, the teacher has room to help children actively construct understanding on what they already know something about.

Constructivist approaches to learning maintain that knowledge is actively constructed through experience and by building on relevant prior knowledge (i.e., Gelman & Williams, 1998; Piaget, 1985). That children performed well in all areas of the task indicates that they used relevant prior knowledge to support their performance. However, though children's performance was above chance, it was not at ceiling in some areas. For example, children's errors in making predictions tended to be systematic (i.e., errors were most frequent for the items of most complex design, such as those with

wheels which did not rotate). Children also showed some difficulty in confirming that their incorrect predictions were indeed wrong.

Teachers can foster learning in these areas by offering children further opportunities to explore the materials in meaningful ways. Children's learning can also be scaffolded by building on what they already know, and providing support at the appropriate level (i.e., Vygotsky, 1978). Direct experience, making mistakes, and searching for solutions are all key components for the assimilation and accommodation of new information (i.e., Piaget, 1985). To support such processes, teachers can provide hints when necessary (i.e., by pointing out when objects do not move down the inclined plane as expected, and asking children why). Teachers can also highlight and discuss design features of objects which enable or prohibit rolling, etc.

The inclined plane diagrams can also be used to scaffold children's learning. Both children's prediction recall, and the accuracy of their judgments, benefitted from use of the diagrams. This suggests that, from a young age, children are perhaps capable of using and understanding forms of diagrammatic representation in physical science to support their predicting and checking during scientific investigations. Use of the diagrams deserves further research attention, as they may be effective pedagogical tools for fostering young children's development of key scientific thinking skills from an early age. Further, such diagrams can be easily translated into practice. This is significant, as little prior work has explored early forms of preschool literacy related to science learning, as well as children's ability to understand symbolic, external representations of predictions and outcomes for scientific investigations. It is worth pursuing whether repeated practice with such relevant diagrams during scientific investigations, in a variety

of contexts, fosters young children's emerging scientific thinking skills. It is also worth considering other possible means of supporting children's representation of, and reflection upon, their scientific investigations. For example, children may benefit comparably from documenting their science investigations using other media formats, such as by drawing in science journals or creating other representations of their thinking.

The children in the study also benefitted from opportunities for repeated practice in predicting and checking. For example, when children's initial predictions were incorrect, they were significantly more likely to recognize their error after testing the item on the ramp, and then revise their prediction for the second attempt than they were to predict incorrectly again. In other words, children appeared to understand when they made a mistake on the first attempt at predicting with an item, and they changed their prediction to the correct one for the second attempt. This has a clear pedagogical implication, namely that young children may benefit from opportunities to reflect on scientific investigations (i.e., in ways such as the physics diagrams used in this study prompt them to do), as well as opportunities to try again as they explore with different items and materials. This, too, is worth further research exploration.

Finally, this study also finds preliminary support for the hypothesis that children's capacity for early scientific thinking may relate to their developing theory of mind and executive functioning capacities, as children who passed the theory of mind and executive function pre-tests tended to also score higher on recalling and judging predictions. In other words, children's theory of mind and executive function pre-test scores positively correlated with their prediction recall and judgment scores.

Importantly, use of the inclined plane diagrams during the science activity benefitted those children with weaker theory of mind and executive function skills, both in terms of the accuracy of children's prediction recall, and their judgments. This again indicates that the diagrams may serve as effective instructional scaffolds (i.e., Vygotsky 1978) for supporting scientific predicting and checking in children as they are acquiring critical theory of mind and executive function skills associated with scientific thinking. It is worth continuing to explore the relationships among theory of mind, executive functioning and scientific thinking, as well as the use of relevant diagrams to support children's development of such skills. To preview, this is one goal for the second study contained in this dissertation.

Although this work found correlations among preschoolers' early scientific thinking skills, and their developing theory of mind and executive functioning abilities, it is important to note that it has not established causal relationships among these variables. It is still possible, for example, that a mediating variable(s), such as children's language development, is responsible for the correlations among these variables. For instance, understanding the kinds of questions and language involved in the pre-tests may have led to the differences in children's performance on the science activity, rather than differences in their theory of mind and/or executive function capacities per se. Thus, more work is needed to clarify the nature of the relationships among young children's capacity for early forms of scientific thinking and their developing theory of mind and executive functioning skills. It is important for future research to continue exploring these ideas.

Chapter 3: Study 2

Preschoolers' Scientific Reasoning: Simple Physics Diagrams of a Balance Scale Aid their Ability to Predict and Check

3.1 Introduction.

 This paper presents the second study contained in the dissertation. It represents an effort to determine whether findings from the prior study replicate in a different context of scientific investigation, and with different test items and materials. To recap, the previous study found that three- and four-year-old children are indeed capable of accurately predicting and checking their predictions in the context of investigating objects traversing an inclined plane. Further, children's abilities to accurately recall, as well as judge the correctness of, their initial predictions benefitted from use of inclined plane diagrams as tools for supporting external representation of, and reflection upon, their scientific investigations. Finally, these abilities (i.e., to recall, and judge the correctness of, initial predictions), also correlated with children's theory of mind and executive functioning scores. (See Liberti-Reuter, 2016 for more details).

 The current study represents a continued exploration of these questions, but in the context of young children's investigations of the relative weights of various objects, with use of a balance scale. Specifically, the study explores children's ability to make appropriate predictions in such context, as well as their capacity to determine whether or not their predictions are confirmed after testing items on the scale. Further, analogous to the inclined plane diagrams developed for the previous study, new diagrams of a balance scale were created to determine whether they are similarly effective in supporting young children's ability to recall, as well as judge the accuracy of, their initial predictions. Finally, this study continues to explore the relationships among children's developing theory of mind and executive functioning skills, and their abilities to engage in early forms of scientific thinking. These goals are briefly outlined below.

3.2 Research Goals.

  1. This study explores young children's ability to predict and check during scientific investigations involving the balance scale, a simple measurement tool commonly found in the preschool classroom. After a brief whole-class introduction to the scale and its function, children engaged in an individual reasoning activity using the balance scale. Children were provided a number of familiar items, two at a time, and were asked to predict whether placing the items on each side of the scale would result in the scale balancing. Children then tested the items on the scale, watched what happened, and finally, were asked to reflect on the outcome and judge the accuracy of their initial predictions. The aim was to explore children's predicting and checking skills in this particular context to check our assumption that the task would be suitable for fostering scientific thinking in the early childhood classroom. An additional goal was to examine the developmental trajectory of such scientific thinking skills over the preschool years.

  The balance scale was chosen as a context for scientific thinking because it is a measurement tool often found in the preschool classroom. The scale is also commonly included in standards for preschool mathematics and science education. For example, the National Council of Teachers of Mathematics has explicit measurement standards for prekindergarten through second grade (NCTM, 2000), and the Framework for K-12 Science Education (NRC, 2012) includes measurement in the science and engineering practice "Using Mathematics and Computational Thinking." The balance scale is also included in New Jersey's Department of Education Preschool Teaching and Learning Standards (2014). For example, mathematics standard 4.3 aims for children to begin to conceptualize measurable attributes of objects through learning opportunities with

measurement tools. Science standard 5.2 relates to observing and investigating matter

and energy, and suggests the provision of a variety of interesting objects and materials, as

well as simple tools (i.e., balance scales, magnifiers), to observe, manipulate, sort and

describe them. Finally, science standard 5.5 concerns providing children experience in

using tools and technology in support of their scientific investigations (i.e., computers,

measurement tools, writing and drawing tools, simple machines, etc.).

As indicated by the content standards described above, the chosen context is

assumed appropriate for supporting both science and mathematics learning in the early

childhood classroom. It was hypothesized that, from a young age, preschoolers would be

capable of appropriately predicting and checking in this context. This is because research

indicates that with just a small amount of input, young children are capable of

understanding and reasoning about some uses of a balance scale (i.e., Halford, Andrews,

Dalton, Boag, & Zielinski, 2002; Jansen & van der Maas, 2002; Kliman, 1987; Siegler &

Chen, 1998).

2. As previously mentioned, the prior study in this dissertation found evidence

that diagrams of an inclined plane serve as effective pedagogical tools for supporting

young children's abilities to both recall, as well as judge the correctness of, their initial

predictions when engaging in scientific investigations involving the motion of objects

down an incline. Thus, an important question remains, namely whether this finding

extends to other contexts, or rather is dependent upon the specific diagrams/stimuli used

in the previous study. Accordingly, the current study explores whether new diagrams,

those of a balance scale, are comparably effective in supporting these critical early

scientific thinking skills in young children.

Thus, abstract diagrams of a balance scale were designed to assist children in externally representing and reflecting upon their predictions and observations as they engaged in the balance scale activity. The diagrams were developed to represent physical science concepts concerning the relative weights of objects using the balance scale, and they may serve as design tools for use by teachers in the early childhood classroom, both to keep track of children's knowledge and to offer support of their thinking about science activities. (See balance scale diagrams in Figure 3.1, below). The goal was to determine whether children are capable of understanding the diagrammatic representation characterized by the stimuli, as well as whether use of the diagrams supports children's ability to predict and check in this physical science context. It was hypothesized that use of the diagrams would extend and support children's developing cognitive capacities (i.e., theory of mind and executive function skills) by helping children to keep predictions in mind, as well as compare predictions and outcomes in order to assess the accuracy of those predictions. In other words, that use of the diagrams would support children's ability to recall, as well as judge the accuracy of, their initial predictions.
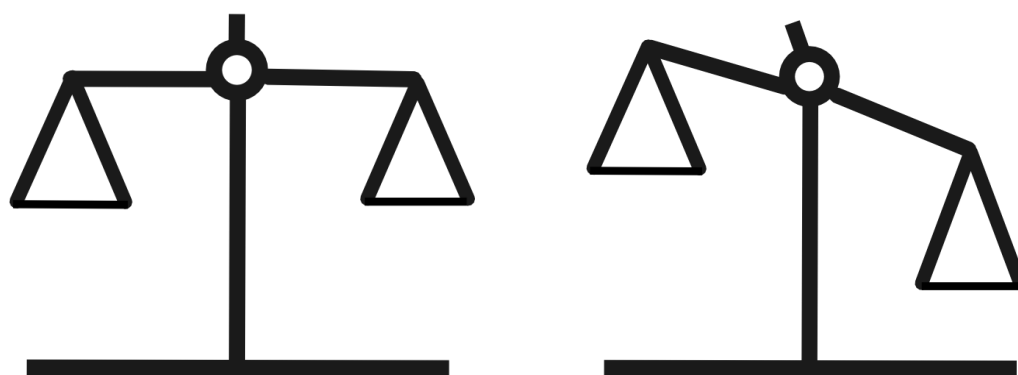
Figure 3.1. *Physics diagrams of a balance scale, representing the concepts of equal and unequal weights.*

3. Finally, it is also important for the field of Early Childhood Education to identify elements of young children's thinking that can buttress the foundation for developing scientific reasoning in the early years. As the previous study in this dissertation found that children's scientific thinking skills positively correlated with their theory of mind and executive function abilities, another goal for this research was to determine whether such correlations replicate in this new context for scientific thinking.

Predicting and checking predictions requires the ability to hold predictions and observations in mind, as well as judge the accuracy of one's initial predictions after engaging in scientific manipulations. It was hypothesized that children with more advanced theory of mind and executive function skills would perform better on the science reasoning activity than those with less advanced skills in these areas. This is because these early developing cognitive skills may relate to children's ability to recall initial predictions after testing items on the scale, as well as to determine whether their predictions were confirmed. In other words, it was hypothesized that children's theory of mind and executive function scores would positively correlate with their science scores. (Again, see Liberti-Reuter, 2016 for a more detailed discussion of these hypotheses).

3.3 Research Questions.

To summarize, the research questions guiding this work include:

- 1). *How accurately do preschool children make and check predictions about the relative weights of objects using a balance scale? Further, how do these skills develop over the preschool years?*
- 2). *Do preschoolers' theory of mind and executive functioning capacities relate to their abilities to recall, as well as judge the accuracy of, their initial predictions in this context?*

- 3). *Does use of simple, abstract diagrams of a balance scale, for encouraging external representation of, and reflection upon, predictions and observations support preschool children's ability to recall, as well as judge the accuracy of, their initial predictions?*

3.4 Experimental Details.

3.4a Participants. The children for the study were drawn from two Central New Jersey preschools. The first is a private preschool situated on an expansive, park-like property with nature trails, a small working farm, and a large park-like playground. Children attending the school were predominantly Caucasian and from families of middle-class socio-economic backgrounds. The school has five classrooms for children ranging in age from 2 ½ to 6 years. Programs include half day and full day preschool classes, as well as a full day kindergarten. All of the teachers, as well as many of the assistant teachers, have a degree in Elementary Education. The school's mission statement highlights the importance of providing a safe and nurturing environment to support children's cognitive, social, emotional and physical development, as well as encouraging children's learning by doing. The school does not follow a specific curriculum; however, children at all levels are regularly introduced to concepts in literacy, mathematics, science, and the arts, with an emphasis on learning through play.

The second school is a preschool program located within a public, Central New Jersey high school. The program is part of the high school's Child Development program. High school students have the option to enroll in the course as an elective, and the course involves them in running the preschool program. Under the direction of the classroom instructor, the high school students assist with the design and implementation of instruction.

The preschool program comprises half day sessions which are open to 3- to 5-year-old children in the community. Preschool children attend free of charge, on a first-come, first-served basis. Children attending the school were predominantly Caucasian and from families of middle-class socio-economic backgrounds. The school does not follow a specific curriculum; however, each week includes learning experiences within the areas of literacy, mathematics, science, social studies, and the arts.

Once the work received IRB approval, parents of all 3- to 5-year-old children at both schools were provided information about the study and offered consent forms. Parents returned signed consent forms for 107 children. Some children were absent on testing dates, and one child did not want to participate, resulting in a final sample of 96 children (47 boys, 49 girls). The mean age was 53 months, with a range of 36 to 67 months. All children spoke English. No child participated in Study 1 in this dissertation.

3.4b Materials. Items for the experiment consisted of a balance scale designed for use by children, as well as a variety of small items of differing weight to test on the scale. Test items were familiar to children, and included objects such as rocks, shells, cotton balls, acorns, crayons, blocks, etc. Other materials included a felt board and the balance scale diagrams described above, printed on image cards, to represent and record children's predictions and observations during the science activity (See Figure 3.1). Figure 3.2, below, illustrates the ten pairs of test items used for the study. Figure 3.3 portrays the balance scale used in the study, and Figure 3.4 illustrates the felt board chart used to record predictions and observations with the balance scale diagrams.

Pair Numbers and Test Items for Each Pair

1. Paint & PlayDoh (Unequal Weights)

2. Two Acorns (Equal Weights)

3. Shell & Rock (Unequal Weights)

4. Wooden Blocks (Unequal Weights)

5. Plastic Blocks (Equal Weights)

6. Two Cotton Balls (Equal Weights)

7. Two Shells (Unequal Weights)

8. Acorn & Pebble (Unequal Weights)

9. Two Crayons (Equal Weights)

10. Two Rocks (Equal Weights)

Figure 3.2. *Pair numbers, and items used for each pair.*

Figure 3.3. *Children's balance scale.*
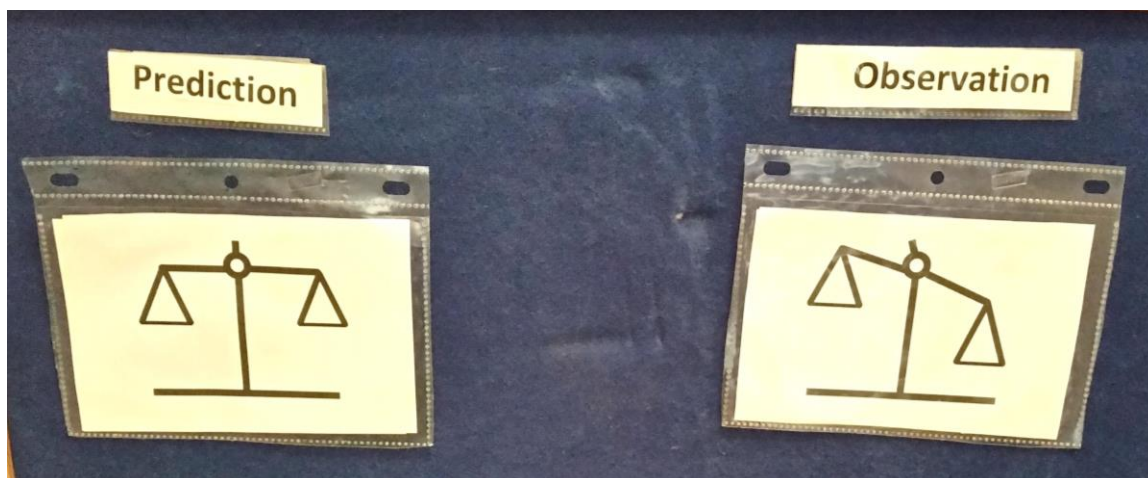
Figure 3.4. *Felt board with balance scale diagrams printed on image cards for representing and reflecting upon predictions and observations. (Note: In this example, the child incorrectly predicted that two items would balance on the scale, but then observed that one side of the scale went down).*

3.4c Methods. The experimenter first spent a few days in the classroom for children to become familiarized and comfortable with her. Next, two pre-tests were administered to obtain initial measures of children's executive function and theory of mind capacities: (1) a standard Dimensional Change Card Sort task (DCCS), and (2) a standard theory of mind task. The DCCS (Frye, Zelazo, & Palfai, 1995) is a commonly used, developmentally-sensitive measure of executive function for young children. This rule-based, card sorting task involves a combination of working memory and inhibition. Children are first shown two boxes with target cards affixed to the boxes (e.g., a blue boat and a red rabbit). Children are presented a series of cards (e.g., red and blue rabbits and boats) and are first asked to sort by shape, and then by color (in counterbalanced order). The final phase asks children to switch rules on each trial. Previous work shows that three-year-olds rarely pass the task, whereas four-year-olds are just above chance

levels, and five-year-olds are near ceiling (i.e., Carlson, 2005). (See Appendix A for full scripts and procedure for the Dimensional Change Card sort pre-test).

A standard and well-established theory of mind task was also administered (i.e., Gopnik & Astington, 1988). Children were presented with a deceptive item (a crayon box containing Band-Aids™). The item was first presented in its deceptive state, and it was assumed that children would represent the item incorrectly. The experimenter then revealed the item's true state by opening the box and showing the child what was inside. The item was then returned to its deceptive state and taken out of the child's reach. The false belief question asked children what a naïve observer (i.e., a classmate who had not yet had a turn to play the game) would think the item contained. The representational change question asked children what they thought the box contained before its true state was revealed. Finally, the appearance-reality task involved a reality question ("What's really and truly in the box, Band-Aids™ or crayons?), and an appearance question ("What does this look like it has in it, Band-Aids™ or crayons?"). (See Appendix B for full scripts and procedure for the Theory of Mind pre-test).

Next, (and on a different day), children were introduced to the balance scale during an interactive circle time classroom lesson for all students led by the experimenter. The lesson was used to familiarize children with the balance scale, including its function, the ways in which it can be used, and the information that it provides about the relative weights of objects. During the lesson, children took turns making predictions about the comparative weights of various items, and the experimenter helped them to check their predictions using the balance scale. For example, children were asked to first "use their muscles" to determine which of two objects was heavier, and then the balance scale was

used to check their judgments. The scale was also used to compare the weights of objects when the felt weights were too similar to determine which was heavier, as well as items of equal weight.

Finally, children engaged in the balance scale reasoning activity. The activity involved interviewing children individually in a quiet area of their classroom, and asking them to reason about the balance scale. Children were shown various objects, two at a time, and were asked to predict what would happen if each item was placed on one side of the scale (i.e., whether the scale would balance, or if one side would go down). Children had the opportunity to handle and explore the items before making a prediction. After making each prediction, children tested the items on the scale, watched what happened, and then revisited their original prediction to determine its accuracy.

Controlling for age, children were randomly assigned to either the diagram condition or the control condition. During the science activity, children in the diagram condition used the felt board and balance scale diagrams to externally represent and reflect upon their predictions and observations. Using the diagram image cards, children recorded each prediction on the felt board, as well as the outcome of testing each set of items on the scale. They were able to then use the diagrams as instructional aids to support their recall of initial predictions, as well as their judgment regarding the accuracy of those predictions. In contrast, children in the control condition did not have access to the diagrams. Instead, they made their predictions and observations verbally. See questions and prompts used for the balance activity in Table 3.1 below. (Also see Appendix D for complete scripts and procedures for the balance scale activity).

Table 3.1. *Prompts/Questions used for the Balance Scale Activity, by Condition.*

Diagram Condition

Make prediction prompt- (Experimenter hands child two items). "First, predict, or guess. If you put them on the scale, will it balance, like this (demonstrate with hands), or will one side go down, like this (demonstrate with hands)? Okay. Let's put your prediction on the chart."

Test and observe outcome prompt- "Now it's time to try it on the scale. (Child places items on each side of the scale). Did it balance, like this (demonstrate with hands), or did one side go down, like this (demonstrate with hands)? (Counterbalanced). Okay. Let's put it on the chart." (Experimenter helps child place observation diagram card on the felt board).

Prediction recall prompt- "Before we tried it on the scale, what was your prediction, or guess? Did you guess balance? Or that one side would go down (counterbalanced)? You can use the chart to help you remember."

Judgment prompt- "Was your prediction right, or do you need to change it (counterbalanced)? You can use the chart to help you remember."

Control Condition

Make prediction prompt- (Experimenter hands child two items). "First, predict, or guess. If you put them on the scale, will it balance, like this (demonstrate with hands), or will one side go down, like this (demonstrate with hands)? Okay."

Test and observe outcome prompt- "Now it's time to try it on the scale. (Child places items on each side of the scale). Did it balance, like this (demonstrate with hands), or did one side go down, like this (demonstrate with hands)? Okay."

Prediction recall prompt- "Before we tried it on the scale, what was your guess, or prediction? Did you guess balance? Or that one side would go down?"

Judgment prompt- "Was your prediction right, or do you need to change it?"

3.5 Analyses.

Scores on the Dimensional Change Card Sort (DCCS) pre-test were assigned

following the procedures described in Frye, Zelazo, & Palfai (1995). Children were first

taught to sort cards according to one dimension (e.g., color) until they correctly sorted at

least four out of five cards according to that dimension. Next, children were taught to

sort according to the other dimension (e.g., shape). Finally, children were asked to sort

four cards, switching rules on each trial. Children received one point for each card

correctly sorted, resulting in a total score between zero and four points.

Scoring for the theory of mind pre-test followed the procedures described by

Gopnik & Astington (1988). Children were scored as having passed the representational

change question if they correctly reported their initial representation of the object (i.e., "I

thought it had crayons"). Children passed the false-belief question if they said that a

classmate who had not seen the box would think it had crayons in it. To pass the

appearance-reality question, children had to answer both parts correctly; they had to

answer that: (1). the box looks like it has crayons in it, but (2). it really contains Band-

Aids™. Thus, each child could receive a score between zero and three points.

The balance scale activity involved reasoning about the outcome of placing

various items on each side of the scale. Children were scored on: (1). the accuracy of

their initial predictions, (2). their ability to accurately recall initial predictions (regardless

of whether those predictions were confirmed or disconfirmed); and (3). their capacity to

judge the correctness of those initial predictions. In other words, children received three

separate scores: (1). a Prediction Accuracy score, based on the total number of correct

predictions made over the course of the session; (2). a Prediction Recall score, based on

the number of initial predictions accurately recalled; and (3). a Judgment Score, based on

the total number of predictions correctly judged as right (i.e., for those which were

confirmed by testing items on the scale), or wrong (i.e., for those which were

disconfirmed by testing items on the scale). Raw scores were then converted to

proportion correct.

Analyses examined overall mean scores, as well as group differences in children's performance based on age, condition and gender. Patterns and consistencies in children's responses were also identified.

Analyses also explored the relationships between children's pre-test scores and their scores on the balance scale reasoning activity, specifically, by examining correlations among children's passing status on the pre-tests and their ability to recall predictions after testing items on the scale, as well as judge the accuracy of those predictions. Finally, analyses also examined whether use of the balance scale diagrams influenced children differently depending on their passing status on the pre-tests. In other words, whether or not the diagrams showed similar effects in children of differing levels of development concerning executive function and theory of mind skills.

3.6 Results

3.6a Accuracy of Children's Initial Predictions. Analyses examined children's ability to accurately predict whether or not placing various items on each side of the balance scale would result in the scale balancing. As the activity involved 10 trials of predicting and checking, each child was assigned a prediction accuracy score (out of 10), based on the total number of correct predictions made over the course of the activity. See distributions of prediction accuracy scores in Figures 3.5 through 3.7 below.
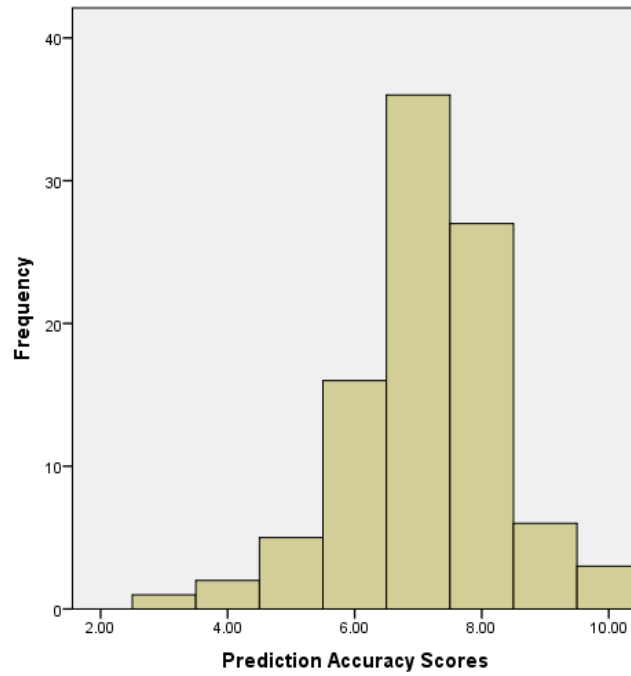
Figure 3.5. *Distribution of prediction accuracy scores for the full sample (n = 96).*
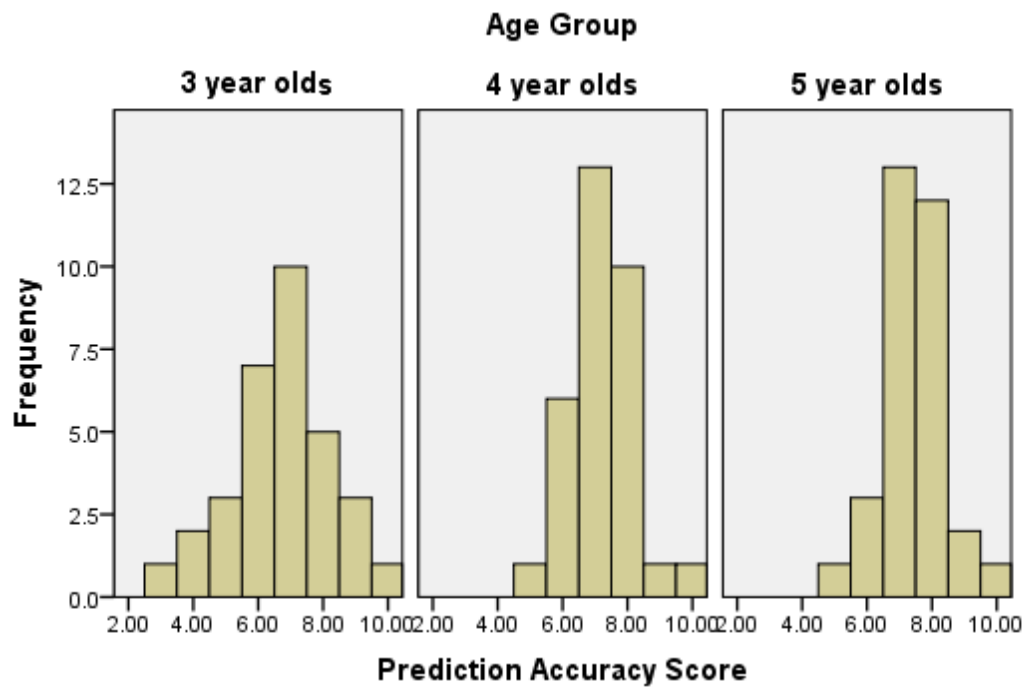
Figure 3.6. *Distribution of prediction accuracy scores, by age group.*
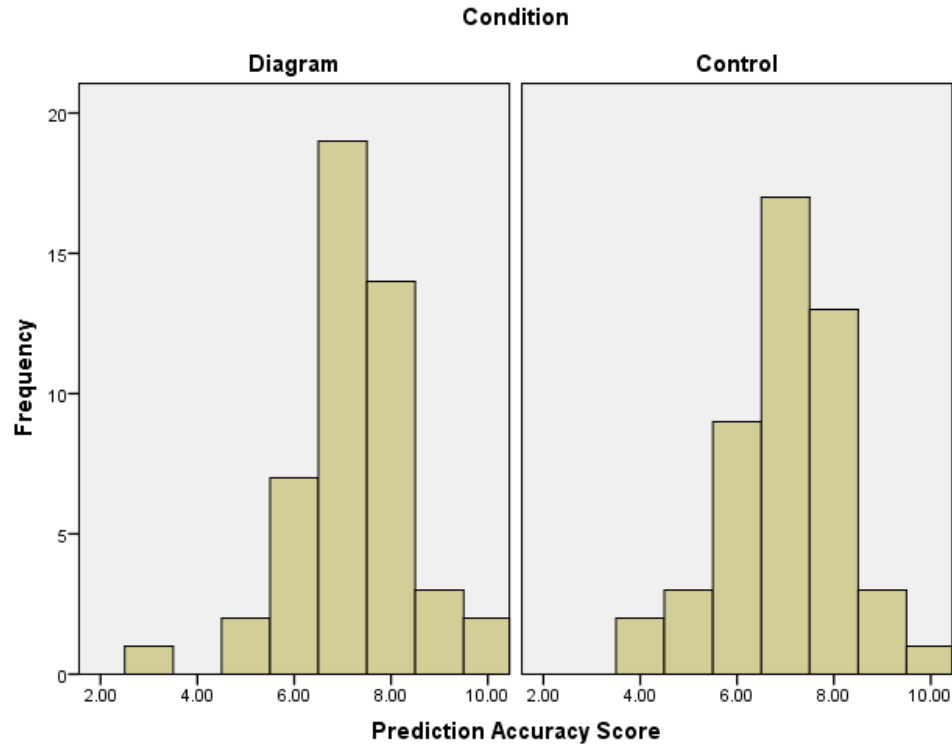
Figure 3.7. *Distribution of prediction accuracy scores, by condition.*

Scores were converted to proportion correct. Overall, the mean prediction accuracy score (mean proportion correct) was 0.71 ($SD = 0.12$). One-sample t-tests (2-tailed) against chance found that children's predictions were above chance levels for all three age groups. The mean for the three-year-olds was 0.67 ($SD = 0.16$), $t(31)= 6.19$, $p < .001$. The mean for the four-year-olds was 0.72 ($SD = 0.10$), $t(31) = 12.46$, $p < .001$. Finally, the mean for the five-year-olds was 0.74 ($SD = 0.98$), $t(31) = 14.05$, $p < .001$. See Figure 3.8 below.

Accuracy of Children's Predictions
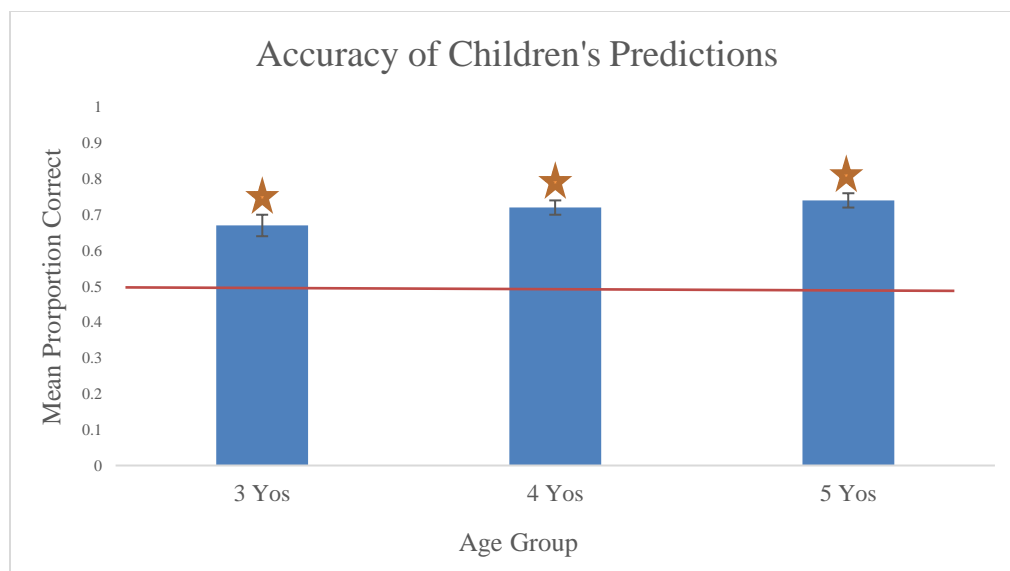
Figure 3.8. *Accuracy of children's predictions (mean proportion correct, by age group).* Note: Error bars on all graphs represent one standard error of the mean[2].

Analyses also examined the effects of condition, age and gender on children's prediction accuracy scores. A Mann-Whitney U test (2-tailed)[1] indicated that scores did not differ between the diagram condition and the control condition, $U = 1046.50$, $p = .42$. Neither was there a difference between boys and girls, $U = 1139.00$, $p = .92$ (2-tailed). Finally, a Kruskal Wallis H test revealed that there was no significant difference in prediction accuracy scores between the three age groups, $\chi^2(2) = 4.92$, $p = 0.09$. (Note: A significant effect for condition was not expected, as use of the inclined plane diagrams was not intended to support children in making correct predictions, but rather to aid children in keeping predictions and outcomes in mind, as well as comparing them). See Table 3.2 below for descriptive statistics.

Table 3.2. *Descriptive Statistics for Prediction Accuracy Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample (*n* = 96)	0.71 (0.12)
3 YOs (*n* = 32)	0.67 (0.16)
4 YOs (*n* = 32)	0.72 (0.10)
5 YOs (*n* = 32)	0.74 (0.10)
Diagram Cond. (*n* = 48)	0.72 (0.12)
Control Cond. (*n* = 48)	0.70 (0.12)
Boys (*n* = 47)	0.72 (0.13)
Girls (*n* = 40)	0.71 (0.12)

In summary, although there were no age, condition, or gender effects, children from all three age groups performed above chance levels in making accurate predictions in this context. However, although scores were above chance levels, they were not at ceiling.

3.6b Individual Item Analysis. Analyses also examined prediction scores for each trial individually to assess whether children's predictions were correct above chance levels for each pair, as well as whether there were any item effects (i.e., if some items were more difficult for children than others).

One sample t-tests (2-tailed) against chance were conducted for each pair. (See Figure 3.2 above for pair numbers and test items contained in each pair). Children performed better than chance (0.50) on pair 1 ($t(95) = 4.74, p < .001, d = .97$), pair 2 ($t(95) = 8.72, p < .001, d = 1.79$), pair 3 ($t(95) = 12.63, p < .001, d = 2.59$), pair 5 ($t(95) = 7.80, p < .001, d = 1.60$), pair 6 ($t(95) = 8.72, p < .001, d = 1.79$), pair 8 ($t(95) = 9.78, p < .001, d = 2.01$), and pair 9 ($t(95) = 7.80, p < .001, d = 1.60$). For each of these analyses, the effect size was found to exceed Cohen's (1988) convention for a large effect ($d = .80$).

Children performed at chance levels (0.50) on pair 4 ($t(95) = -1.02$, $p = .31$), pair 7 ($t(95) = -0.61$, $p = .54$), and pair 10 ($t(95) = -0.82$, $p = 0.42$). Cohen's (1988) effect size values suggested low practical significance ($d = -.21$, $-.13$, and $-.17$, respectively). See Table 3.3 for overall means and standard deviations for each pair, as well as by age group. (Note: The one-sample t-tests against chance were also carried out for each age group separately. The results were the same for each age group [i.e., all age groups were above chance on the same pair numbers], so results are reported for the full sample).

Table 3.3. *Mean and SD for Children's Performance on each Pair for all Age Groups.*

	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7	Pair 8	Pair 9	Pair 10
3s	0.75(0.44)	0.69(0.47)	0.84(0.37)	0.56(0.50)	0.69(0.47)	0.69(0.47)	0.53(0.51)	0.75(0.44)	0.72(0.46)	0.50(0.51)
4s	0.69(0.47)	0.84(0.37)	0.88(0.34)	0.47(0.51)	0.84(0.37)	0.84(0.37)	0.47(0.51)	0.94(0.25)	0.84(0.37)	0.41(0.50)
5s	0.72(0.46)	0.97(0.18)	0.97(0.18)	0.31(0.47)	0.91(0.30)	0.97(0.18)	0.41(0.50)	0.88(0.34)	0.88(0.34)	0.47(0.51)
Tot.	0.71(0.45)	0.83(0.37)	0.90(0.31)	0.45(0.50)	0.81(0.39)	0.83(0.37)	0.47(0.50)	0.85(0.35)	0.81(0.39)	0.46(0.50)

3.6c Recalling Initial Predictions. Analyses also explored children's ability to recall their initial predictions after testing items on the scale. Each child was assigned a prediction recall score (out of a possible 10 points), based on the total number of predictions accurately recalled, regardless of whether those predictions were confirmed or disconfirmed by testing the items on the scale. See distributions of prediction recall scores in Figures 3.9 through 3.11 below.
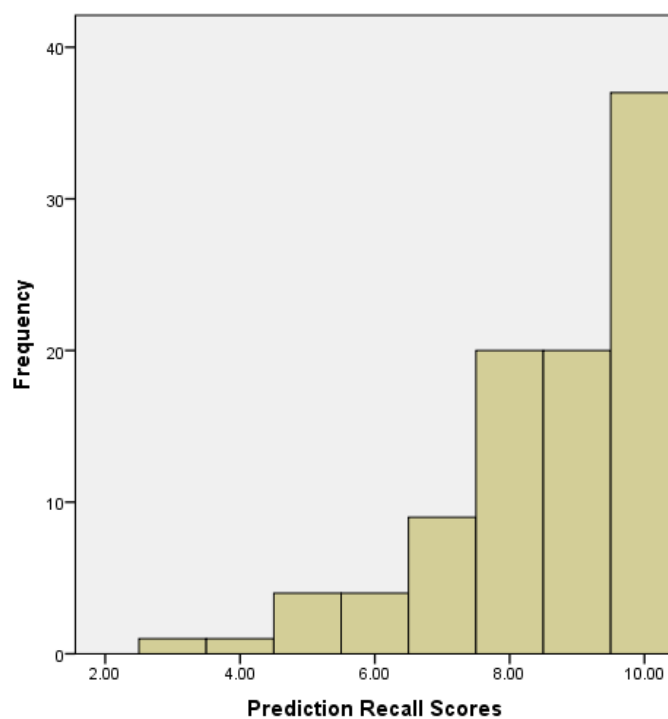
Figure 3.9. *Distribution of prediction recall scores for the full sample (n = 96).*
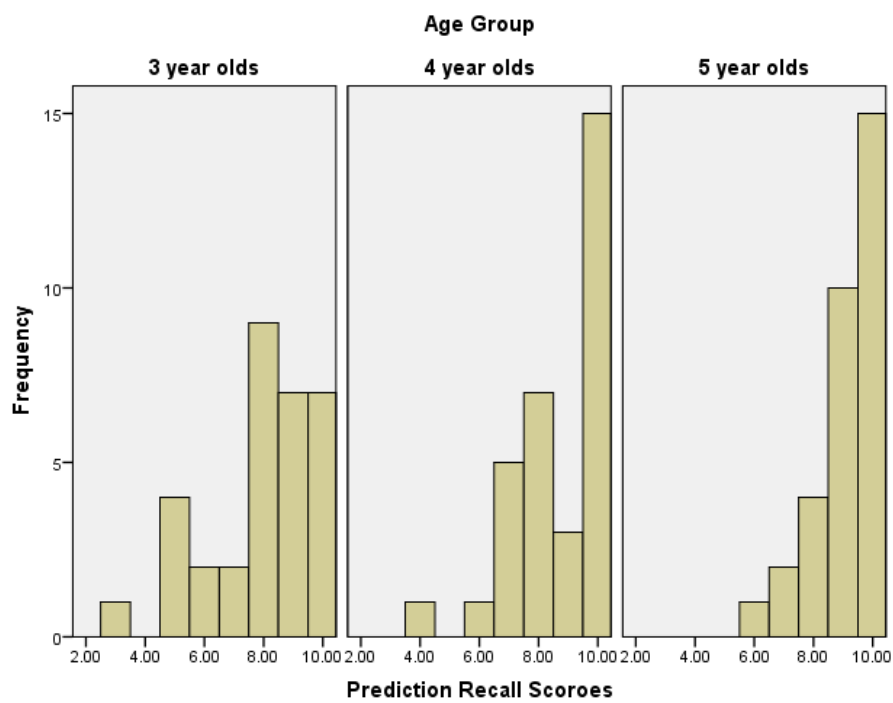
Figure 3.10. *Distribution of prediction recall scores, by age group.*
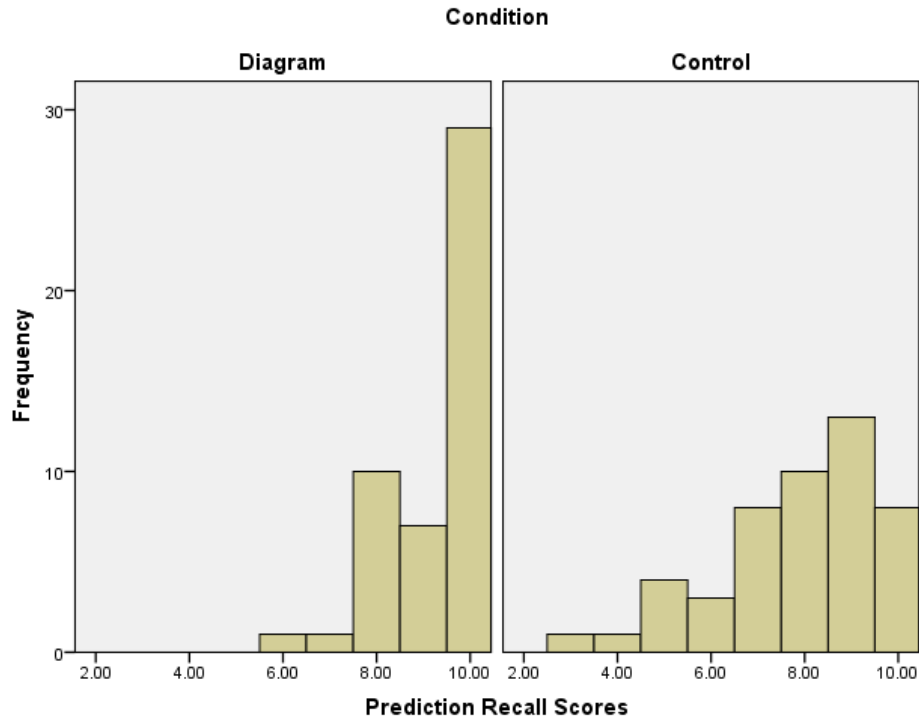
Condition

Figure 3.11. *Distribution of prediction recall scores, by condition.*

Scores were then converted to proportion correct. Overall, the mean prediction recall score (mean proportion correct) was 0.86 ($SD = 0.16$). One-sample t-tests (2-tailed) against chance found that, for all three age groups, children's prediction recall scores were reliably above chance levels. The mean was 0.79 ($SD = 0.18$) for the three-year-olds, 0.87 ($SD = 0.15$) for the four-year-olds, and 0.91 ($SD = 0.11$) for the five-year-olds, ($t(31) = 8.99$, $p < .001$; $t(31) = 13.61$, $p < .001$; $t(31) = 21.81$, $p < .001$, respectively). Further, the prediction recall scores for 77 children (80%) were above chance levels according to the binomial criterion of correctly recalling at least 8 out of 10 predictions during the activity (binomial sign test, $p = 0.0547$). By age, 72% of three-year-olds, 78% of four-year-olds, and 91% of five-year-olds passed the task under this

criterion. Taken together, these results indicate that from a young age, children in the study were skilled in accurately recalling their predictions after testing items on the scale.

Analyses also examined the effects of condition, age and gender on children's prediction recall scores. A Mann-Whitney U test (2-tailed) found a significant difference between the diagram condition (*Mdn* = 1.00, mean rank = 60.98) and the control condition (*Mdn* = 0.80, mean rank = 36.02), *U* = 553.00, *p* < .001. Further, a Kruskal Wallis H test revealed a significant difference in scores among the three age groups, $\chi^2(2)$ = 8.005, *p* = 0.02. The median prediction recall score was 0.80 (mean rank = 38.25) for the three-year-olds, 0.90 (mean rank = 50.33) for the four-year-olds, and 0.90 (mean rank = 56.92) for the five-year-olds. Post-hoc comparisons with Bonferroni corrections found a significant difference between the three- and five-year-olds (*p* = .016). There was no difference between the three- and four-year-olds (*p* = .21), nor between the four- and five-year-olds (*p* = .97). Finally, a Mann-Whitney U test (2-tailed) found no gender differences, *U* = 1066.50, *p* = .52. See Table 3.4 below for descriptive statistics. Also see age and condition effects in Figure 3.12 below.

Table 3.4. *Descriptive Statistics for Prediction Recall Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample (*n* = 96)	0.86 (0.16)
3 YOs (*n* = 32)	0.79 (0.18)
4 YOs (*n* = 32)	0.87 (0.15)
5 YOs (*n* = 32)	0.91 (0.11)
Diagram Cond. (*n* = 48)	0.93 (0.10)
Control Cond. (*n* = 48)	0.79 (0.17)
Boys (*n* = 47)	0.85 (0.16)
Girls (*n* = 49)	0.87 (0.15)
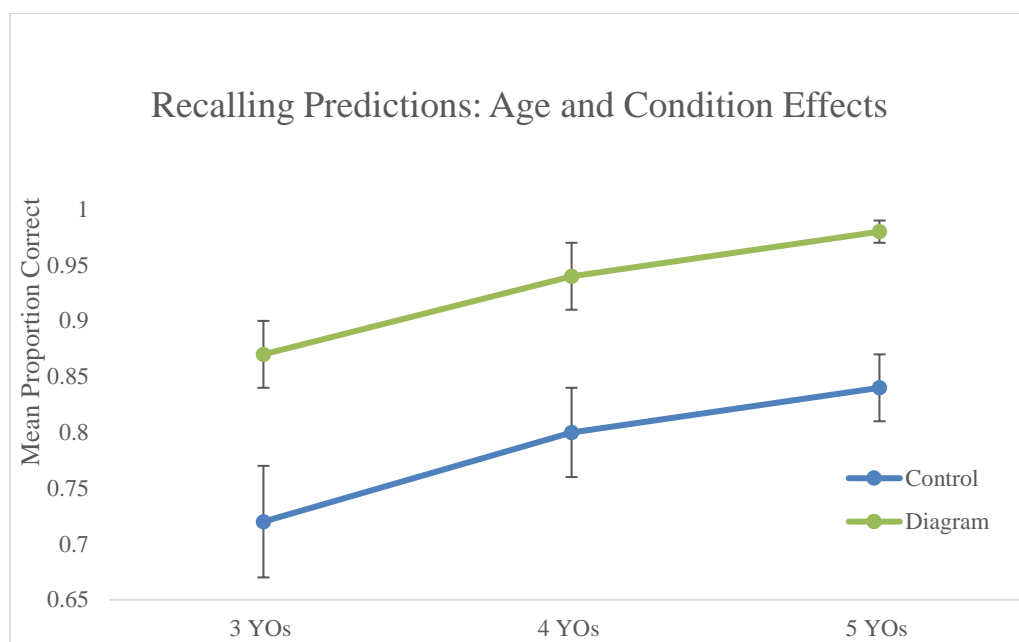
Figure 3.12. *Condition and age effects on children's mean prediction recall scores.*

In summary, from a young age, children were skilled in accurately recalling their predictions after testing items on the balance scale. This skill also increased over the preschool years, as the five-year-olds significantly outperformed the three-year-olds. Importantly, use of the balance scale diagrams reliably supported children's ability to accurately recall predictions after testing items on the balance scale. The effect was salient for all three age groups.

3.6d Accuracy of Children's Judgments. Analyses also examined children's ability to judge whether or not their predictions were confirmed by testing items on the scale. As the balance scale activity involved 10 trials of predicting and checking, children received 1 point for each correct judgment made regarding the accuracy of their predictions. In other words, 1 point was awarded each time a child agreed that his/her

correct prediction was in fact right, as well as each time a child agreed that his/her

incorrect prediction was indeed wrong. Thus, children could receive a judgment score

between 0 and 10 points. See distributions of judgment scores in Figures 3.13 through

3.15 below.
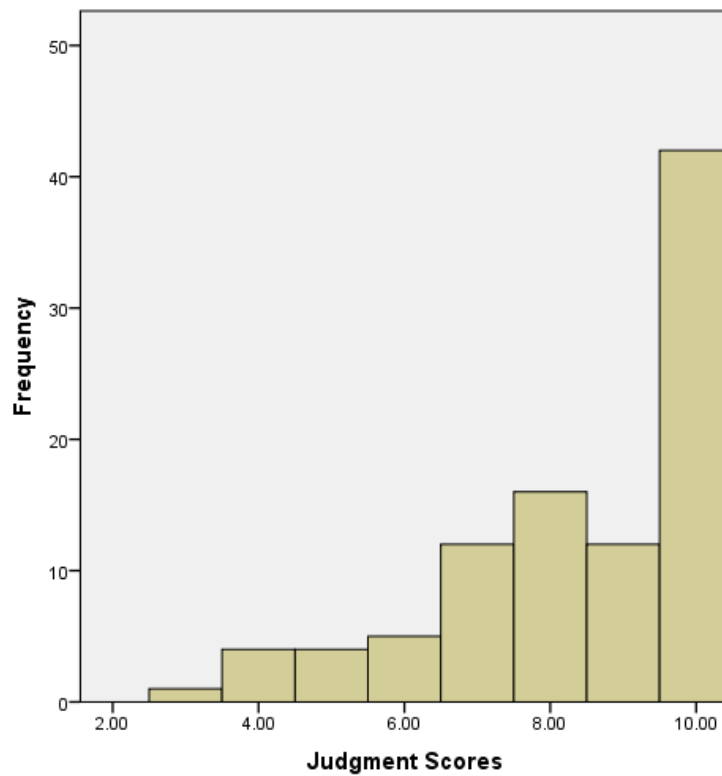
Figure 3.13. *Distribution of judgment scores for the full sample (n = 96).*
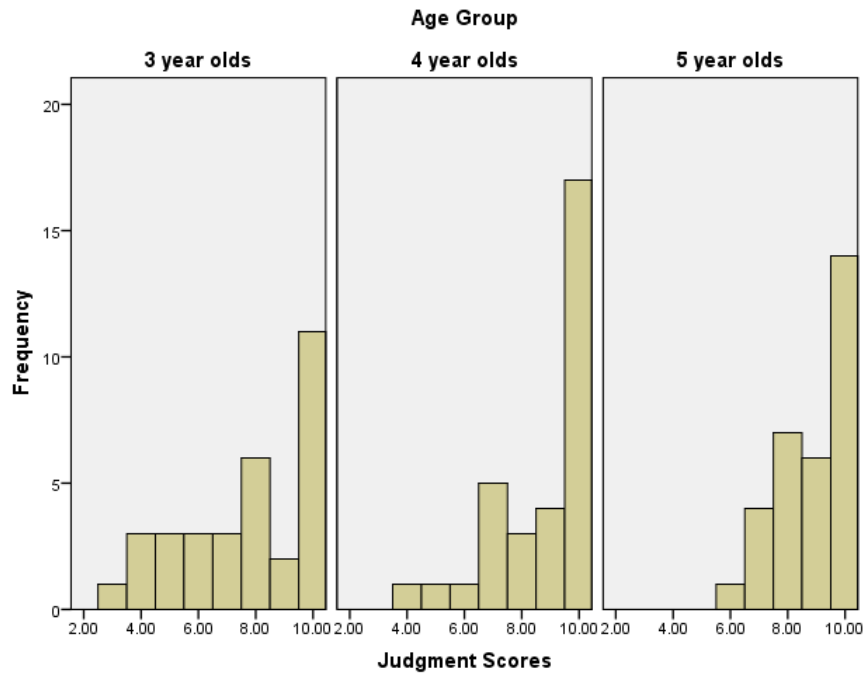
Figure 3.14. *Distribution of judgment scores, by age group.*
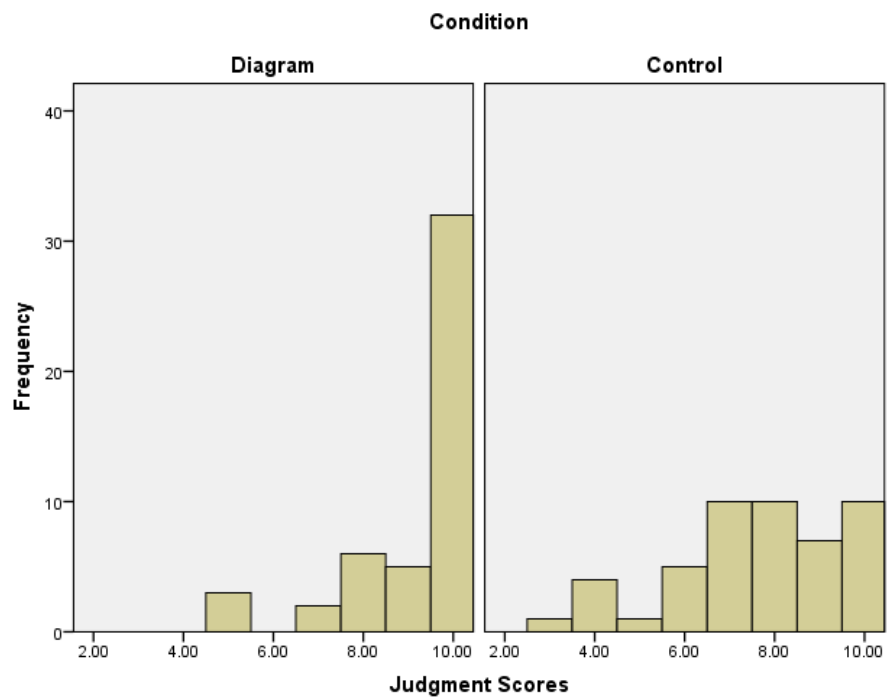
Figure 3.15. *Distribution of judgment scores, by condition.*

Scores were then converted to proportion correct. Overall, the mean judgment score (mean proportion correct) was 0.84 (SD = 0.18). One-sample t-tests (2-tailed) against chance indicated that, for all three age groups, children's judgment scores were reliably above chance levels. The mean was 0.77 (SD = 0.23) for the three-year-olds, 0.88 (SD = 0.17) for the four-year-olds, and 0.89 (SD = 0.12) for the five-year-olds ($t(31)$ = 6.67, $p < .001$; $t(31)$ = 12.59, $p < .001$; and $t(31)$ = 18.09, $p < .001$, respectively). Further, 70 children (73%) had passing judgment scores according to the binomial criterion of making 8 or more correct judgments out of 10 (binomial sign test, p = 0.0547). By age, 59% of three-year-olds, 75% of four-year-olds, and 84% of five- year-olds performed better than chance according to this criterion. This indicates that from a young age, the children in the study were skilled in judging the accuracy of their predictions about the relative weights of objects using the balance scale.

Analyses also examined the effects of condition, age and gender on children's judgment scores. A Mann-Whitney U test (2-tailed) found a significant difference between the diagram condition (Mdn = 1.00, mean rank = 61.02) and the control condition (Mdn = 0.80, mean rank = 35.98), U = 551.00, $p < .001$. Although there was a trend for judgment scores to increase with age, a Kruskal Wallis H test exploring differences in scores among the three age groups reached only marginal significance, $\chi^2(2)$ = 5.46, p = 0.065. Finally, no gender differences were found, U = 1046.50, p = .42 (2-tailed). See Table 3.5 below for descriptive statistics. Also see age and condition effects in Figure 3.16 below.

Table 3.5. *Descriptive Statistics for Children's Judgment Scores, by Group.*

Group	Mean Proportion Correct (SD)
Full Sample (*n* = 96)	0.84 (0.18)
3 YOs (*n* = 32)	0.77 (0.23)
4 YOs (*n* = 32)	0.88 (0.17)
5 YOs (*n* = 32)	0.89 (0.12)
Diagram Cond. (*n* = 48)	0.92 (0.14)
Control Cond. (*n* = 48)	0.76 (0.19)
Boys (*n* = 47)	0.83 (0.18)
Girls (*n* = 49)	0.85 (0.18)
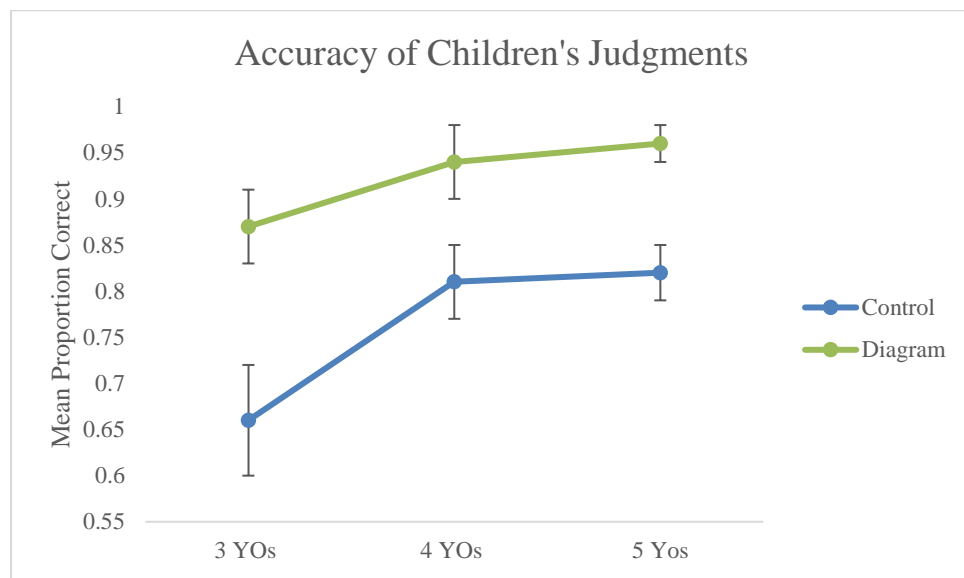
Figure 3.16. *Condition and age effects on children's mean judgment scores.*

In summary, preschool children in this study were able to judge the accuracy of their predictions about the relative weights of items using the balance scale. Importantly, use of the inclined plane diagrams again reliably supported children's ability to judge the correctness of initial predictions after testing items on the scale. The effect was particularly salient among the youngest participants.

Analyses also explored whether the accuracy of children's judgments related to the correctness of their predictions. In other words, analyses examined how well children

judged correct versus incorrect predictions (i.e., whether children were more skilled in judging one versus the other). Considering trials with correct initial predictions, and those with incorrect initial predictions as separate groups, the total number of times those predictions were accurately judged was tallied, as was the total number of times those predictions were judged inaccurately. The proportion of correct and incorrect predictions judged accurately was then calculated.

Table 3.6 below illustrates the proportions of children's predictions that were judged accurately, based on whether or not those predictions were confirmed by testing items on the scale. More specifically, the table includes the proportion of trials in which children both made correct initial predictions, as well as accurately judged those predictions (i.e., agreed that they were in fact right), out of the total number of trials in which children made correct initial predictions. The table also includes the proportion of trials in which children both made incorrect initial predictions, as well as accurately judged those predictions (i.e., agreed that those predictions were indeed wrong), out of the total number of trials in which children made incorrect initial predictions.

Table 3.6. *Proportion of Correct Judgments for both Correct and Incorrect Initial Predictions.*

	Diagram		Control	
	Correct P	**Incorrect P**	**Correct P**	**Incorrect P**
3 YOS	0.89(0.19)	0.79(0.29)	0.62(0.37)	0.59(0.48)
4 YOs	0.96(0.14)	0.87(0.30)	0.87(0.23)	0.65(0.31)
5 YOs	0.98(0.08)	0.81(0.36)	0.94(0.10)	0.50(0.35)
Full Sample	0.95(0.15)	0.82(0.31)	0.81(0.29)	0.58(0.38)

First, in looking only at those trials in which children made correct initial predictions, one sample t-tests against chance (2-tailed) found that, for all three age groups, when children's predictions were correct, children were also above chance levels (M = 0.76, 0.91, and 0.96, respectively) in judging the accuracy of those predictions, $t(31) = 4.51, p < .001; t(31) = 11.87, p < .001;$ and $t(31) = 28.42, p < .001$.

Analyses also examined the influence of condition, age and gender on the proportion of correct predictions judged accurately. A Mann-Whitney U test (2-tailed) found a significant difference between the diagram condition (Mdn = 1.00, mean rank = 55.36) and the control condition (Mdn = 1.00, mean rank = 41.64), U = 822.50, p = .003. Further, a Kruskal Wallis H test revealed a significant difference in scores among the three age groups. The median was 0.94 (mean rank = 37.84) for the three-year-olds, 1.00 (mean rank = 52.55) for the four-year-olds, and 1.00 (mean rank = 55.11) for the five-year-olds, $\chi^2(2) = 10.848, p = 0.004$. Post hoc comparisons with Bonferroni corrections found a significant difference between the three- and four-year-olds (p = .028), and between the three- and five-year-olds (p = .007). There was no difference between the four- and five-year-olds (p = 1.00). Finally, no gender differences were found, U = 1023.00, p = .25 (2-tailed). See condition and age effects in Figure 3.17 below.
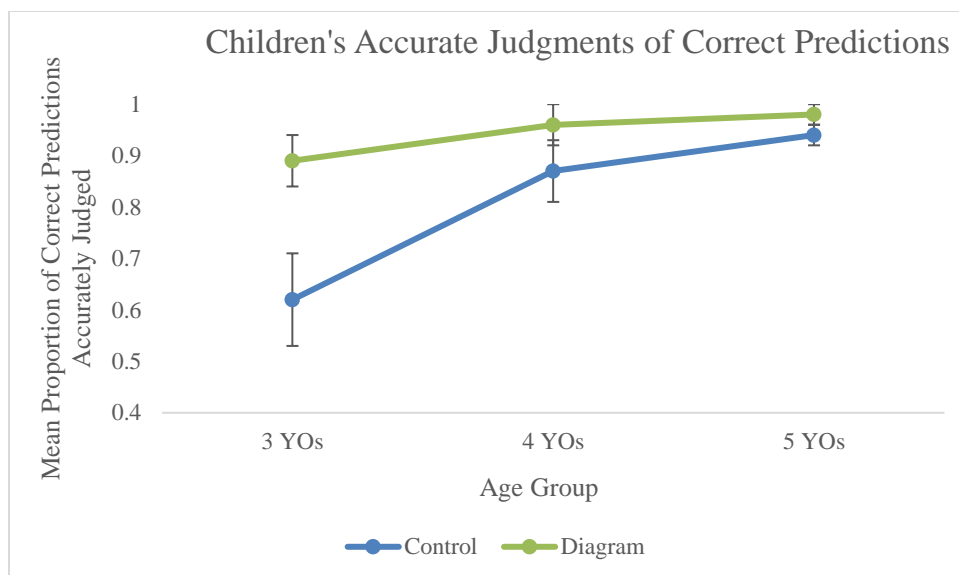
Figure 3.17. *Age and condition effects on children's judgments of their correct predictions.*

In sum, children were skilled in judging the accuracy of their correct predictions, with this skill increasing with age over the preschool years. Importantly, use of the balance scale diagrams again effectively supported children's ability to judge their correct predictions, as children in the diagram condition reliably outperformed the control. This effect was especially salient for the youngest age group.

Next, analyses examined the accuracy of children's judgments for only those trials in which children made incorrect initial predictions. (Note: three children, 1 three-year-old, 1 four-year-old, and 1 five-year-old, made only correct initial predictions throughout the task; thus, these 3 children were not included in the following analyses).

One sample t-tests against chance (2-tailed) found that, for all three age groups, when children's predictions were incorrect, children were above chance levels in judging the accuracy of those predictions (i.e., in agreeing that those predictions were in fact

wrong). The mean proportion of incorrect predictions accurately judged was 0.69 ($SD =$ 0.40) for the three-year-olds, 0.75 ($SD = 0.32$) for the four-year-olds, and 0.66 ($SD =$ 0.38) for the five-year-olds, ($t(30) = 2.72$, $p = .01$; $t(30) = 4.35$, $p < .001$; and $t(30) =$ 2.31, $p = .03$, respectively).

Analyses also examined the influence of condition, age and gender on the proportion of incorrect predictions judged accurately. A Mann-Whitney U test (2-tailed) found a significant difference between the diagram condition ($Mdn = 1.00$, mean rank = 55.71) and the control condition ($Mdn = 0.50$, mean rank = 38.48), $U = 680.50$, $p = .001$. A Kruskal Wallis H test found no difference in scores among the three age groups, $\chi^2(2)$ = 0.848, $p = 0.66$. Finally, a Mann-Whitney U test (2-tailed) found no difference between the boys and girls, $U = 1009.00$, $p = .56$. See condition and age effects in Figure 3.18 below.
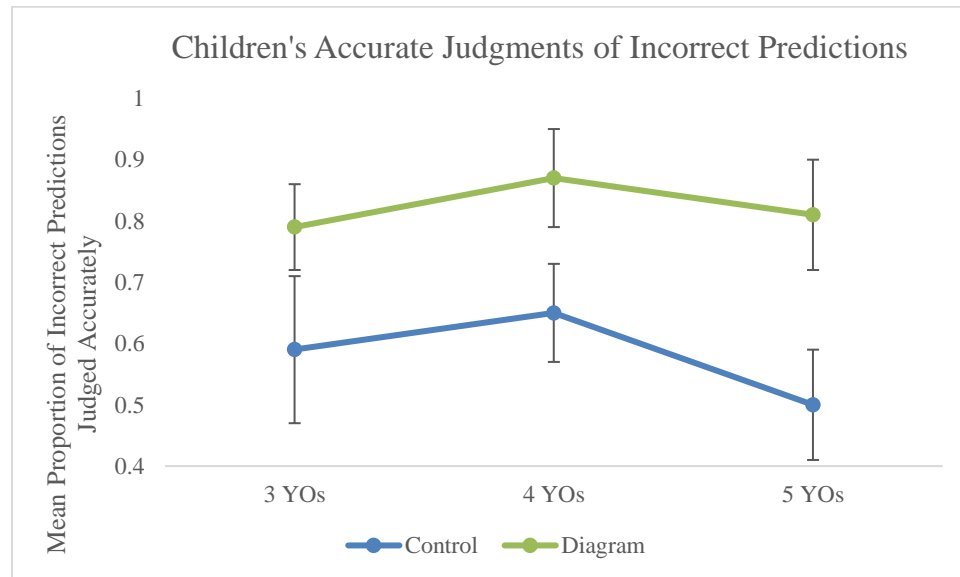
Figure 3.18. *Age and condition effects on children's judgments of their incorrect predictions.*

In summary, children from all three age groups were above chance levels (though not at ceiling) in accurately judging the accuracy of their incorrect predictions (i.e., or agreeing that those predictions were in fact wrong). Importantly, use of the balance scale diagrams again supported children's ability to judge their incorrect predictions, as children in the diagram condition reliably outperformed the control.

Finally, analyses also explored whether children from each age group showed a difference in their ability to judge the accuracy of their correct versus incorrect initial predictions. Among the three-year-olds, the mean proportion of correct predictions accurately judged was 0.75 ($SD = 0.33$), while the mean proportion of incorrect predictions accurately judged was 0.69 ($SD = 0.40$). A Wilcoxon Signed Ranks test (2-tailed) found no difference between the two groups, $Z = 0.68$, $p = .50$. For the four-year-olds, the mean proportion of correct predictions accurately judged was 0.91 ($SD = 0.20$), compared to a mean proportion of incorrect predictions accurately judged of 0.75 ($SD = 0.32$). A Wilcoxon Signed Ranks test (2-tailed) indicated that the four-year-olds reliably judged correct predictions more accurately than they did incorrect predictions. The mean rank for judging correct predictions was 7.00 ($Mdn = 1.00$), and the mean rank for judging incorrect predictions was 8.25 ($Mdn = 1.00$), $Z = 2.22$, $p = .027$. Finally, for the five-year-olds, the mean proportion of correct predictions accurately judged was 0.96 ($SD = 0.09$), while the mean proportion of incorrect predictions accurately judged was 0.66 ($SD = 0.38$). A Wilcoxon Signed Ranks test (2-tailed) indicated that, like the four-year-olds, five-year-olds were reliably better at judging correct predictions compared to incorrect ones. The mean rank for judging correct predictions was 4.00 ($Mdn = 1.00$),

and the mean rank for judging incorrect predictions was 9.31 (*Mdn* = 0.67), Z = 3.45, p = .001. Thus, three-year-old children showed no difference in their ability to determine whether their correct and incorrect predictions were confirmed by testing items on the scale. However, both the four- and five-year-old children were reliably more skilled in judging the accuracy of their correct predictions, compared to those which were not confirmed by testing items on the scale.

3.6e Patterns in Responses. Analyses also explored patterns and/or consistencies in children's responses for the balance scale reasoning activity. Again, the activity asked children to: (1). predict whether various pairs of items would make the scale balance, (2). recall predictions after testing items on the scale, and (3). judge the accuracy of those predictions. Note, as it was possible for children to demonstrate certain patterns and/or consistencies in any of these areas, the response patterns described below are not mutually exclusive.

Twenty-eight children (29%) achieved perfect prediction recall and judgment scores on the activity. Of these, 5 were three-year-olds, 13 were four-year-olds, and 10 were five-year-olds. Further, 23 were in the diagram condition, whereas 5 were in the control. Another 7 children (7%) made only 1 error on the task. Of these, 2 were three-year-olds, 2 were four-year-olds, and 3 were five-year-olds. Further, 5 were in the diagram condition, and 2 were in the control. Taken together, this group of 35 children (36%) demonstrated a clear understanding of all areas of the task. Further, the majority of these children (80%) were in the diagram condition, suggesting that perhaps use of the balance scale diagrams fostered their high performance on the activity.

When asked to recall their predictions after testing items on the scale, 20 children (21%) always stated the true observed outcome rather than their initial prediction. For example, in cases in which these children predicted that the scale would balance and then observed that one side went down, the children always stated that they correctly predicted that one side of the scale would go down. Instead of providing their initial prediction, these children always stated what actually happened. Of these 20 children, 8 were three-year-olds, 8 were four-year-olds, and 4 were five-year-olds; further, 5 were in the diagram condition whereas 15 were in the control. Thus, the majority of children of children showing this pattern were in the control condition.

When asked to state the outcome of testing sets of items on the scale, another 2 children (both five years old and in the diagram condition) always stated that their prediction occurred, rather than stating the true outcome. For example, in cases in which these children predicted that the scale would balance, and then tested the items on the scale and observed that the scale did not balance, these children stated that the scale did in fact balance.

Further, 15 children (16%) showed response biases when asked to judge the accuracy of their predictions, either by always saying that their predictions were correct ($n = 12$), or always stating that their predictions were incorrect ($n = 3$). Of these, 5 were three-year-olds, 5 were four- year-olds, and 5 were five-year-olds. Further, only 3 were in the diagram condition, whereas 12 were in the control condition. Another 6 children (6%) appeared not to understand the question regarding the judgement of their predictions, as they randomly responded that predictions were confirmed or needed to be revised. Of these, 4 were three years old, and 2 were four years old. Further, 2 were in

the diagram condition, whereas 4 were in the control. The remaining 32 children's (33%)

errors revealed no systematic patterns or biases. See Table 3.7 below for a summary.

Table 3.7. *Patterns and Biases in Children's Responses for the Balance Scale Activity.*

Response Pattern	Number of Cases (% of Total)
All correct prediction recall and judgments	28 (29%)
Only 1 error, either in recalling or judging predictions	7 (7%)
Recalling Predictions: Always stated true outcome	20 (21%)
Judging Predictions: Always stated that predictions were right, or wrong	15 (16%)
Stating Outcome: Always stated that prediction occurred	2 (2%)
Random errors; no systematic patterns or biases	32 (33%)

Relationships between Pre-test Scores and Science Scores.

3.6f Pre-test Scores. Thirty-six children (38%) passed the theory of mind pre-test

under the strict criterion of scoring 3 out of a possible of 3 points, whereas 60 (63%) did

not pass. Of those who passed the pre-test, 3 were three years old, 12 were four years

old, and 21 were five years old. Further, 45 children (47%) passed the DCCS pre-test

under the strict criterion of scoring 4 out of a possible of 4 points, whereas 51 (53%) not

pass. Of those who passed the pre-test, 11 were three years old, 12 were four years old,

and 22 were five years old. Mann-Whitney U tests (2-tailed) found no difference

between the diagram and control conditions in terms of the number of children in each

group who passed the pre-tests; for the DCCS pre-test, $U = 936.00$, $p = .07$, and for the

theory of mind pre-test, $U = 1152.00$, $p = 1.00$.

3.6g Correlations between Pre-test Scores and Science Scores. Spearman's rho

correlations were used to explore the relationships between children's passing status on

the pre-tests, and their prediction recall and judgment scores. Results found that children

who passed the theory of mind pre-test tended to have higher prediction recall scores $r_s =$

.34, p = .001, as well as higher judgment scores r_s = .31, p = .002. Similarly, children

who passed the DCCS pre-test also tended to have higher prediction recall scores r_s =

.31, p = .002, as well as higher judgment scores DCCS pre-test r_s = .31, p = .002. Thus,

children who passed the theory of mind and DCCS pre-tests tended to more accurately

recall, as well as judge the accuracy of, predictions compared to those children who did

not pass the pre-tests. See Table 3.8 below for descriptive statistics.

Table 3.8. *Descriptive Statistics for Science Scores, based on Passing Status on the Pre-tests.*

Group	Mean Prediction Recall Score (SD)	Mean Judgment Score (SD)
Passed ToM (n = 36)	0.93 (0.09)	0.92 (0.11)
Failed ToM (n = 60)	0.82 (0.17)	0.80 (0.20)
Passed DCCS (n = 45)	0.91 (0.12)	0.90 (0.16)
Failed DCCS (n = 51)	0.81 (0.18)	0.79 (0.19)

3.6 h Effect of Diagrams on Children's Prediction Recall Scores, Based on their

Passing Status on the Pre-tests. Analyses also explored whether use of the balance scale

diagrams influenced children's *prediction recall scores* differently depending on

children's passing status on the pre-tests. For the theory of mind pre-test, the mean

prediction recall score for those children who passed the pre-test was 0.98 (SD = 0.05)

for those in the diagram condition (n = 18), compared to 0.88 (SD = 0.10) for those in the

control condition (n = 18). A Mann-Whitney U test (2-tailed) found a significant

difference between the two groups. The diagram condition (Mdn = 1.00, mean rank =

23.78) reliably outperformed the control (Mdn = 0.90, mean rank = 13.22), U = 67.00, p

= .001. Among children who did not pass the theory of mind pre-test, the mean

prediction recall score for those in the diagram condition (n = 30) was 0.90 (SD = 0.11),

compared to a mean score of 0.73 (SD = 0.19) for those in the control (n = 30). A Mann-

Whitney U test (2-tailed) again found a significant difference between the groups. The diagram condition (*Mdn* = 0.90, mean rank = 38.62) again reliably outperformed the control (*Mdn* = 0.75, mean rank = 22.38), *U* = 206.50, *p* < .001. Thus, use of the balance scale diagrams was associated with reliably higher prediction recall scores, independent of children's passing status on the theory of mind pre-test. The effect was especially salient among children who failed the pre-test. See Figure 3.19 below.
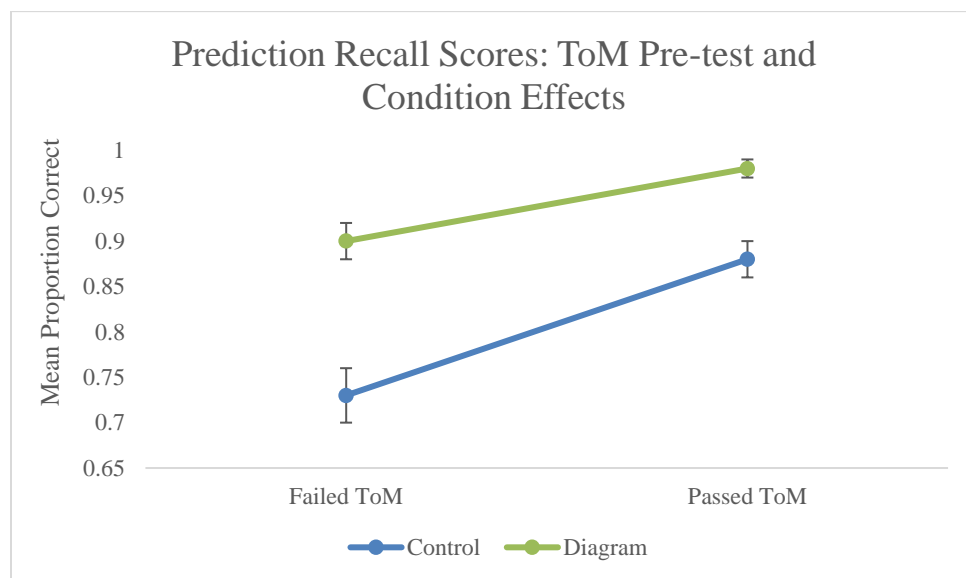
Figure 3.19. *Condition and ToM pre-test effects on prediction recall scores.*

Analyses also explored whether use of the balance scale diagrams influenced children's prediction recall scores differently depending on children's passing status on the Dimensional Change Card Sort pre-test. Among the children who passed the DCCS pre-test, the mean prediction recall score for those in the diagram condition (*n* = 27) was 0.95 (*SD* = 0.10) compared to a mean of 0.85 (*SD* = 0.12) for those in the control condition (*n* =18). A Mann-Whitney U test (2-tailed) confirmed a significant difference between the two groups, with the diagram condition (*Mdn* = 1.00, mean rank = 28.07)

outperforming the control (*Mdn* = 0.90, mean rank = 15.39), *U* = 106.00, *p* = .001.

Among those who did not pass the DCCS pre-test, the mean prediction recall score for

those in the diagram condition (*n* = 21) was 0.90 (*SD* = 0.10), whereas the mean for the

control (*n* = 30) was 0.75 (*SD* = 0.19). A Mann-Whitney U test (2-tailed) again found a

significant difference between the two groups, with the diagram condition (*Mdn* = 0.90,

mean rank = 33.07) outperforming the control (*Mdn* = 0.80, mean rank = 21.05), *U* =

166.50, *p* = .004. Thus, use of the balance scale diagrams was associated with higher

prediction recall scores among children who both passed or failed the DCCS pre-test,

compared to those children who did not have access to the diagrams. The effect was

again especially salient among children who failed the pre-test. This finding suggests

that use of the diagrams may benefit children, and especially those with less-well

developed executive functioning skills, in terms of supporting their ability to recall initial

predictions while engaging in scientific investigations. See Figure 3.20 below.
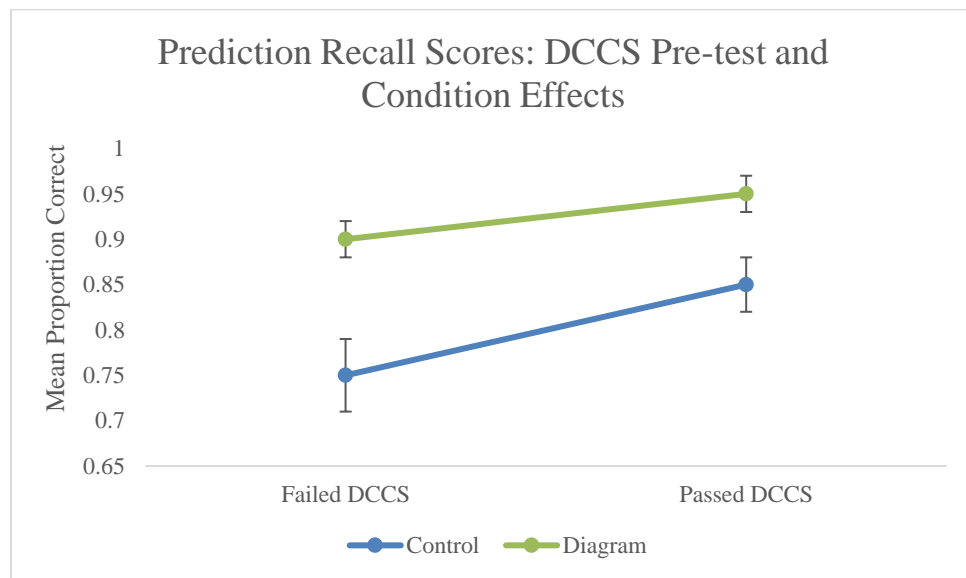
Figure 3.20. *Condition and DCCS pre-test effects on prediction recall scores.*

3.6i Effect of Diagrams on Children's Judgment Scores, Based on their Passing Status on the Pre-tests. Analyses also explored the effect of the balance diagrams on children's *judgment scores*, based on their passing status on the pre-tests. First, among children who passed the theory of mind pre-test, the mean judgment score for those in the diagram condition was 0.97 (SD = 0.06, n = 18), compared to a mean of 0.87 (SD = 0.12) for those in the control condition (n = 18). A Mann-Whitney U test (2-tailed) found a significant difference between the groups. The diagram condition (*Mdn* = 1.00, mean rank = 23.03), reliably outperformed the control (*Mdn* = 0.90, mean rank = 13.97), U = 80.50, p = .004. Among children who failed the theory of mind pre-test, the mean judgment score for children in the diagram condition (n = 30) was 0.89 (SD = 0.16) compared to a mean of 0.70 (SD = 0.19) for children in the control condition (n = 30). A Mann-Whitney U test (2-tailed) again found the difference between the groups to be significant, with the diagram condition (*Mdn* = 1.00, mean rank = 38.95) reliably outperforming the control (*Mdn* = 0.70, mean rank = 22.05), U = 196.50, p < .001. Thus, independent of children's passing status on the theory of mind pre-test, use of the balance scale diagrams during the science activity was associated with significantly higher judgment scores. The effect was especially salient for children who failed the pre-test. This indicates that the balance scale diagrams may serve as effective instructional tools for supporting young children's ability to judge the accuracy of their scientific predictions, especially among those with a less well-developed theory of mind. See Figure 3.21 below.
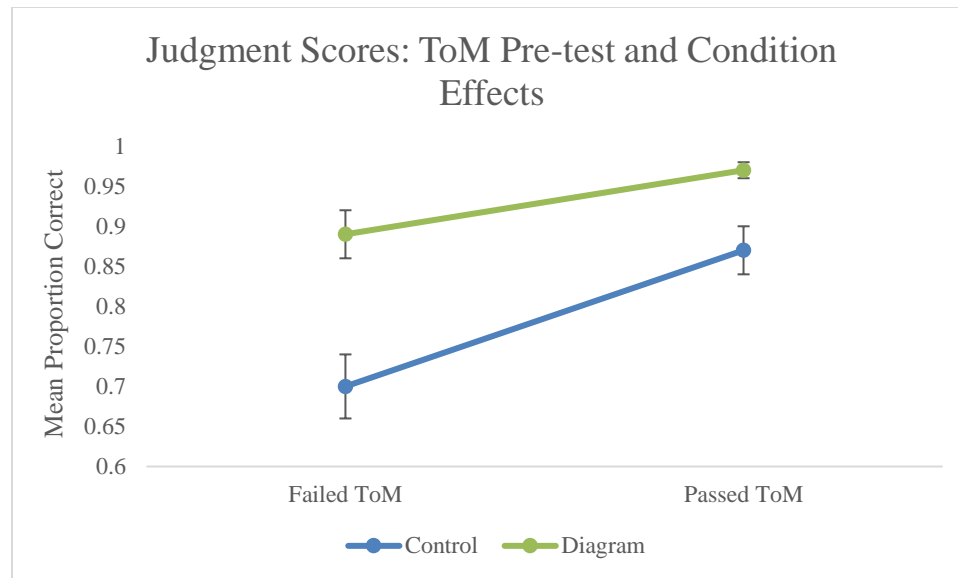
Figure 3.21. *Condition and ToM pre-test effects on judgment scores.*

Finally, analyses also examined the effect of use of the balance scale diagrams on children's judgment scores, based on their passing status on the DCCS pre-test. Among children who passed the DCCS pre-test, the mean judgment score for those in the diagram condition ($n = 27$) was 0.98 ($SD = 0.05$), compared to a mean of 0.78 ($SD = 0.19$) for those in the control condition ($n = 18$). A Mann-Whitney U test (2-tailed) found a significant difference between the two groups. Scores for the diagram condition (*Mdn* = 1.00, mean rank = 29.15) were significantly higher compared to those of the control condition (*Mdn* = 0.80, mean rank = 13.78), $U = 77.00$, $p < .001$. Among children who failed the DCCS pre-test, the mean judgment score was 0.85 ($SD = 0.18$) for those in the diagram condition ($n = 21$), compared to a mean of 0.75 ($SD = 0.19$) for those in the control condition ($n = 30$). A Mann-Whitney U test (2-tailed) exploring the difference between the two groups reached marginal significance. There was a strong tendency for scores of those in the diagram condition (*Mdn* = 0.90, mean rank = 30.55) to

be higher compared to those in the control condition (*Mdn* = 0.80, mean rank = 22.82), *U* = 219.50, *p* = .06. Thus, children who passed the DCCS pre-test benefitted from use of the balance scale diagrams in terms of judging the correctness of their initial predictions. However, the effect was less salient among children with weaker executive function skills. See Figure 3.22 below.
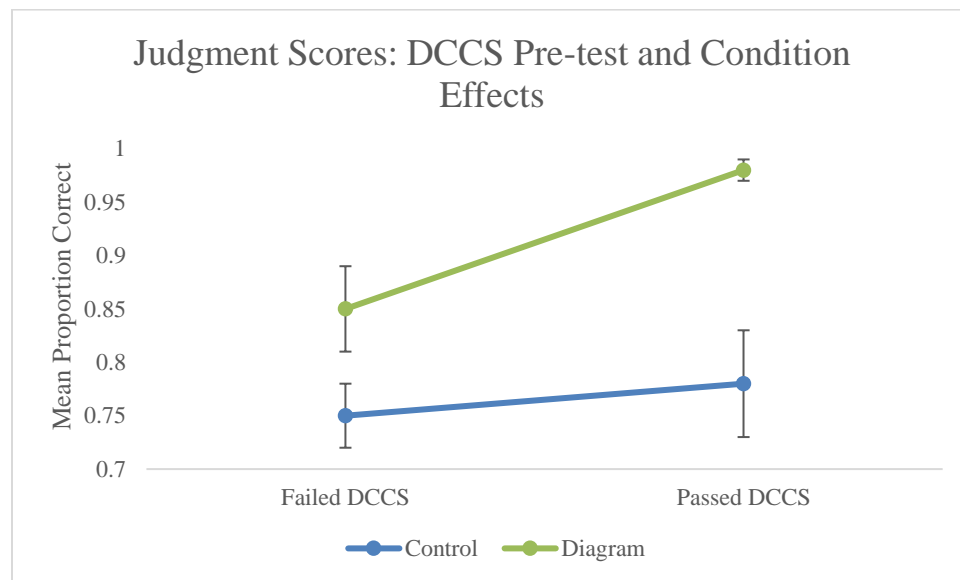
Figure 3.22. *Condition and DCCS pre-test effects on mean judgment scores.*

3.7 Conclusions and Implications for Early Childhood Education

This study aimed to determine whether findings from the previous study contained in this dissertation replicate in a different context of scientific investigation, and with different test items and materials. Findings did indeed replicate. Specifically, the research results presented here provide further evidence that preschool children, ages three to five years, can use the scientific practices of predicting and checking. When asked to predict about the relative weights of objects using a balance scale, even the three-year-olds' predictions were above chance levels. The children in the study also accurately recalled their initial predictions after testing items on the balance scale, with this skill increasing over the preschool years. Children were also capable of judging the correctness of their initial predictions. Importantly, children again benefitted from the option to use relevant physics diagrams during the science investigations. Finally, positive correlations were again found among children's developing executive function and theory of mind skills, and their scientific thinking scores.

These findings indicate that exploring the relative weights of objects using a balance scale is another suitable context for investigating the ability of preschool-aged children to engage in scientific investigations. The context also constitutes a developmentally-appropriate set of conversational options, objects and diagrams that can support science learning in the early childhood classroom. Since the materials are neither too easy nor too difficult, the teacher has room to help children actively construct understanding on what they already know something about.

Again, it is important that the findings regarding the effectiveness of the physics diagrams replicated in this new context for scientific investigation, and with a new set of

diagrams. Specifically, both children's prediction recall, and the accuracy of their judgments, again benefitted from use of the balance scale diagrams. This provides confirming evidence that, from a young age, children are perhaps capable of using and understanding forms of diagrammatic representation in physical science to support their predicting and checking during scientific investigations. As follows, the diagrams may be effective pedagogical tools for fostering young children's development of key scientific thinking skills from an early age.

Importantly, such diagrams are tools that teachers can easily incorporate into the kinds of activities often already happening in the early childhood classroom. This is significant, as little prior work has explored early forms of preschool literacy related to science learning, as well as children's ability to understand symbolic, external representations of predictions and outcomes for scientific investigations. It is worth pursuing whether repeated practice with such relevant diagrams during scientific investigations, in a variety of contexts, fosters young children's emerging scientific thinking skills. It is also worth exploring other means of supporting children's external representation of, and reflection upon their scientific investigations. For example, it is possible that children may similarly benefit from documenting their investigations in science journals (i.e., by drawing, and having adults help them write what they want to say) or by creating other models to represent of their thinking (i.e., Brenneman & Lauro, 2008; Gelman et al., 2009). Similarly, the Reggio Emilia approach demonstrates young children's ability to document, record and represent ideas using a variety of forms of documentation (i.e., Katz, 1998).

Finally, this study also finds further support for the hypothesis that children's capacity for early scientific thinking may relate to their developing theory of mind and executive functioning capacities, as children who passed the theory of mind and executive function pre-tests tended to score higher on scientific thinking measures. In other words, children's theory of mind and executive function pre-test scores positively correlated with their prediction recall and judgment scores. Further, use of the balance scale diagrams during the science activity benefitted children's prediction recall and judgment scores, independent of their passing status on the theory of mind and executive function pre-tests. However, the effects were particularly salient among the children with weaker theory of mind and executive function skills. This again indicates that the diagrams may serve as effective instructional scaffolds (i.e., Vygotsky, 1978) for supporting children's scientific predicting and checking as they are acquiring critical theory of mind and executive function skills associated with scientific thinking. It is worth continuing to explore the relationships among theory of mind, executive functioning and scientific thinking, as well as the use of relevant diagrams to support children's development of such skills in the early childhood classroom.

Although this work found positive correlations among preschoolers' early scientific thinking skills, and their developing theory of mind and executive functioning abilities, it is important to note that this work has not established causal relationships among these variables. It is still possible, for example, that a mediating variable(s), such as children's language development, is responsible for the correlations among these variables. For instance, understanding the kinds of questions and language involved in the pre-tests may have led to the differences in children's performance on the science

activity, rather than differences in their theory of mind and/or executive function capacities per se. Thus, more work is needed to clarify the nature of the relationships among young children's capacity for early forms of scientific thinking and their developing theory of mind and executive functioning skills. It is important for future research to continue exploring these ideas.

In summary, this work provides confirming evidence that from a young age, preschool children are indeed capable of appropriately predicting and checking, but in a different context for scientific investigation. This indicates that it is developmentally appropriate to ask children as young as three years old to consider the relative weights of various items using a balance scale. It is also appropriate to provide young children with opportunities to use basic forms of scientific thinking skills in the preschool classroom, such as through learning activities which actively engage children in making and checking predictions in scientific domains, especially those in which they have some relevant prior knowledge. Further, such developing skills may be scaffolded (i.e., Vygotsky, 1987) by offering use of simple physics diagrams which buttress children's developing theory of mind and executive function skills, specifically by supporting children's external representation of, and reflection upon, relevant components of their scientific investigations. Providing such opportunities in the early childhood classroom may foster young children's development of early forms of critical thinking and scientific reasoning skills, which may help to better prepare children for learning in the later school years.

Chapter 4: Study 3

Preschoolers' Collaboration on a Physics Problem Involving Objects' Motion on an

Inclined Plane

4.1 Introduction.

Peer learning is an educational approach that involves students working with peers to achieve educational goals; it is variously described as cooperative learning, collaboration, and peer tutoring (O'Donnell & King, 1999). The past 30 years have seen a major increase in the use of peer learning in schools. Peer learning approaches to education have grown in popularity as schools have moved away from transmission models of learning, those which emphasize knowledge transmission from teacher to student, and toward constructivist approaches to learning which stress children's discovery learning and highlight the role of social activity in knowledge acquisition (e.g., Gelman, Meck, Romo, Meck, & Fritz, 1995; Phillips, 1995). As such, peer learning has become a central technique for employing constructivist educational approaches. One key reason for putting students into groups is to provide students with opportunities to learn from one another (Webb & Farivar, 1999). Another reason arises from the responsibility that schools have in preparing students for life after school in the workplace and in communities. Skills related to the ability to work collaboratively with others are valued in the larger cultural context in which schools exist. Thus, providing students with opportunities to learn with peers is seen as a means of enhancing learning outcomes, as well as providing experiences necessary for preparing students for life after formal schooling in an increasingly diverse and multicultural world.

Much of the work on peer learning stems from the constructivist theories of Piaget and Vygotsky (Tudge, 1992). Both theories view peer learning as a means to enhance learning outcomes. Piaget (1959) maintained that cognitive growth depends on active interaction with, and manipulation of one's environment. Learning involves

refining existing cognitive systems as individuals reflect on and organize experiences to adapt to their environment, and as they apply their current cognitive systems in order to make meaning in new situations. Piaget suggested that peer interactions serve as important contexts for students to modify their existing cognitive systems, because as students reflect on the perspectives and reactions of peers, they may revise their current cognitive system to make new meanings (De Lisi & Golbeck, 1999). Thus, peer interaction may raise cognitive conflict by bringing to light discrepancies between children's own and others' knowledge, resulting in disequilibration (Fawcett & Garton, 2005). An array of research grounded in Piagetian constructivist theory indicates that working with a peer results in greater cognitive benefit than does working alone (e.g., Druyan, 2001; Light & Littleton, 1994; Slavin, 1992).

Vygotsky's theory, in contrast, views development as a process through which children acquire mastery over cultural tools through their interactions with more competent others in their environment. These more competent others help children to learn appropriate ways of using tools that are important to their cultural group. A resulting implication is that children's learning can be fostered by their interactions with more competent peers. Thus, researchers in the Vygotskian tradition maintain that cognitive development is most likely to occur when two participants of differing initial levels of competence work collaboratively on a task to arrive at shared understanding (i.e., Garton, 1992; Johnson & Johnson, 1994). Work grounded in this framework has supported the notion that cognitive development depends on active social interaction with a more competent partner (i.e., Garton, 1992; Tudge, Winterhoff, & Hogan, 1996).

Together, these constructivist perspectives indicate that shared problem solving activities with peers may provoke and extend children's development of knowledge. Also emerging from this work is the development of collaborative communities that foster children's learning. In such communities, children and adults take on "varying but coordinated responsibilities to foster children's learning" (i.e. Rogoff, 1990; Rogoff, Turkanis, & Bartlett, 2001). However, many questions remain unanswered in terms of the process by which children can influence one another's learning and development. For example, although children working collaboratively have been shown to perform at higher levels than children working alone (i.e., Underwood, Underwood, & Wood, 2000), peer collaboration does not always result in individual cognitive change (i.e., Tudge & Winterhoff, 1993). Further, the cognitive benefits of working with a peer seem to depend on a complex array of factors, such as the comparative ability level of partners (Garton & Pratt, 2001), age (Hogan & Tudge, 1999), and the specific task at hand (Phelps & Damon, 1989).

In many preschool classrooms, children's opportunities for self-generated knowledge construction and peer learning tend to occur in "play" activities, such as in the dramatic play area of the classroom, constructive play in the "block area," or outdoors in social and cooperative play on the playground. Young children are natural collaborators and problem solvers within the context of such play. However, collaboration and cooperation among young children need not be restricted to play outside the classroom; it is also critical for children to learn how to work with others and how to grow intellectually in a wide array of social and educational settings. With appropriate

structuring of the learning environment, such forms of collaboration and cooperation may also be extended to school curriculum content areas, such as science learning.

Social interaction has been defined as a key component of scientific practice and productive learning in general, and it currently plays a central role in K-12 science learning (Duschl, Schweingruber, & Shouse, 2007). Students have been shown to obtain social, cognitive and affective benefits when teachers successfully foster a classroom community of learners in which students aim to contribute to a communal understanding of scientific problems. For example, Gelman et al.'s (1995) science-into-ESL program was successful in actively engaging high school ESL students in their own learning about doing science; students completed science experiments in teams, shared data, and produced public reports together. Others have worked with younger children in elementary classrooms to promote learning through social interaction by attempting to establish similar classroom versions of scientific communities, with the goal of supporting students in working together to build knowledge and understanding (i.e., Brown, 1997; Brown & Campione, 1990, 1994; Scardamalia & Bereiter, 1994). Such classroom interactions are not easy to facilitate and often require professional development for teachers, as well as time to develop into shared classroom norms (Duschl, Schweingruber, & Shouse, 2007). Still, it is worth the effort since students can learn to participate successfully in learning communities as early as the elementary school years.

Although an array of research has explored school-aged children's learning of science with peers, little work has examined the role of peer interaction in science learning during the preschool years. As peer interaction and collaboration have been

identified as critical components of science learning and science education, it is important to gain a better understanding of young children's collaborative skills in the context of science learning activities within the early childhood classroom.

4.2 Research Goals.

1. Research has shown that even toddlers demonstrate collaborative skills within controlled laboratory settings (i.e., Tomasello & Hamann, 2012). Still, the sequence of emerging skills enabling the coordination of mental action between two children given common goals in more naturalistic settings is unclear. One goal for the current study was to examine the nature of preschool children's collaboration and social interactions as they worked in pairs to solve simple physics problems. Children were asked to work with a classmate as they explored problems focused on the motion of objects on an incline plane. The aim was to check our assumption that the task would be suitable for fostering both peer collaboration and scientific thinking in the early childhood classroom.

The choice of problem was based on recent work on science education that demonstrates individually competent interactions (e.g., NRC, 2007; Duschl, 2012. See further detail below). Specifically, children were asked to work together to construct an object that rolls down the inclined plane, and then an object that slides down (in counterbalanced order). Children were also asked how to increase the objects' speed down the incline. Finally, children were asked which object traversed the incline at a faster rate, the rolling or sliding object. Of interest was whether pairs of children are capable of working together to solve the problems by drawing on their implicit knowledge. It was hypothesized that children would in fact work together to solve the

problems. It was also hypothesized that children would demonstrate an early understanding of some relevant physics principles, described in further detail below.

Investigating objects' motion an inclined plane, or "the simple mechanics of solid bounded objects" (Michaels, Shouse, & Schweingruber, 2008, p. 38), exposes children to a number of concepts within the domain of physical science. Such concepts about motion are also routinely included in standards for early science education. For example, part of the Next Generation Science Standards kindergarten performance expectations for *Motion and Stability: Forces and Interactions* (NGSS Lead States, 2013) encourage teachers: (a) to have children investigate the outcomes of pushes and pulls on the motion of an object; and (b) analyze whether their solutions changed the speed and/or the direction of an object as they expected. Further, when children explore the motion of objects down an inclined plane, they are introduced to physical science concepts associated with the motion and position of objects. This is one way to demonstrate the crosscutting concept, *Structure and Function* (NRC, 2012). This type of scientific investigation also meets the criteria for an age-appropriate science learning activity as defined by Worth (2010). This is because it involves: (1). concepts which are important to science, (2). phenomena drawn from the environment which are available for direct exploration, (3). concepts which can be explored in-depth and from multiple perspectives over time, and (4). phenomena and concepts which are interesting and engaging to young children and their teachers.

Investigating the motion of objects traversing an inclined plane is a particularly interesting choice because it provides the opportunity to offer young children intuitions about motion and its laws, without any use of advanced language or mathematics.

However, it still allows for the introduction of new, relevant vocabulary. For example, such context exposes children to the physics principle that as the angle of the inclined plane is increased, the acceleration of the object also increases (i.e., as the angle increases, the component of force parallel to the incline increases while the component of force perpendicular to the incline decreases; the parallel component of the weight vector causes the acceleration, and so accelerations increase at greater angles of incline). Another physics principle relates to objects *sliding* down a (frictionless) ramp. In such cases, two forces act on the object: (1) its weight pulls down toward the center of the earth; and (2) the ramp exerts upward force perpendicular to the surface of the ramp (i.e., the "normal" force). The result is that mass cancels out of Newton's Second Law, meaning that any object, regardless of size or mass, will slide down the (again, frictionless) incline with the same acceleration, or at the same rate (and again, the rate is dependent upon the angle of the incline).

Finally, a number of physics principles are also involved in considering the motion of objects that *roll* down an incline, such as the influence of mass and radius on the object's linear and rotational acceleration. For example, physics principles tell us that the rotational acceleration of an object does not depend on the object's mass, although it does depend on its radius. Further, the linear acceleration of the object is really what is of interest, and it depends not on the object's radius or mass, but rather on how the mass is distributed. This means that all hoops, for example, regardless of mass or size, roll at the same rate down the inclined plane. The same is true for empty cans (i.e., all roll down the incline at the same rate, regardless of mass or size), but an empty can will always roll faster than a hoop. Further, all solid spheres will roll down the incline with

the same acceleration, but every solid sphere, regardless of mass or size, will roll faster than any solid cylinder.

Although these physics concepts and principles are discussed in greater depth than a preschool child would be expected to understand, providing young children with opportunities to explore such concepts in an engaging and collaborative context may lay the foundation to support deeper learning of such concepts in the later school years.

2. Also of interest was whether there is a developmental function for collaborative skills and behaviors over the preschool years (i.e., between the ages of three and five years). Children need to learn to consider each other's perspectives, to become less egocentric in their actions and thinking, and to communicate their ideas effectively to successfully work together to complete the science learning activity. These capacities are related to children's metacognitive and executive functioning skills, as well as theory of mind and social cognition, all of which develop considerably during the preschool years (i.e., DeLuca & Leventer, 2008; Gopnik & Astington, 1988; Kuhn, 2000). Thus, it is possible that older preschool children (i.e., 4 ½- and 5-year-olds) are better able to collaborate, since they have enhanced metacognitive and executive functioning abilities compared to younger preschoolers (i.e., 3- to 4 -year-olds), as well as heightened ability to distinguish the self from others and to take into account different perspectives due to their more developed theory of mind and social cognition (i.e., Tomasello, Kruger, & Ratner, 1993). Cooper's (1980) work supports such hypothesis. In a cooperative problem-solving task involving the use of balance scales to locate matching pairs of blocks, she found that 4-year-olds better coordinate their language with the task, and they use more directive language to assist their partners compared to 3-year-olds. Further, 5-

year-olds are better at offering special expertise, opinions and hypotheses, compared to both 3- and 4-year-olds.

However, it is also conceivable that the differences between younger and older preschoolers will be minimal, as even young preschoolers have been shown capable of distinguishing the self from others and taking into account different perspectives. For example, Shatz & Gelman (1973) show that 4-year-old children can consider the perspectives of others in that they adjust their speech with regard to the changing capacities of different-aged listeners. Thus, it is also possible that differences between the older and younger preschoolers, in terms of collaborative skills and behaviors, may not be large.

4.3 Research Questions.

To summarize, the research questions guiding this work include:

1). *What is the nature of preschool children's collaboration and peer interactions as they work with a peer on a collaborative, physical science learning activity?*

2). *How do children's collaborative abilities develop over the preschool years?*

3). *Do children demonstrate an early understanding of relevant physics principles associated with the science activity?*

4.4 Experimental Details.

4.4a Participants. The children for the study were drawn from a preschool program located within a public, Central New Jersey high school. The preschool program is part of the high school's Child Development program. High school students

have the option to enroll in the course as an elective, and the course involves the high school students running the preschool program. For example, under the direction of the classroom instructor, high school students assist in the design and implementation of instruction. The program comprises half day sessions which are open to 3- to 5-year-old children in the community. Preschool children attend free of charge, on a first-come, first-served basis. The school does not follow a specific curriculum; however, each week includes learning experiences within the areas of literacy, mathematics, science, social studies, and the arts. Children attending the school were predominantly Caucasian and from families of middle-class socio-economic backgrounds.

Once the work received IRB approval, parents of all children were provided information about the study and consent forms. Only those children whose parents returned signed consent forms participated in the study. Thirty-nine forms were returned. As the activity involved children working in pairs, one child could not be included because an odd number of forms were returned. Another pair had to be dropped due to video recording equipment failure. Thus, the final sample involved thirty-six children (20 girls, 16 boys). The mean age was 55.32 months, with a range of 43.62 - 67.10 months. All children spoke English. No child participated in Studies 1 or 2 in this dissertation.

4.4b Materials. The picture book, *Swimmy,* written by Leo Lionni. A wooden ramp of adjustable height. An assortment of TinkerToy™ pieces. (See Figure 4.1 below).

Figure 4.1. *Inclined plane and TinkerToy™ pieces.*

4.4c Methods. The experimenter first spent a few days in the classroom for children to become familiarized and comfortable with her. Next, children were introduced to the concept of *collaboration* through a whole-class, interactive circle time lesson led by the experimenter. The experimenter read a picture book aloud to the class (*Swimmy,* by Leo Lionni) which centers on the concept of collaboration. The class then engaged in a discussion about what collaboration means, and how collaboration can be facilitated. For example, children were told that collaboration means working together to solve a problem or to achieve a goal. Children were asked to share examples from their past of working with others to achieve a goal. The discussion also focused on behaviors for promoting collaboration, or working with others. For example, "to collaborate, it is necessary to listen to one another, share, take turns, help one another, etc."

Next (and on a different day), children were randomly paired with a same-age, same-gender peer. (Note: a median split by age was used to create younger [3 ½ - to 4-year-olds] and older dyads [4 ½ - to 5-year-olds]. The mean age for the younger group

was 50.1 months [range = 43.6 – 57.6]; the mean for the older group was 60.8 months [range = 57.2 – 67.1]. The classroom teacher reviewed all pairs and confirmed that there were no potential conflicts, such as children who tend to not get along well. Further, data collection took place late in the school year, so children knew one another well).

In a corner of their classroom, children were interviewed in pairs and presented with a problem about the movement of a set of interesting objects designed to foster the awareness of the physical properties of objects down an inclined plane. Children were reminded of their circle time discussion about collaboration, and the concept of collaboration was briefly reviewed. Children were then told that they were going to play a game that involved collaboration, or working together. Next, they were introduced to the materials (see Figure 4.1 above) and shown how the angle of the inclined plane could be adjusted by raising or lowering the height of the ramp.

The interviewer then posed questions about specific physical science principles relevant to the materials and asked children to work together in solving the problems. Specifically, children were asked to work together using the TinkerToy™ pieces provided to construct an object that could either roll or slide down the inclined plane (in counterbalanced order). Next, children tested their object(s) on the inclined plane to observe its motion on the ramp. Afterwards, children were asked how increase the objects' speed down the incline, and they were able to test their ideas and predictions using the provided materials. After completing this procedure with both items, children were finally asked which object traversed the ramp at a faster rate, the rolling object or the sliding object.

Children had no prior experience with the inclined plane, and throughout the session they were encouraged to work together to solve the problems. Sessions lasted between nine and eighteen minutes. All sessions were video and audio recorded, and the experimenter took written notes throughout the session on children's behaviors and interactions. See questions and prompts in Table 4.1 below. Also see Appendix E for full scripts and procedures.

Table 4.1. *Prompts and Questions used for the Collaborative Science Activity.*

Review Collaboration: We read this book the other day (*Swimmy,* by Leo Lionni). What was it about? Collaboration, or working together, right? What are some things you need to do when you collaborate, or work together? Today you are going to play a game. For this game, it is really important that you collaborate, or work together. Can you do that?

Construct Object 1: I brought Tinker Toys™ and this ramp with me. The ramp can move up and down like this, see (demonstrate)? Your job is to work together to build something that will *roll* down the ramp. You don't want it to slide down, you want it to roll. You can use any of the pieces, but you need to work together as a team. Let me know when you're finished and then we will test it on the ramp.

Test Object 1 on Ramp: Did it roll? Why/ why not?

Increase Speed Question: How can we make it roll down faster? (If neither child suggests increasing the angle of the incline, ask "What would happen if I move the ramp up, like this?").

Construct Object 2: Are you ready for your next job? This time you need to work together to build something that will *slide* down the ramp. You don't want it to roll down, you want it to slide. Remember that you must collaborate, or work together to build. Let me know when you're finished and then we will test it on the ramp.

Test Object 2 on Ramp: Did it slide? Why/ why not?

Increase Speed Question: How can we make it slide down faster? (If neither child suggests increasing the angle of the incline, ask "What would happen if I move the ramp up, like this?").

Comparing Speed Question: You worked together to make two objects. One rolled down the ramp, and one slid down. Which one moved faster down the ramp? The one that rolled down, or the one that slid (counterbalanced)?

4.5 Analyses.

The data analysis process began with a transcription of the videotapes. These were used to code children's conversations and their context. Children's verbal and nonverbal activity during the course of the session were included in the transcripts. Next, the number of utterances spoken by each child during the session was counted, as well as categorized based on to whom the talk was directed. The proportion of talk that children directed to peers versus the experimenter was then calculated, as was the proportion of children's talk with an unclear recipient. Counts were also made of the frequency and length of verbal exchanges between children.

An overall inspection of the transcripts made it clear that they could be sorted into two groups: *collaborative builders* and *non-collaborative builders.* Collaborative builders included dyads with at least one instance of children working together to collaboratively construct an object to test on the ramp; all others were sorted as non-collaborators. (It is important to note that although non-collaborative builders worked individually to construct objects to test on the ramp, it does not mean that children did not interact with one another. In contrast, all dyads demonstrated forms of collaborative behaviors as they engaged in the task. For example, it was possible for non-collaborative builders to observe and imitate one another, discuss their own and/or their partner's constructions and actions, ask one another for help, provide explanations, etc. Thus, these children engaged in collaborative behaviors, although they did not build a single object collaboratively).

Age and gender differences were explored, as well as differences between the collaborative and non-collaborative builders, in both the frequency and proportion of talk between children, as well as in the amount of time dyads spent on the task.

Video data were also coded to look for any further evidence of children's early forms of collaborative and communicative skills and behaviors. The dyad was treated as the unit of analysis, and only interactions or exchanges (verbal or gestural) between the two children were analyzed, with the possible exception of when one child responded to the interviewer and the other child also became involved in the exchange so that the children appeared to be interacting with each other in some way. The data were coded according to the categories described below.

Coding Categories:

- **Explain/Instruct/Model**- a child demonstrates and/or provides an explanation for how to do or construct something, or instructs his/her partner on how to do or construct something. (e.g., "That (TinkerToy ™ piece) goes on the side. That's how I made it"; "We need it a lot smaller. Try these ones (pieces), try these").

- **Imitate-** a child engages in a behavior that his/her partner recently carried out, such as replicating an object that his/her partner constructed, doing something with the materials recently done by his/her partner, or moving an object down the ramp in a manner similar to his/her partner's previous action.

- **<u>Request Help and/or Information</u>**- a child makes a request and/or asks a question indicating a need for help and/or information (e.g., "How do you make that?"; "I can't get this one in.").

- **<u>Respond</u>**- a child responds to his/her partner's question or request for assistance, such as by providing help (either physical or verbal). For example, a child might give a piece to his/her partner in response to the partner's request, or a child might ask for clarification in response to the partner's question by asking him/her to repeat the question.

- **<u>Agree/Uptake of an idea</u>**- a child accepts an idea suggested by his/her partner (e.g., a child says, "Let's do X", and his/her partner responds, "Good idea!").

Analyses examined mean frequencies and proportions for each of the coding categories, as well as explored group differences in children's collaborative behaviors based on age, gender and whether or not children worked together to construct objects to test on the inclined plane. Also included are typical examples from transcripts of children's collaborative behaviors and interactions (both verbal and non-verbal) as they worked together to solve the problems.

Finally, children's scientific thinking and knowledge related to the task were also measured by scoring each child's responses to the science questions posed by the experimenter intermittently throughout the session. These questions included: *How can we make the object(s) move down the ramp faster?* (Note: this question was asked twice, once for the object that was built to roll down the ramp, and again for the object that was constructed to slide down). And finally, after testing both items on the ramp (i.e., the

rolling object and the sliding object), and experimenting with ways to increase the speed of objects down the ramp, children were asked, *Which object moved down the ramp faster, the one that rolled down or the one that slid down*? (Note: as these questions were asked intermittently through the session, children were able to hear one another's responses).

For the two questions about how increase objects' speed down the incline, children were scored as correct if they suggested increasing the angle of the incline to increase objects' speed down the ramp. Analyses examined the frequencies for each of the different responses provided by children, as well as the proportion of children who provided the correct response (each time the question was asked). We also explored whether children's scores improved from their first attempt at answering this question to the second attempt.

For the question about whether the rolling or sliding object traversed the ramp at a faster rate, children were scored as correct if they stated that the rolling object was faster. All other responses were scored as incorrect. Analyses examined the proportion of correct responses, as well as age and gender differences in children's scores.

4.6 Results

4.6a Extent of Collaboration on Object Construction. After repeated readings of sets of transcripts, each dyad was sorted as *collaborative builders* or *non-collaborative builders*. This was based on whether the children in the dyad worked collaboratively to construct a single object to test on the inclined plane at least once (for any given amount of time) over the course of the session, or in contrast, if each child worked individually to

build his/her own object(s) to test on the ramp. Of the 18 dyads, 8 (44%) were sorted as

collaborative builders and 10 (56%) as non-collaborative builders. Mann-Whitney U

tests (2-tailed) confirmed that there were no age or gender differences among the two

groups. For age, $U = 40.50$, $p = 1.00$, and for gender, $U = 36.00$, $p = .76$. See Table 4.2

below for the number of dyads from each age and gender group comprising the

collaborative and non-collaborative builders. Neither were there differences in the

average amount of time dyads from each group spent on the task (see analysis below).

Table 4.2. *Collaborative and Non-Collaborative Builders, Dyads by Age and Gender.*

Group	Collaborative Builders	Non-Collaborative Builders
Younger Boys	$n = 2$ dyads	$n = 2$ dyads
Younger Girls	$n = 2$ dyads	$n = 3$ dyads
Older Boys	$n = 2$ dyads	$n = 2$ dyads
Older Girls	$n = 2$ dyads	$n = 3$ dyads
Total	**$n = 8$ dyads**	**$n = 10$ dyads**

Thus, overall, the science activity fostered collaborative building among children

in only approximately half of the sample. This was an unexpected result. Potential

reasons for this finding are discussed in the conclusions section below.

4.6b Amount of Time Spent on Task. The amount of time that children spent

engaged in the activity ranged from 9 to 18 minutes, with a mean of 11.44 minutes ($SD =$

3.32). Analyses explored whether there were group differences in the average amount of

time spent on the task based on age, gender, as well as based on whether children worked

collaboratively to construct objects to test on the inclined plane (i.e., collaborative versus

non-collaborative builders). Mann-Whitney U tests (2-tailed) found no group

differences. For collaboration (i.e., collaborative versus non-collaborative builders), $U =$

36.00, $p = .72$. For age, $U = 38.00$, $p = .83$. Finally, for gender, $U = 23.00$, $p = .13$.

Thus, there were no group differences in the amount of time dyads spent working to solve

the problems.

   4.6c Examples from Transcripts: Collaborative Behaviors and Conversations

Between Children. Although just under half of the dyads (45%) worked together to

collaboratively build a single object, it was evident that all dyads demonstrated a number

of collaborative and communicative behaviors as they engaged in the task. Below are

some examples from transcripts illustrating children's collaborative interactions and

conversations as they engaged in the activity together, including how they were coded

using the scheme described above. Such examples represent typical peer interactions

from the total sample, and they include boy and girl dyads, older and younger dyads, as

well as collaborative and non-collaborative builders. These examples illustrate that

although not all dyads worked collaboratively to build a single object, they were still able

to carry out collaborative interactions as they engaged in the task.

Example 1. (Pair 8. Younger girls; Collaborative builders).

   (Children are working to build a single object together).
   Child 1- Let's put this red one right here. E/I/M
   Child 2- And I'm gonna put this (piece) there. (She tries to connect a piece but
         has difficulty). I can't. RH/I
   Child 1- Here, I'll help you. R
   Child 2- Is it done? (She notices a missing piece). Uh oh. (Attempts to fix it).
   Child 1- (Both girls work to add wheels to the object). Let's try the side. E/I/M
   Child 2- Like this? RH/I
   Child 1- Yeah, good. R
   Child 2- Now let's try it down (the ramp), okay? E/I/M
   Child 1- Yeah! A/U
   Child 2- (Sets object on ramp but it doesn't move).
   Child 1- Maybe if we push it?
   Child 2- Yeah. (Pushes object so it moves down the ramp). A/U
   Child 1- It worked!

Example 2. (Pair 9. Older boys; Collaborative builders).

> Child 1- I'm making this. But I'm not putting it like this; I'm putting it up
> (demonstrates with his object). But it's hard to make. E/I/M
> Child 2- (Watches as Child 1 builds, and then hands a piece to Child 1).
> Child 1- Put that one right here. E/I/M
> Child 2- Okay, my turn (and follows Child 1's instruction). A/U
> Child 1- Then put theses (pieces) here. E/I/M
> Child 2- (Attempts to connect a piece to the object).
> Child 1- There's one here and down here. (Points to wheels on the object). E/I/M
> Child 2- We need another one (stick). Here's one. (Holds up green stick and
> hands it to Child 1).
> Child 1- No, that won't really work. Try that one over there. E/I/M
> Child 2- Okay. (Picks up the piece and hands it to Child 1). A/U
> Child 1- (Connects the piece to the object, and then a piece falls off). Oh no!
> Child 2- Do you need some help? R
> Child 1- Now (we need) one long one right here. E/I/M
> Child 2- What? RH/I
> Child 1- We need one more orange one right here. R, E/I/M
> Child 2- I can't find another orange one.
> Child 1- (Picks up a green stick and attaches it to the object).
> Child 2- Yeah that one. I know. A/U

Example 3. (Pair 19. Younger girls; Non-collaborative builders).

> Child 1-Can I have that back? (Child 2 has an object that Child 1 built). RH/I
> Child 2- Here. (Returns the object). R
> Child 1- Thank you. If you want, you can make a small one like this one. There's
> another one (piece) if you want to make one like that (i.e., like the object
> she just returned). (Child 1 hands Child 2 a piece she needs to replicate
> the object).
> Child 2- (Takes the piece from Child 1 and connects it to her wheel). We need
> another one of these (pieces). A/U, RH/I
> Child 1- (Looks for the piece Child 2 needs, and finds it in an object built earlier.
> She hands the object to Child 2). This has one. That goes on the side.
> Like this. That's how I made it. R, E/I/M
> Child 2- Okay. (Successfully replicates Child 1's object). A/U, IM
> [Girls place both of their objects on the ramp, without releasing them].
> Child 2- Let's see what happens if we both work together!
> [Girls both release their objects at the same time and they roll down.]

Example 4. (Pair 1. Older girls; Collaborative builders).

       Child 1- How did you make that? RH/I
       Child 2- I used the thing that I did to roll and then I put these on the end. R; E/I/M
       Child 1- (Begins replicating Child 2's design). Wait, the blue one has to go like
             there, right? IM; RH/I
       Child 2- Yes. R
       Child 1- And now I just need to… (Child 2 interjects).
       Child 2- And the red one has to go on the bottom and you put two sticks right
             there. E/I/M

Example 5. (Pair 9. Older boys; Collaborative builders).

       (Working collaboratively to build an object).
       Child 1- Too big. It's too big.
       Child 2- How about we make it shorter?
       Child 1- Okay. Here's a red (stick). Make it shorter. A/U, E/I/M
       Child 2- Take this one out (referring to a piece on the original object). E/I/M
       Child 1- Okay. And I found another red one. A/U
       Child 2- We need another (stick).
       Child 1- What if it's on the top? What if we go like that? (Demonstrates where to
             put the piece on top of their object). E/I/M
       Child 2- Oh yeah. We have to make it together (connect the two pieces). A/U
       Child 1- Yeah! And I want to be the green one, that's my guy. A/U
       Child 2- No, this one goes right here. E/I/M
       Child 1- You need another one here (piece). E/I/M

Example 6. (Pair 10. Younger boys; Non-collaborative builders).

       Child 1- (Working to make a rolling object). It can't roll. RH/I
       Child 2- I think you just need a wheel. Like this. (Demonstrates by dropping
             wheel on the ramp, but it falls over and slides down). That's how you roll
             it. (He tries again, and the wheel rolls down). That's it. R; E/I/M
       Child 1- Oh, that's how you roll it! (Imitates Child 2 and puts a single wheel on
             the ramp, but it falls over and slides down). It didn't work. A/U; IM; RH/I
       Child 2- No, you just need to do this. This is how you roll it. This way.
             (Demonstrates). R; E/I/M
       Child 1- Oh. (Tries again, once unsuccessfully, then successful). It worked! IM

       4.6d Examining Individual Children's Talk. For each dyad, the number of

utterances made by each individual child was tallied, as well as categorized based on to

whom the talk was directed. Children's talk could be directed to either their peer or the

experimenter; otherwise it was considered ambiguous. The frequencies were then

converted to proportions to control for differences in the length of time each dyad spent

on the task. See Table 4.3 below for descriptive statistics.

Table 4.3. *Proportions for Children's Talk, by Group.*

	Talk Btwn Children	**Children to Exp.**	**Child Talk Ambig.**
Collaborative Builders			
Mean (SD)	0.42 (0.15)	0.35 (0.07)	0.23 (0.09)
Range	0.18 - 0.63	0.25 – 0.45	0.10 – 0.36
Non-Collaborative Builders			
Mean (SD)	0.23 (0.12)	0.49 (0.09)	0.28 (0.05)
Range	0.04 - 0.42	0.33 – 0.58	0.22 – 0.38
Older Dyads			
Mean (SD)	0.41 (0.16)	0.38 (0.11)	0.21 (0.07)
Range	0.20 - 0.63	0.25 – 0.57	0.10 – 0.32
Younger Dyads			
Mean (SD)	0.22 (0.11)	0.48 (0.08)	0.30 (0.05)
Range	0.04 - 0.36	0.38 – 0.58	0.26 – 0.38
Boy Dyads			
Mean (SD)	0.35 (0.19)	0.40 (0.11)	0.25 (0.10)
Range	0.14 - 0.63	0.25 – 0.54	0.10 – 0.36
Girl Dyads			
Mean (SD)	0.29 (0.14)	0.45 (0.11)	0.26 (0.05)
Range	0.04 - 0.49	0.30 – 0.58	0.21 – 0.38
Full Sample			
Mean (SD)	0.31 (0.16)	0.43 (0.11)	0.26 (0.07)
Range	0.04 - 0.63	0.25 – 0.58	0.10 – 0.38

As can be seen in the table above, when looking at the sample as a whole,

approximately one-third of children's talk was directed to peers, while just under half was

directed to the experimenter. The rest of children's utterances lacked a clear recipient.

This suggests that the task encouraged some interaction between children, while also

supporting interaction between children and the experimenter. (It is important to note

that the relatively high proportion of talk directed to the experimenter results in part from the structure of the task. As the experimenter directed questions to children intermittently throughout the session, much of children's talk directed to the experimenter consisted of their responses to her questions). However, it is also important to note that the variation between dyads in terms of between-peer talk was fairly large ($SD = 0.16$). For example, some dyads exhibited a high proportion of between-peer talk (i.e., for two dyads, the proportion of between-child talk was greater than 60%), whereas others exhibited much less (i.e., for three dyads, the proportion of between-child talk was less than 15%).

Mann-Whitney U tests (2-tailed) were conducted to explore differences in the proportions of children's talk based on whether dyads were sorted as collaborative or non-collaborative builders. Results indicated that collaborative builders demonstrated a significantly higher proportion of between-children talk ($Mdn = 0.40$, mean rank = 12.88, $n = 8$ dyads) than did those sorted as non-collaborative builders ($Mdn = 0.22$, mean rank = 6.80, $n = 10$ dyads), $U = 13.00$, $p = .016$. This pattern of speech suggests that children who worked collaboratively to build items to test on the incline were perhaps more jointly engaged in the task than were the non-collaborative builders. Further, the non-collaborative builders directed a significantly higher proportion of talk to the experimenter ($Mdn = 0.53$, mean rank = 12.40) compared to the collaborative builders ($Mdn = 0.37$, mean rank = 5.88), $U = 11.00$, $p = .01$. This pattern of speech indicates that the non-collaborative builders may have relied more on the experimenter for feedback and support as they engaged in the task than did the collaborative-builders, who may have looked to their peers more for such support (i.e., as they were more jointly engaged in the task). There was no difference between the two groups in the proportion

of talk with an ambiguous recipient, $U = 23.00$, $p = .13$. Of course, it is necessary to replicate these patterns with larger samples.

The same pattern was found for the older versus younger dyads. Specifically, the older dyads ($n = 9$) exhibited a higher proportion of between-children talk than did the younger dyads $(n = 9)$. The median for the older dyads was 0.42 (mean rank = 12.61), compared to a median of 0.20 (mean rank = 6.39) for the younger dyads, $U = 12.50$, $p = .013$. This again indicates that older children were perhaps more jointly engaged in the task than were younger children. Further, younger children directed a higher proportion of their talk to the experimenter than did older children. The median for the younger children was 0.47 (mean rank = 12.11), compared to median of 0.34 (mean rank = 6.89) for the older children, $U = 17.00$, $p = .038$. This pattern of speech again suggests that the younger children may have relied more on the experimenter for support and feedback as they engaged in the activity than did the older children. The older children, in contrast, may have looked more to their peers for such support, as their pattern of speech indicates that they were perhaps more jointly engaged in the task. Finally, the younger children also exhibited a significantly higher proportion of ambiguous talk compared to the older children. The median for the younger group was 0.28 (mean rank = 12.94), compared to a median of 0.22 (mean rank = 6.06) for the older dyads, $U = 9.50$, $p = .006$. Again, it is necessary to replicate these patterns with larger samples.

Finally, no gender differences were found in terms of children's patterns of speech. In comparing the proportion of talk between children by gender, $U = 36.00$, $p = .72$. For talk directed to the experimenter, $U = 26.50$, $p = .23$; and for ambiguous talk, $U = 35.50$, $p = .69$.

4.6e Number and Length of Uninterrupted Verbal Exchanges Between Children.

The previous analyses examined patterns in individual children's speech (i.e., how much of each child's speech was directed to peers, the experimenter, and ambiguous). This section discusses patterns in children's conversations with one another. Specifically, analyses explored the frequency and duration of conversations occurring between children as they engaged in the activity, as well as group differences in each. In other words, the number of instances in which children engaged in verbal exchanges involving only their peer (i.e., the number of between-peer conversations) was tallied, as was the total number of verbal utterances exchanged between children during those conversations. (Note: See above for examples from transcripts of the between-child conversations discussed here).

Overall, the mean number of between-child conversations occurring over the course of the session was 3.00 (SD = 1.64), with a range of 0.00 to 7.00. Further, the mean number of verbal exchanges comprising those conversations was 6.99 (SD = 3.43), with a range of 0.00 to 12.57.

Mann-Whitney U tests (2-tailed) were conducted to explore group differences in the number of between-child conversations. Results found a significant difference between the older (n = 9 dyads) and younger (n = 9 dyads) groups. The median number of between-child conversations for the older group was 4.00 (mean rank = 11.89), compared to a median of 2.00 (mean rank = 7.11) for the younger group, U = 19.00, p = .05. Collaborative builders (n = 8 dyads) also engaged in significantly more between-child conversations than did non-collaborative builders (n = 10 dyads). The median for the collaborative builders was 4.00 (mean rank = 12.19), compared to a median of 2.50

(mean rank = 7.35) for the non-collaborative builders, $U = 18.50$, $p = .05$. Finally, no gender differences were found, $U = 30.00$, $p = .36$. See Figure 4.2 below for mean frequencies, by group.
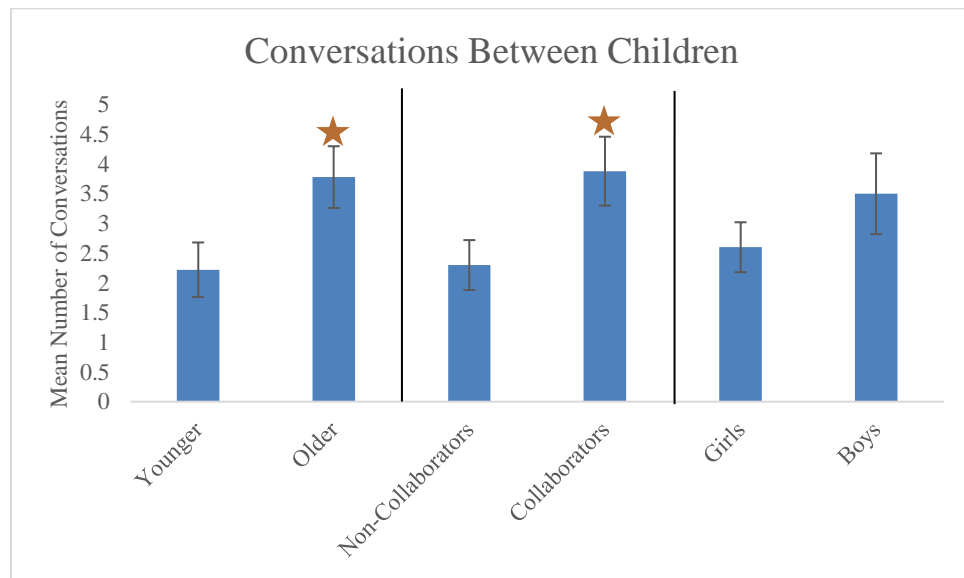
Figure 4.2. *Mean number of conversations between children, by group.*
Note: Error bars on all graphs represent one standard error of the mean[2].

Mann-Whitney U tests (2-tailed) were also conducted to explore group differences in the number of verbal exchanges comprising those between-peer conversations. A significant difference was found between the younger and older dyads. The median number of verbal exchanges comprising older dyads' between-peer conversations was 8.00 (mean rank = 12.22), compared to a median of 5.50 (mean rank = 6.78) for the younger dyads, $U = 16.00$, $p = .03$. There was no difference between the collaborative and non-collaborative builders, $U = 22.50$, $p = .12$. Nor was there a difference between boy and girl dyads, $U = 35.00$, $p = .66$. See Figure 4.3, below, for mean frequencies, by group.
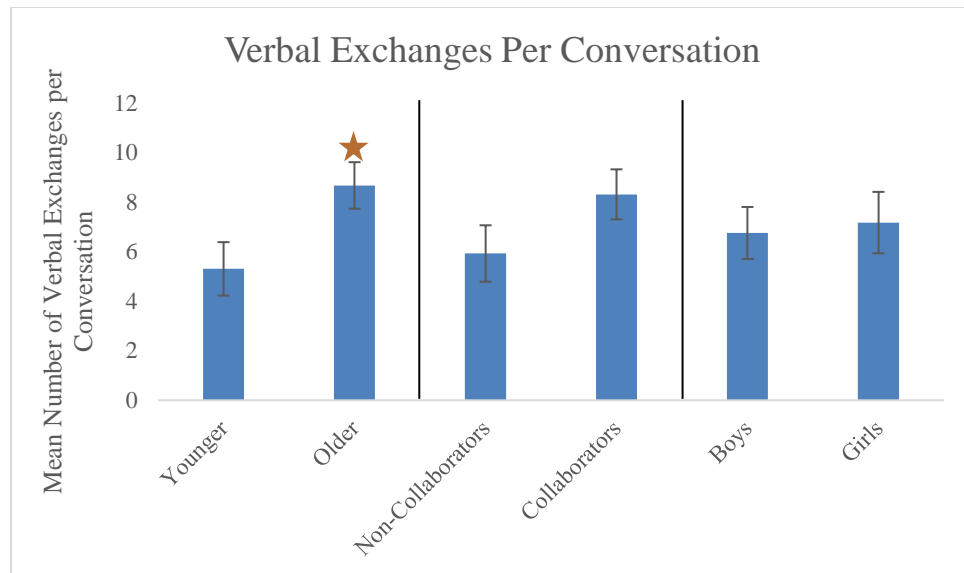
Figure 4.3. *Mean number of verbal exchanges per conversation between children, by group.*

In summary, dyads from the older half of the sample engaged in more frequent conversation with peers, and those conversations consisted of significantly more verbal exchanges between children, compared to the dyads from the younger half of the sample. Further, the dyads who worked collaboratively to construct objects carried out significantly more conversations between children than did the non-collaborative builders. This is despite the fact that both groups included equal numbers of dyads from the upper and lower age groups. However, the number of verbal exchanges comprising the conversations between children was similar in both groups. Finally, boy and girl dyads showed no differences in terms of the frequency of conversation occurring among peers, nor in the number of verbal exchanges between children comprising such conversations. However, it is important to note that the sample size is small, so it is necessary for future work to replicate these findings with data from larger and more diverse samples.

4.6f Frequencies for Collaborative Behaviors. To more closely examine the
frequency and extent of these early forms of collaborative skills and behaviors, the video
data were coded using the coding scheme described above. Further, interrater reliability
was established by having two coders code a random sample of transcripts (22%) from
this database. Agreement was relatively high, with raters agreeing on 144 out of 172
codes applied (0.84). All disagreements were resolved through discussion. See Table 4.4
below for mean frequencies for each coding category, by group.

Table 4.4. *Mean Frequencies for Each Coding Category, by Group.*

Group	E/I/M	IM	RH/I	R	A/U	Total
Collaborators (n=8 dyads)	10.13	1.63	9.00	3.25	4.25	28.25
Non-Collaborators (n=10 dyads)	7.90	1.50	3.00	1.30	1.40	15.10
Older Half (n=9 dyads)	10.33	1.78	6.11	2.11	3.11	23.44
Younger Half (n=9 dyads)	7.44	1.33	5.22	2.22	2.22	18.44
Boys (n=8 dyads)	10.00	1.88	7.88	3.75	3.75	27.25
Girls (n=10 dyads)	8.00	1.30	3.90	0.90	1.80	15.90
Full Sample (n= 18 dyads)	**8.97**	**1.57**	**5.85**	**2.26**	**2.76**	**21.40**

Mann-Whitney U tests (2-tailed) were conducted to explore group differences in
the total number of codes applied for each dyad. The collaborative-builders exhibited a
strong tendency to engage in more collaborative behaviors than did non-collaborative
builders. This difference reached marginal significance, $U = 19.00$, $p = .06$. However,
no age or gender differences were found. For age, $U = 21.00$, $p = .08$; for gender, $U =
29.00$, $p = .33$.

Mann-Whitney U tests (2-tailed) were also conducted to explore age, gender, and
collaboration (collaborative builders, non-collaborative builders) effects on the mean
frequency of each individual coding category. First, in looking at age, the only difference
found between the older and younger dyads was that older dyads requested help and/or

information from peers significantly more often than did younger dyads. The median for the older group was 6.00 (mean rank = 12.00), compared to a median of 2.00 (mean rank = 7.00) for the younger group, $U = 18.00$, $p = .045$. There was a tendency for older dyads to engage in more explaining, instructing and modeling to peers compared to the younger dyads, but this difference did not reach significance. See Figure 4.4 below.
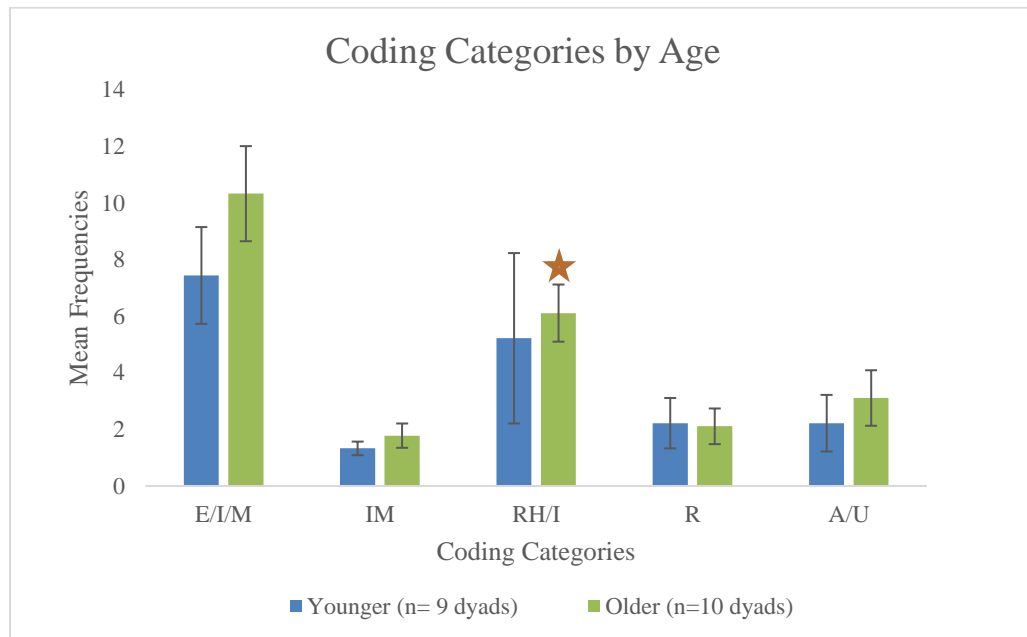
Figure 4.4. *Mean frequencies for each coding category, by age group.*

Next, in looking at collaboration, only one group difference was found. Specifically, the collaborative builders requested help and/or information from peers significantly more often than did non-collaborative builders. The median for the collaborative builders was 6.50 (mean rank = 12.69), compared to a median of 3.00 (mean rank = 6.95) for the non-collaborative builders, $U = 14.50$, $p = .02$. See Figure 4.5 below.
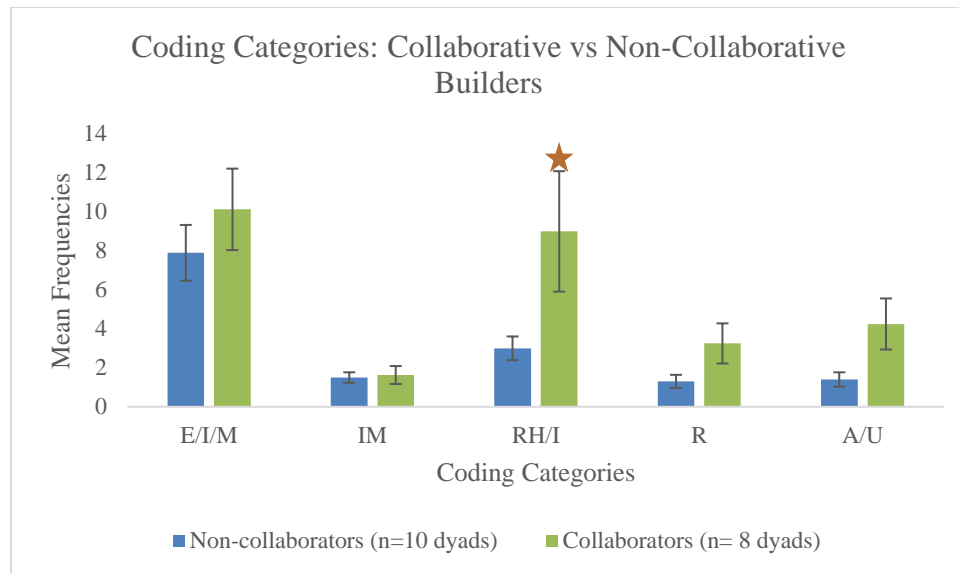
Figure 4.5. *Mean frequencies for each coding category, for collaborative and non-collaborative builders.*

Finally, for gender, again only one group difference was found. Specifically, the boy dyads responded to peers' questions and/or requests for help significantly more often than did girl dyads. The median for the boy dyads was 2.50 (mean rank = 13.75), compared to a median of 1.00 (mean rank = 6.10) for the girl dyads, $U = 6.00$, $p = .002$. See Figure 4.6 below.
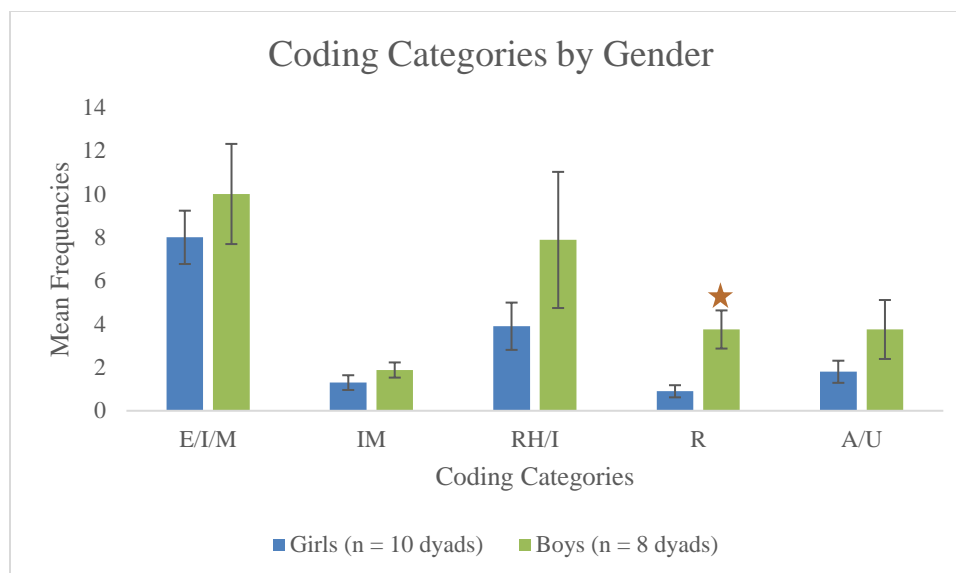
Figure 4.6. *Mean frequencies for each coding category, by gender.*

(Note: The above analyses were also conducted controlling for the amount of time dyads spent on the task. Results were nearly identical[3]. The only difference was that when controlling for time, older dyads were also found to engage in significantly more explaining, instructing and modeling (E/I/M) to peers compared to younger dyads, $U = 18.00$, $p = .047$. All other results were the same).

To control for variations between pairs in the total amount of time spent on the task, the frequency data for each coding category were also converted to percentage values by converting the raw frequency scores to a proportion of the total. See Table 4.5 below for mean proportions for each coding category, by group.

Table 4.5. *Mean Proportions for Each Coding Category, by Group.*

Group	E/I/M	IM	RH/I	R	A/U
Collaborators (n=8 dyads)	0.37	0.08	0.30	0.11	0.14
Non-Collaborators (n=10 dyads)	0.50	0.11	0.22	0.08	0.08
Older Half (n=9 dyads)	0.44	0.07	0.28	0.08	0.12
Younger Half (n=9 dyads)	0.45	0.13	0.23	0.10	0.09
Boys (n=8 dyads)	0.36	0.11	0.27	0.15	0.11
Girls (n=10 dyads)	0.51	0.09	0.25	0.05	0.10
Full Sample (n= 18 dyads)	**0.44**	**0.10**	**0.26**	**0.09**	**0.11**

Mann-Whitney U tests (2-tailed) were conducted to explore differences in proportions for each coding category by age, gender, and collaboration. No age differences, nor differences between-collaborative and non-collaborative builders, were found in the proportions for each coding category. For gender, the only significant difference was that the proportion for code R (responding to a peer's request for help and/or information) was significantly higher among boy dyads (n = 8) compared to girl dyads (n = 10). The median for the boy dyads was 0.14 (mean rank = 14.38), compared to a median of 0.06 (mean rank = 5.60) for girls, U = 1.00, p < .001. Thus, the patterns for the proportions for each coding category across groups were rather similar.

4.6g Individuals' Responses to the Posed Science Questions: *How to Increase Objects' Speed*. Recall that one science question asked children how to make the objects traverse the incline at a faster rate. Further, this question was asked twice over the course of the session, once for the object that children constructed to roll down the incline, and once for the object built to slide down. Figure 4.7, below, illustrates the frequencies for children's responses to these questions about how to increase objects' speed down the incline, for each time this question was asked.
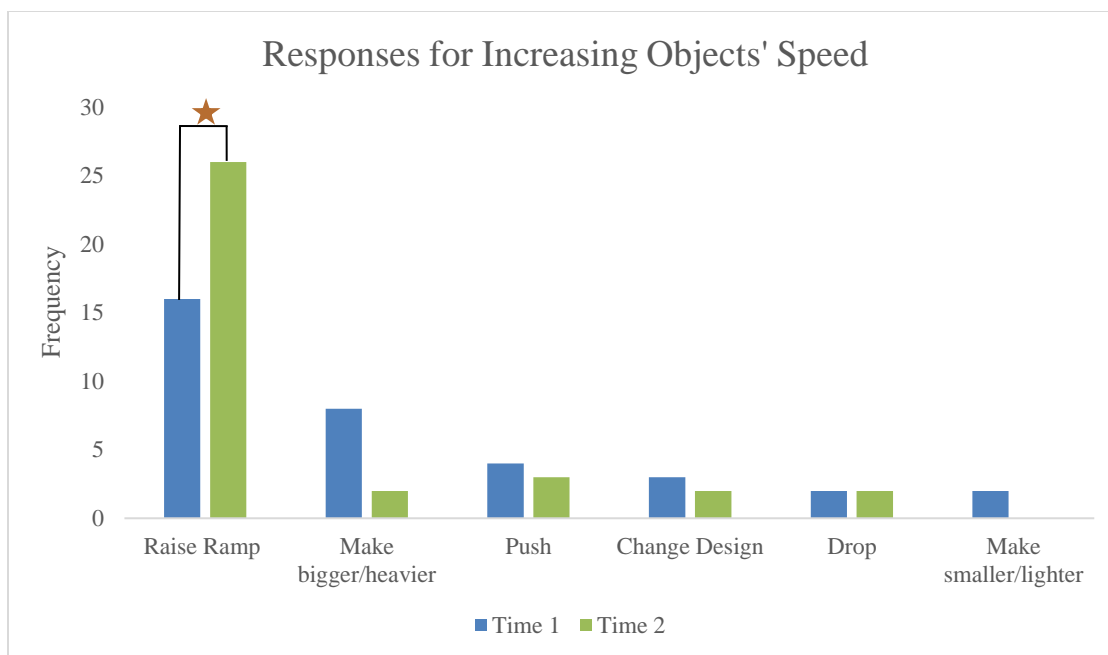
Figure 4.7. *Dyads' responses for how to increase objects' speed down the inclined plane. (Note: as this question was asked twice, the bars represent individuals' responses for the first and second time the question was asked).*

As can be seen in Figure 4.7 above, the most frequent response provided by children for how to increase objects' speed down the ramp was the correct answer of raising the height of the ramp, or increasing the angle of the incline. Twenty-seven children (75%) provided this correct answer at least once (i.e., for the two times the question was asked). This indicates that the majority of children demonstrated an understanding of the relevant physics principle that increasing the angle of the incline on the ramp results in an increase in objects' speed down the incline.

Other suggestions provided for how to increase objects' speed down the ramp included: making the object bigger or heavier (adding more pieces); making the object smaller or lighter; placing the object on the ramp and then pushing it down; altering the design of the object in some non-specified way; and finally, dropping the object onto the ramp.

Analyses examined whether scores increased from children's first attempt at answering the question to the second attempt. A total of 16 children (44%) provided the correct response of increasing the angle of the incline to increase objects' speed down the incline the first time the question was asked, compared to 26 children (72%) for the second time. A paired samples t-test (2-tailed) confirmed that this difference was significant; $t(35) = 3.25$, $p = .003$. This suggests that children learned from their first attempt at experimenting with ways to increase objects' speed down the ramp that raising the angle of the incline did in fact work as intended. Further, with the exception of the correct response of increasing the angle of the incline to increase objects' speed down the ramp, the frequency all other responses decreased from the first time the question was asked to the second time. This again indicates that children learned from their experience with testing predictions the first time to determine which approach was indeed effective, as well as which were unsuccessful, in increasing objects' speed down the incline.

Finally, Chi-Square tests were also conducted to explore age and gender differences in children's responses to these two questions. Specifically, analyses explored group differences in the number of children who provided the correct answer at least once (out of the two times the question was asked). The difference between older and younger children reached significance. A total of 16 children (89%) from the older group at least once provided the correct response of increasing the angle of the incline to increase objects' speed down the ramp, compared to 11 (61%) from the younger group, $\chi^2(1) = 3.704$, $p = .05$. The difference between boys and girls was not significant, $\chi^2(1) = 0.60$, $p = .44$.

In summary, the majority of children (75%) appropriately suggested raising the angle of the incline on the ramp to increase objects' speed, and children were significantly more likely to provide this correct answer the second time the question was asked (i.e., after testing various predictions the first time), compared to the first. Further, children from the older half of the sample were more likely to provide this correct answer compared to the younger children.

4.6h Individuals' Responses to the Posed Science Questions: *Comparing Speed for Rolling vs. Sliding Objects.* After testing both objects on the inclined plane (i.e., the one children constructed to slide down the incline, and the one built to roll down), as well as experimenting with ways to increase the speed of those objects down the incline, children were asked which object traversed the ramp at a faster rate, the one that rolled down or the one that slid.

A total of 26 children (72%) provided the correct response that the rolling object traversed the ramp at a faster rate compared to the sliding object. Children from both the younger ($n = 18$) and older ($n = 18$) half of the sample performed above chance levels (.50). For the older group, the mean number correct was 0.72 ($SD = 0.46$), $t\,(17) = 2.046$, $p = .057$. The mean was the same for the younger group ($M = 0.72$, $SD = 0.46$), $t(17) = 2.046$, $p = .057$. Chi Square tests found no age or gender differences in terms of the number of children from each group correctly answering this question. For age, $\chi^2(1) = 0.00$, $p = 1.00$. For gender, $\chi^2(1) = 1.36$, $p = .24$. In sum, the majority of children demonstrated an understanding of the relevant physics principle that an object's mechanism for motion effects its rate of speed down the incline. See Figure 4.8 below.
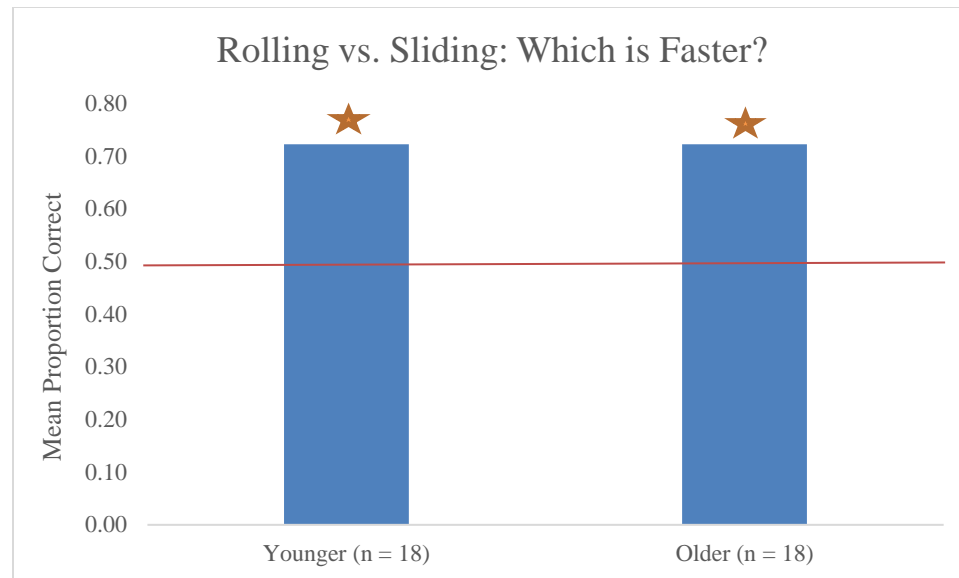
Figure 4.8. *Mean proportion of correct responses by age group, in comparing the speed of rolling versus sliding objects.*

4.6i Examples of Rare, Non-Collaborative Interactions. Finally, there were also instances in which children were not engaged in collaborative interactions as they worked on the science activity together. Most frequently, this occurred when children were focused on building their own individual constructions to test on the ramp, without paying attention to what their partners were doing. In such cases, children were not collaborating because they were focused on their own individual work. There were also a very small number of cases in which children carried out non-collaborative behaviors, or behaviors detrimental to fostering successful cooperation (i.e., ignoring or refusing a peer's attempt to collaborate, not answering a peer's question and/or request for help, etc.). Again, this was rare, occurring only five times in total. Thus, although children often spent time focusing on their own work, it was rare for them to exhibit non-collaborative behaviors.

4.7 Conclusions and Implications for Early Science Education

The research results presented here show that preschool children ages 3 ½ to 5 years can engage in collaborative and communicative behaviors when working with a same-aged peer to solve simple physics problems. Specifically, after coding for specific collaborative and communicative behaviors, it was found that pairs of children frequently demonstrated such behaviors as they worked together on the task, doing so (on average), 21 times over the course of the session, or at a rate of approximately 2 collaborative behaviors per minute. Children often provided explanations to peers, and/or instructed and modeled to peers how to do or make something. Children also often requested help and/or information from their partners. Although somewhat less frequently, children also accepted peers' ideas and/or suggestions, responded to peers' questions and/or requests for assistance, and imitated peers in some way. Further, the patterns across groups (i.e., based on age, gender, and whether pairs built objects collaboratively) were rather similar.

The largest difference found between older and younger dyads was that older children tended to talk with peers more frequently compared to the younger children. A higher proportion of the older children's talk was directed to peers, and they engaged in significantly more between-peer conversations compared to the younger dyads. (Their conversations also consisted of more verbal exchanges between children). It is possible that older children's enhanced cognitive skills (i.e., metacognitive and executive functioning skills, as well as theory of mind and social cognition, all of which develop considerably during the preschool years) supported their more frequent and elaborate forms of communication.

Further, the more frequent and extended patterns of communication between older children suggests that they were perhaps more jointly engaged in the task than were the younger children (who tended to talk more to the experimenter). For example, older children may have relied more on their peers for feedback and support as they engaged in the task compared to younger children, whom in contrast tended to rely more on the experimenter for such support and feedback. It is also possible that older children simply preferred talking with peers more than the younger children. Thus, it is necessary to continue exploring these questions, as well as replicate these patterns with larger samples.

However, overall, the patterns and frequencies for engaging in collaborative and communicative behaviors were rather similar across the age groups. This indicates that children as young as 3 ½ years have some capacity to distinguish the self from others and consider another's perspective, as well as to communicate ideas effectively when working together with a peer to solve simple physics problems.

In addition to the ability to engage in collaborative and communicative behaviors, children in the study also exhibited an early understanding of relevant physics principles. The majority of children in the study demonstrated an understanding of the principle that an object's mechanism for motion down the incline has an effect on the object's speed (i.e., rolling objects traverse the ramp at a greater rate of speed compared to sliding objects). The majority of children also showed some understanding of the physics principle that increasing the angle of the incline on the ramp results in an increase in objects' speed. Notably, children were significantly more likely to answer this question correctly the second time it was asked compared to the first time. This suggests that children learned from their first attempt at experimenting with ways to increase objects'

speed down the ramp to determine which approach was indeed effective, as well as which were unsuccessful. That children benefitted from the repeated experience with exploring ways to increase objects' speed down the ramp is an important finding, with a clear pedagogical implication. Specifically, young children may benefit from opportunities to try again as they engage in scientific investigations involving the exploration of different items and materials. This is worth further research exploration.

Taken together, these findings suggest that the task represents a suitable context for cultivating collaborative and communicative behaviors among preschool children. The context also constitutes a developmentally-appropriate set of conversational options and objects that can support science learning in the early childhood classroom. Since the materials are neither too easy nor too difficult, the teacher has room to help children actively construct understanding on what they already know something about. Providing such opportunities for children to actively construct knowledge with peers in an engaging context has the potential to lay the foundation for science learning in the later school years, as well as support children's developing abilities to work with others.

Finally, although all dyads exhibited a number of collaborative and communicative behaviors as they worked together to solve the problems, only approximately half of the dyads worked collaboratively to construct a single object to test on the inclined plane during the activity. (In other words, collaborative builders worked together to build one object to test on the ramp, while non-collaborative builders each built their own objects). This was an unexpected finding, as it was hypothesized that children would be motivated to build collaboratively as they engaged in the task together. (Again, it is also important to note that the non-collaborative builders still engaged in

similar patterns of collaborative and communicative behaviors compared to those who built collaboratively).

There are a number of potential reasons for why dyads may not have been overly inclined to work together to build objects to test on the inclined plane. First, the instructions may have been ambiguous to children. Children were asked to work together to construct an item to roll or slide down the inclined plane. It is possible that they interpreted the directions to mean that they should each build their own structure to test on the ramp. Second, children were not introduced to the materials before engaging in the task. In reviewing video data, it became clear that many children found the TinkerToy™ building pieces to be novel and exciting, with children often immediately becoming engaged in exploring with the materials individually. This novelty effect may have hindered children's collaborative building by increasing their desire to explore with the new and exciting materials on their own.

A third potential reason for dyads' lesser tendency to build collaboratively is the abundance of materials. This allowed for children to construct complex objects on their own, and may have encouraged children to build individually as there were enough materials for children to be successful in building on their own. It is possible that providing dyads with fewer materials would result in more collaborative building; for example, by providing so few materials that it is necessary for children to work together to be successful (i.e., only a small number of wheels, connector pieces, etc.).

Another possible explanation for children's lesser tendency to build collaboratively as they worked on the task is that the experimenter was in close proximity throughout the duration of each session. This was necessary for the experimenter to

guide children through the science activity (i.e., by asking them questions, providing directions and prompts, etc.), as well as to closely observe, and take comprehensive fieldnotes on, children's collaborative interactions and behaviors as they engaged in the task. This close proximity with the experimenter may have led to a "teacher effect," where children assumed the experimenter to take on the role of instructor. For example, this may have increased children's tendency to rely on the experimenter for guidance or assistance when working to solve the physics problems, rather than on their peers, as it is commonplace for students to look to teachers for assistance and support on various classroom activities.

Finally, children engaged in the activity in the block-building area in the back of their classroom. Although they were in a separate area of the classroom from the rest of their classmates, children at times became distracted by the noise of other activities going on simultaneously within the classroom. This distraction may have reduced children's collaborative building by interrupting children's shared focus. Thus, specific task features, such as the specific instructions used, the quantity and novelty of materials, the proximity of the experimenter, and background noise/distractions may have hindered children's tendency to build collaboratively as they engaged in the task. Future studies are being planned to explore whether children's tendency to build collaboratively with peers increases when these possible inhibiting factors are eliminated and/or reduced.

In summary, although collaboration, communicating ideas, developing arguments, and working with peers are important features of the curriculum bands and concepts in later elementary and secondary school, these notions tend to receive little attention in the early years. This work is important as it closely examines the features of preschool

children's peer interactions as they work together to solve problems in science. Such

work can be helpful in informing educators about preschool children's social and

cognitive abilities, so that such abilities can then be considered as features in designing

educational environments and pedagogy in science that benefit from, as well as support,

children's collaborative abilities. In other words, such understandings can be used to

shed light on pedagogy and learning activities for the early childhood classroom which

may support goals for both early science education, and for working with others. Future

work should continue to explore task features which may better promote sustained and

effective collaboration among young children in science learning contexts within the

preschool classroom, as well as the development of collaborative skills in naturalistic,

classroom contexts over the early school years.

Limitations. As previously discussed, a number of potential factors may have

inhibited children's tendency to build collaboratively as they engaged in the task together

(i.e., the instructions used, novelty effect, ample materials, background distractions,

proximity of experimenter; see above). Future studies are being planned to reduce and/or

eliminate these possible inhibiting factors to determine whether it results in an increased

tendency for children to work collaboratively.

Another limitation of the study is that participants were not asked to provide

background information related to participants' prior experience with other children (i.e.,

whether they have siblings [and their ages]; the number of years they have been in

school; previous forms of childcare they have experienced and for how long; etc.). Such

factors may influence children's tendency to collaborate on various classroom activities,

as well as children's development of collaborative skills, as they directly relate to

participants' exposure to other young children. However, it is important to note that the study took place late in the school year (late May through early June), so all children in the sample completed nearly a full school year together and were familiar with working with one another during various classroom activities.

Finally, the sample in this study is small, and not very diverse. Future studies will ask participants for such demographic information, as well as determine whether the results replicate with larger samples from more diverse populations. Further, the current study paired children in same-age, same-gender dyads. Future studies will also explore different ways of pairing and/or grouping children for collaborative work (i.e., same-versus mixed-gender pairs; same- versus mixed-age pairs; pairs based on friendships; pairs based on prior knowledge or experience, such as by expert versus novice status, etc.), as well as the effects of various pairing formats on children's collaborative skills and behaviors.

Chapter 5: Conclusions and Implications for Education

5.1 Conclusions and Implications for Education.

As previously discussed, there is still much to learn about what constitutes appropriate frameworks that blend science education with developmentally-appropriate learning environments. An important goal for the construction of early science is a better understanding of appropriate learning experiences and expectations for young children. The research presented in this dissertation aims to explore possible conditions of science learning in the early childhood classroom, as well as arrive at possible clues as to how teachers can help move students forward. Specifically, the dissertation focuses on three dimensions of science learning in the early childhood classroom: (1) the learner, (2) instructional tools and pedagogy, and (3) developmentally-appropriate contexts for supporting science learning (both individually, and with peers).

In terms of the learner, one goal for this dissertation was to examine some dimensions of young children's early scientific reasoning and inquiry skills in the context of potentially relevant, developing general reasoning abilities. As young children undergo rapid cognitive changes during the preschool years, it is important to explore how these cognitive changes may influence scientific thinking. Two features of cognitive functioning have been carefully studied: (1) the demonstration of an epistemic awareness through an emerging theory of mind (i.e., Flavell, 1999; Perner, Leekam & Wimmer, 1987), and (2) the rapid improvement in executive functioning capacity (i.e., De Luca & Leventer, 2008; Espy, 2004). Both continue to develop through childhood and adolescence, but changes in early childhood are especially striking and have been neglected as regards their potential role in scientific reasoning. This follows from the

general stage theory assumption that preschool children are perception bound and unable to learn and think about the abstract concepts involved in science.

One key finding from this work is that preschool-aged children's theory of mind and executive functioning skills positively correlated with their ability to recall, as well as judge the correctness of, their initial predictions during scientific investigations. This indicates that young children's acquisition of scientific thinking may relate to their developing theory of mind and executive functioning skills. An array of theoretical work hypothesizes the positive relationships among these variables (i.e., see Duschl, Schweingruber, & Shouse, 2007; Gropen, Clark-Ciarelli, Hoisington, & Ehrlich, 2011; Kuhn, 2000; Kuhn & Pearsall, 2004). Further, the few existing empirical studies exploring the relationships among these variables reveal similar findings (i.e., see Nayfeld, Fuccillo, & Greenfield, 2013; Zaitchik, Iqbal, & Carey, 2014).

This finding is important, as it can help to identify elements of young children's thinking that can add to the foundation for developing scientific reasoning skills in the early years. Although this work has not demonstrated causal relationships among these variables, the possible relationships should be explored further. If such relationships are confirmed, it is possible that educators can more effectively cultivate young children's scientific thinking skills by simultaneously supporting their emerging theory of mind and executive function skills. For example, the recording charts and physics diagrams used in the first and second studies of this dissertation represent instructional tools which effectively support children's scientific thinking skills by buttressing their emerging theory of mind and executive function abilities. The charts and diagrams aid children in attending to the critical components of their scientific investigations, while helping them

to overcome working memory load, as well as recognize discrepancies between initial predictions and observed outcomes (both of which relate to children's theory of mind and executive function skills).

Thus, a second key finding from this work is that the recording charts and relevant physics diagrams effectively supported children's scientific thinking in two different contexts for scientific investigation. This is significant, as such forms of documenting and representing ideas are central in doing science. Little prior work has explored early forms of preschool literacy related to science learning, as well as children's ability to understand symbolic representations of predictions and outcomes for scientific investigations. Also important is the finding that the diagrams' effects were especially salient among younger children (i.e., three-year-olds), as well as among children with less-well developed theory of mind and executive function skills. As follows, the diagrams may be considered developmentally-appropriate instructional scaffolds (i.e., Vygotsky, 1978) for promoting scientific thinking, especially among children who are not quite ready to engage in such forms of thinking on their own.

Finally, in terms of developmentally-appropriate contexts for supporting science learning (both individually, and with peers), the research presented here provides evidence that young children are indeed capable of engaging in basic forms of scientific thinking in the early childhood classroom. The chosen contexts for scientific investigation explored in this dissertation are suitable ones for fostering such forms of thinking in the preschool classroom. Specifically, learning activities which ask children to make and check predictions in scientific contexts in which they already have some relevant knowledge are developmentally-appropriate for cultivating such abilities in the

early childhood classroom. Infant literature illustrates that early knowledge of the physical world develops from infancy (i.e., Baillargeon, 1995). Thus, teachers should not shy away from introducing basic physical science concepts during the early years. Rather, such contexts enable educators to build upon what young children already know about the physical world.

The contexts also constitute appropriate sets of conversational options, objects and instructional tools that can encourage physical science learning in the early childhood classroom. Many science and mathematics standards for early childhood education include exploring with ramps and balance scales. Thus, the materials are ones that teachers and children are comfortable with and already often use in the early childhood classroom. This allows the findings to more easily be translated into practice (i.e., teachers can build upon the kinds of activities they already do with children in the classroom). Further, since the materials are neither too easy nor too difficult, the teacher has room to help children actively construct understanding on what they already know something about. For example, educators can support children in making focused predictions, as well as in reflecting back upon their predictions to determine whether or not the predictions were confirmed. Educators can achieve this by using diagrams and recording charts like those explored in the first two studies of this dissertation, as well as other possible forms of documentation and reflection. For example, children can be encouraged to record/document and then reflect back upon their science investigations through use of drawings, photographs, science journals, or other models/external representations of their predictions and observations. In doing so, educators can support children's use of various forms of text to foster early forms of scientific and critical

thinking, which aligns with current policy for improving early STEM education (i.e., Early Childhood STEM Working Group, 2017).

Cognitive psychology also tells us that it is important to use structurally related materials to enhance students' learning. Redundant structural examples, such as the three different (but related) physics activities used in this dissertation, can help to move students' learning forward in a constructivist manner. For example, educators can foster children's learning by encouraging them to compare and contrast different (but conceptually related) learning activities, as well as by supporting children in eliciting relevant prior knowledge that can be further built upon in new, structurally-related examples.

Another important finding is that preschool children's scientific thinking profited from opportunities to try again as they engaged in experimental manipulations. For example, the children in the first study benefitted from opportunities for repeated practice in predicting and checking. In other words, children appeared to understand when they made a mistake on the first attempt at predicting with an item, and they changed their prediction to the correct one for the second attempt. Further, the children in the third study benefitted from repeated opportunities to explore methods for increasing objects' speed down the inclined plane. Children had two opportunities to explore approaches for increasing objects' rate of speed down the ramp, and they were significantly more likely to provide the correct answer the second time the question was asked compared to the first time. This suggests that children learned from their first attempt at experimenting with ways to increase objects' speed down the ramp, and determined which approach was indeed effective, as well as which were unsuccessful.

Taken together this represents an important finding, that children's scientific thinking benefitted from opportunities for repeated experience/practice. This has a clear pedagogical implication that can easily be translated into practice, namely that young children may benefit from opportunities to reflect on scientific investigations (i.e., in ways such as the recording charts and physics diagrams used in the first study prompt them to do), as well as opportunities to try again as they engage in scientific investigations involving the exploration of different items and materials.

The final important finding from this work is that children as young as 3 ½ years are capable of working with a peer to solve simple problems in science. Science is a process, and it almost always involves collaboration with others. The Next Generation Science Standards (NGSS) are built on the key assumption that "science is fundamentally a social enterprise, and scientific knowledge advances through collaboration and in the context of a social system with well-developed norms" (NGSS Lead States, 2013, p. 27). Thus, learning with peers has been identified as a critical component of science learning and science education, with science standards highlighting the importance of providing students opportunities for collaboration, discussion and reflection around science. However, very little work explores collaboration in the context of science learning experiences in children younger than school age, making it unclear whether preschool children have skills to support their ability to work with peers on science learning activities in the early childhood classroom.

The final study presented in this dissertation reveals that preschool-aged children are capable of demonstrating a number of collaborative and communicative behaviors as

they work with a peer to solve simple physics problems focused on objects traversing an inclined plane. For example, children explained, instructed and modeled to peers how to do things as they worked on the task. They asked one another questions and/or for assistance, and responded to such requests from peers. Children also imitated one another, as well as agreed with peers' ideas or suggestions. Moreover, there was little difference between younger and older preschoolers in terms of the patterns of children's collaborative and communicative behaviors. This indicates that children as young as 3 ½ years are capable of working together on science learning activities in the preschool classroom, while also demonstrating an understanding of some basic, relevant physics principles associated with the task.

Taken together, this indicates that it can be appropriate to have young children work together on science learning activities in the early childhood classroom. Further, the chosen context is a suitable one for fostering children's acquisition of both scientific content knowledge and thinking skills, as well as early collaborative and communicative abilities. In such contexts, children can support one another as they work together to actively construct understanding on what they already know something about. As learning with peers has been identified as a critical component of science learning and science education, in doing so, educators may help to better prepare young children for learning in the later school years.

References

Anderson, R. D., & Mitchner, C. P. (1994). Research on science teacher education. In D. L. Gabel (Ed.), *Handbook of science teaching and learning* (pp. 32–37). New York: Macmillan.

Ashbrook, P. (2016). *Science learning in the early years: Activities for PreK-2*. Arlington, VA: National Science Teachers Association Press.

Au, T. K. F. (1994). Developing an intuitive understanding of substance kinds. *Cognitive Psychology*, *27*(1), 71-111.

Baillargeon, R. (1995). Physical reasoning in infancy. In M. S. Gazzaniga (Ed.-in-chief), *The Cognitive Neurosciences* (pp. 181-204). Cambridge, MA: The MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*(1), 37-46.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78(2)*, 647-663.

Bowman, B. T. (1999). Policy implications for math, science and technology in early childhood education. In *Dialogue on Early Childhood Science, Mathematics, and Technology Education* (pp. 40-49). Washington, DC: Project 2061, American Association for the Advancement of Science.

Bowman, B. T., Donovan, M. S., & Burns, M. S. (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: The National Academies Press.

Brenneman, K. (2011). Assessment for preschool science learning and learning environments. *Early Childhood Research & Practice*, *13*(1), n1.

Brenneman, K. (2014). Science in the Early Years. The Progress of Education Reform. Volume 15, Number 2. *Education Commission of the States*.

Brenneman, K., & Lauro, I. (2008). Science journals in the preschool classroom. *Early Childhood Education Journal, 36(2),* 113-119.

Brenneman, K., Stevenson-Boyd, J.S., & Frede, E. (2009). *Early mathematics and science: Preschool policy and practice (*Preschool Policy Brief No. 19). New Brunswick, NJ: National Institute for Early Education Research.

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist, 52(4),* 399-413.

Brown, A. L., & Campione, J. C. (1994). *Guided discovery in a community of learners.* Cambridge, MA: The MIT Press.

Brown, A. L., & Campione, J. C. (1990). Interactive learning environments and the teaching of science and mathematics. In M. Gardner, J. Greeno, F. Reif, A. Schoenfeld, A. diSessa, & E. Stage (Eds.), *Toward a scientific practice of science education* (pp. 111–140). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Brown, E. T. (2005). The influence of teachers' efficacy and beliefs regarding mathematics instruction in the early childhood classroom. *Journal of Early Childhood Teacher Education, 26*(26), 239-257.

Bryant, P. E. & Trabasso, T. (1971). Transitive inference and memory in young children. *Nature, 232,* 456-458.

Callanan, M. A., & Oakes, L. M. (1992). Preschoolers' questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development*, *7*(2), 213-233.

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616.

Chalufour, I. & Worth, K. (2006). Science in kindergarten. In D.F. Gullo (Ed.), *K Today: Teaching and learning in the kindergarten year* (pp. 95-106). Washington, DC: National Association for the Education of Young Children.

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

Cooper, C. R. (1980). Development of collaborative problem solving among preschool children. *Developmental Psychology*, *16*(5), 433-440.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078.

De Lisi, R., & Golbeck, S. L. (1999). Implications of Piagetian theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

De Luca, C. R., & Leventer, R. J. (2008). Developmental trajectories of executive functions across the lifespan. In V. Anderson, R. Jacobs, & P. Anderson (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 281-304). New York: Psychology Press.

Driver, R. (1989). Students' conceptions and the learning of science. *International Journal of Science Education, 11(5),* 481-490.

Druyan, S. (2001). A comparison of four types of cognitive conflict and their effect on cognitive development. *International Journal of Behavioural Development, 3,* 226-236.

Duschl, R. A. (2012). The second dimension- crosscutting concepts: Understanding a framework for K-12 science education. *The Science Teacher*, *9(2)*, 34-38.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: National Academies Press.

Early Childhood STEM Working Group. (2017). *Early STEM Matters: Providing High Quality STEM Experiences for All Young Learners.* A Policy Report by the Early Childhood STEM Working Group. Chicago: U Chicago STEM Education.

Espy, K. A. (2004). Using developmental, cognitive, and neuroscience approaches to understand executive control in young children. *Developmental Neuropsychology*, *26*(1), 379-384.

Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children' problem-solving ability. *British Journal of Educational Psychology, 75(*2), 157-169.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, *50*(1), 21-45.

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly, 19(1),* 138-149.

French, L., Conezio, K., & Boynton, M. (2000). *Using Science as the Hub of an Integrated Early Childhood Curriculum: The Science Start![TM] Curriculum*. Arlington, VA: National Science Foundation.

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483-27.

Garton, A. E. (1992). *Social interaction and the development of language and cognition*. East Sussex: Erlbaum.

Garton, A. E., & Pratt, C. (2001). Peer assistance in children's problem solving. *British Journal of Developmental Psychology, 19,* 307-318.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly (19),* 150-158.

Gelman, R., & Brenneman, K. (2012). Moving young "scientists-in-waiting" onto science learning pathways: Focus on observation. In J. Shrager, S. Carver (Eds.), *The journey from child to scientist: Integrating cognitive development and the education sciences* (pp. 155-169). Washington, DC US: American Psychological Association.

Gelman, R., Brenneman, K., Macdonald, G., & Roman, M. (2009). *Preschool pathways to science (PrePS): Facilitating scientific ways of thinking, talking, doing and understanding.* Baltimore, MD: Brookes Publishing Co.

Gelman, R., Meck, G., Romo, L., Meck B., & Fritz, C. (1995). Integrating science concepts into intermediate English as a second language (ESL) instruction. In R. F. Macias, & R. G. Gracia Ramos (Eds.), *Changing schools for changing students: An anthology of research on language minorities, schools & society* (pp. 181-203). Santa Barbara, CA: UC Linguistic Minority Research Institute.

Gelman, R., & Williams, E. M. (1998). Enabling constraints for cognition development and learning: Domain specificity and epigenesis. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (5th Ed., pp. 575-630). New York: Wiley.

Gilbert, J. K., & Watts, D. M. (1983). Concepts, misconceptions and alternative conceptions: Changing perspectives in science education. *Studies in Science Education, 10,* 61-98.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance reality distinction. *Child Development*, 26-37.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development,* 1205-1222.

Greenfield, D. B., Jirout, J., Dominguez, X., Greenberg, A., Maier, M., & Fuccillo, J. (2009). Science in the preschool classroom: A programmatic research agenda to improve science readiness. *Early Education and Development*, *20*(2), 238-264.

Gropen, J., Clark-Chiarelli, N., Ehrlich, S., & Thieu, Y. (2011). Examining the efficacy of foundations of science literacy: Exploring contextual factors. *American Law and Economics Review, 8(2),* 249-281.

Gropen, J., Clark-Chiarelli, N., Hoisington, C., & Ehrlich, S. B. (2011). The importance of executive function in early science education. *Child Development Perspectives*, *5*(4), 298-304.

Halford, G. S., Andrews, G., Dalton, C., Boag, C., & Zielinski, T. (2002). Young children's performance on the balance scale: The influence of relational complexity. *Journal of Experimental Child Psychology*, *81*(4), 417-445.

Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. *Cognition, 50*, 171-188.

Hogan, D., & Tudge, J. (1999). Implications of Vygotsky's theory for peer learning. In A. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39-65). New Jersey: Erlbaum Press.

Howitt, C., Upson, E., & Lewis, S. (2011). 'It's a mystery!': A case study of implementing forensic science in preschool as scientific inquiry. *Australasian Journal of Early Childhood*, *36*(3), 45.

Hughes, C., & Graham, A. (2002). Measuring executive functions in childhood: Problems and solutions? *Child and Adolescent Mental Health*, *7*(3), 131-142.

Jansen, B. R., & van der Maas, H. L. (2002). The development of children's rule use on the balance scale task. *Journal of Experimental Child Psychology*, *81*(4), 383-416.

Johnson, D. W., & Johnson, R. T. (1994). Collaborative learning and argumentation. In P. Kutnick & C. Rogers (Eds.), *Groups in schools* (pp. 66-86). London: Cassell.

Johnson-Pynn, J. S., & Nisbet, V. S. (2002). Preschoolers effectively tutor novice classmates in a block construction task. *Child Study Journal*, *32*(4), 241-55.

Katz, L. (1998). What can we learn from Reggio Emilia? In C. Edwards, L. Gandini, & G. Forman (Eds.), *The hundred languages of children* (pp. 19–37). Norwood, NJ: Ablex.

Klahr, D., and Z. Chen. 2003. Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development, 74*(5), 1275-1296.

Kliman, M. (1987). Children's learning about the balance scale. *Instructional Science, 15*(1), 307-340.

Kloos, H., & Somerville, S. C. (2001). Providing impetus for conceptual change: The effect of organizing the input. *Cognitive Development*, *16*(2), 737-759.

Korkman, M. (2000). Executive functions in children: Assessment, development, and disorders. *Revue de Neuropsychologie, 10*(3), 471- 487.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science*, *9*(5), 178-181.

Kuhn, D., & Pearsall, S. (2000). Developmental origins of scientific thinking. *Journal of Cognition and Development, 1,* 113-129.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65-100.

Liberti-Reuter, J. (2016). *Simple physics diagrams of an inclined plane support preschoolers' scientific predicting and checking.* Unpublished manuscript.

Light, P. L., & Littleton, K. (1994). Cognitive approaches to group work. In P. Kutnick & C. Rogers (Eds.), *Groups in schools* (pp. 87-103). London: Cassell.

McClure, E. R., Guernsey, L., Clements, D. H., Bales, S. N., Nichols, J., Kendall-Taylor, N., & Levine, M. H. (2017). *STEM starts early: Grounding science, technology, engineering, and math education in early childhood*. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Michaels, S., Shouse, A., & Schweingruber, H. (2008). *Ready, set, science!: Putting research to work in K-8 science classrooms.* Washington, DC: The National Academies Press.

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

National Research Council. (2001). *Eager to learn: Educating our preschoolers.* Committee on Early Childhood Pedagogy. In B. T. Bowman, M. S. Donovan and M. S. Burns (Eds.), Commission on Behavioral and Social Sciences and Education. Washington, DC.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, DC: National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Washington, DC: National Academy Press.

National Research Council. (2015). *Transforming the workforce for children birth through age 8: a unifying foundation*. Washington, DC: National Academies Press.

National Science Teacher Association (NSTA). (2014*). NSTA position statement: Early childhood science education.* Arlington, VA: National Science Teachers Association.

Nayfeld, I., Brenneman, K., & Gelman, R. (2011). Science in the classroom: Finding a balance between autonomous exploration and teacher-led instruction in preschool settings. *Early Education & Development*, *22*(6), 970-988.

Nayfeld, I., Fuccillo, J., & Greenfield, D.B. (2013). Executive functions in early learning: Extending the relationship between executive functions and school readiness to science. *Learning and Individual Differences, 26,* 81-88.

Next Generation Science Standards (NGSS) Lead States. (2013). *Next generation science standards: For states, by states*. Washington, D.C.: The National Academies Press.

New Jersey Department of Education (2014). *Preschool teaching and learning standards.* Trenton, NJ: New Jersey Department of Education.

O'Donnell, A. M., & King, A. (Eds.). (1999). *Cognitive perspectives on peer learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125-137.

Phelps, E., & Damon, W. (1989). Problem solving with equals: Peer collaboration as a context for learning mathematics and spatial concepts. *Journal of Educational Psychology, 81(4),* 639-646.

Phillips, D. C. (1995). The good, the bad and the ugly: The many faces of constructivism. *Educational Researcher, 24*(7), 5-12.

Piaget, J. (1959). *The language and thought of the child* (3rd ed.). London: Routledge and Kegan Paul.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development.* Chicago, IL: University of Chicago Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515-526.

Quinn, M. E., Taylor, K., & Taylor, S. (2004). *Let's talk science. Wings of discovery: A framework for early learning* (2nd ed.). Toronto, Ontario, Canada: GTK Press.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.

Rogoff, B., Turkanis, C. G., & Bartlett, L. (Eds.). (2001). *Learning together: Children and adults in a school community.* Oxford: Oxford University Press.

Rutherford, F. J., & Ahlgren, A. (1989). *Science for all Americans: A Project 2061 report on literacy goals in science, mathematics, and technology.* Washington, DC: American Association for the Advancement of Science.

Ryan, S., Whitebook, M., & Cassidy, D. (2014). *Strengthening the math-related teaching practices of the early care and education workforce: Insights from experts*. Berkeley, CA: Center for the Study of Child Care Employment, University of California, Berkeley.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The Journal of the Learning Sciences*, *3*(3), 265-283.

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology*, *40*(2), 162.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development, 38*(5)*,* 1-38.

Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, *36*(3), 273-310.

Slavin, R. E. (1992). When and why does cooperative learning increase achievement/theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145-173). New York: Cambridge University Press.

Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M., & Breinlinger, K. (1994). Early knowledge of object motion: Continuity and inertia. *Cognition*, *51*(2), 131-176.

Tomasello, M., & Hamann, K. (2012). Collaboration in young children. *The Quarterly Journal of Experimental Psychology, 65*(1), 1-12.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495-511.

Tu, T. (2006). Preschool science environment: What is available in a preschool classroom? *Early Childhood Education Journal, 33*(4), 245-251.

Tudge, J. R. H. (1992). Processes and consequences of peer collaboration: A Vygotskian analysis. *Child Development, 63,* 1364-1379.

Tudge, J. R. H., & Winterhoff, P. A. (1993). Vygotsky, Piaget and Bandura: Perspectives on the relations between the social world and cognitive development. *Human Development, 36,* 61-81.

Tudge, J. R. H., Winterhoff, P. A., & Hogan, D. M. (1996). The cognitive consequences of collaborative problem solving with and without feedback. *Child Development, 67,* 2892-2909.

Underwood, J., Underwood, G., & Wood, D. (2000). When does gender matter? Interactions during computer-based problem solving. *Learning and Instruction, 10(5),* 447-462.

Vitiello, E. V. (2009). EF and approaches to learning: Relationships to school readiness in Head Start preschoolers (Doctoral dissertation, University of Miami). Retrieved from http://etd.library.miami.edu/theses/available/etd-07132009-155459/

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

Wandersee, J. H., Mintzes, J. J. & Novak, J. D. (1994). Research on alternative conceptions in science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177-210). New York: Macmillan.

Webb, N. M., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 117-149). Mahwah, NJ: Lawrence Erlbaum Associates.

Wellman, H. M., & Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In D. Kuhn & R. Siegler (Vol. Eds.), N. Eisenberg (Series Ed.), *Handbook of Child Psychology: Cognition, Perception and Language* (5th Ed., pp. 523-573). New York: Wiley.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523-541.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.

Worth, K. (2010). Science in early childhood classrooms: Content and process. Early Childhood Research and Practice, Collected Papers from the SEED (STEM in Early Education and Development) Conference. Retrieved from http://ecrp.uiuc.edu/beyond/seed/worth.html

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: an individual differences study. *Child Development, 85*(1), 160-175.

Appendix A

Scripts and Procedure for Dimensional Change Card Sort Pre-test (Used for Studies 1 & 2)

Introduction Phase I (Shape)- Show 2 cards, a red boat and a blue rabbit: "Here's a red boat, and here's a blue rabbit. We are going to play a game. This is the shape game. All the boats go in this box, and all the rabbits go in that box. We don't put any boats in that box. We put all the boats over here and only rabbits go over there. This is the shape game."

Experimenter then sorts 2 cards interactively, explaining the basis of her behavior to child. "Okay, now I'm going to show you some boats and some rabbits." Show child 5 cards in a random sequence (no more than 2 of the same in a row). "If it's a boat, then it goes here; if it's a rabbit, then it goes there. This is a (i.e., red boat). Where does this go?" The child is given feedback (i.e., "That's right, the boats go here. Or, "No, that's not right, you have to put all the boats here; that's how you play the shape game). The child is asked to place the cards into one of the boxes, or point to a box.

Introduction Phase II (Color)- "Now we're going to play a new game." Show 2 cards, a red boat and a blue rabbit: "Here's a red boat, and here's a blue rabbit. We are going to play a game. This is the color game. The color game is different from the shape game. All the red ones go in this box, and all the blue ones go in that box. We don't put any red ones in that box. We put all the red ones over here and only blue ones go over there. This is the color game."

Experimenter then sorts 2 cards interactively, explaining the basis of her behavior to child. "Okay, now I'm going to show you some red ones and some blue ones." Show child 5 cards drawn in a random sequence (no more than 2 of the same in a row). "If it's a red one, then it goes here; if it's a blue one, then it goes there. This is a (i.e., red boat). Where does this go?" The child is given feedback (i.e., "That's right, the blue ones go here." Or, "No, that's not right, you have to put all the blue ones here; that's how you play the color game). The child is asked to place the cards into one of the boxes, or point to a box.

Test Phase I- "We are going to play the color game." For each of 5 test cards, "If it's a red one, then it goes here; if it's a blue one, then it goes there. This is a (i.e., red boat); where does this go?" Feedback is not be given. After child makes a response, experimenter says, "Okay," and proceeds to the next card.

Test Phase 2- "Okay, now we're going to switch and play a different game, the shape game. You have to pay attention." Five test cards are presented exactly as in the Test Phase I. The child is told the rule on each trial. "If it's a boat, then it goes here; if it's a rabbit, then it goes there. This is a (i.e., red boat); where does this go?" Feedback is not be given. After the child makes a response, experimenter says, "Okay" and proceeds to the next card.

Test Phase 3- Four test trials administered. On each trial, including the first, the child is required to switch to the other rule. For each switch, "Okay, now we're going to switch

again and play a different game, the (e.g., shape) game. You have to pay attention." The

experimenter states the relevant rules and labels the cards for each trial (i.e., "Okay, now

we're going to switch again and play a different game, the color game. You have to pay

attention. If it's a red one, then it goes here, if it's a blue one, then it goes here. This is a

(i.e., red boat); where does this go? Feedback is not given. After child makes a response,

experimenter says, "Okay" and moves on to the next card.

Appendix B

Scripts and Procedure for the Theory of Mind Pre-test (Used for Studies 1 & 2)

Experimenter presents Crayon Box to child (box is closed, and child does not touch).

"What is this? What's in it?"

Experimenter opens the box and shows child the contents of the box (Band-aids™).

"What's in it?" If no answer, "It has Band-aids™ in it."

Experimenter closes box and moves out of child's reach.

1. **False Belief Question: "**(Mary/John; use classmate's name) hasn't played this game yet. S/he hasn't seen this (box) before, and s/he hasn't looked inside. If s/he comes into the room right now, what will s/he think is in it, Band-Aids™ or crayons?"

2. **Representational Change Question**: "When you first saw this (box), before you had a chance to look inside, what did you think was in it, Band-Aids™ or crayons?"

3. (Counterbalanced order):

(a). **Reality Question**: "What's really and truly inside the box, Band-Aids™ or crayons?"

(b). **Appearance Question**: "What does it look like it has in it, Band-Aids™ or crayons?

Appendix C

Scripts and Procedure for the Ramp Predict and Check Activity (Used for Study 1)

Diagram Condition

<u>**Introduction**</u>: "We are going to play a game. I brought this ramp and some toys (TinkerToy™ objects). For the game, you will predict, or guess, what will happen if we try different things on the ramp. A prediction is like a guess. You will predict, or guess, whether each will roll down the ramp, or slide down. Let me show you.

<u>**Demonstration Items**</u>: (Randomly choose one of the two demo items [ball or block]). Note: for the first item, the experimenter demonstrates the procedure while making a correct prediction. In other words, she correctly predicts whether the item will roll or slide down the ramp. For the second item, the experimenter demonstrates the procedure while making an incorrect prediction. In other words, she incorrectly predicts whether the item will roll or slide down.

<u>First Demo Item (ball)-</u> This is a ball. If I put it on the ramp like this (demonstrate by setting ball on top of ramp, but without releasing it), I predict, or guess, that it will roll down. My guess is that the ball will roll down the ramp. Now I am going to put my prediction, or guess, on this chart. See the two pictures? Which one shows rolling down the ramp? (Wait for child to point. If incorrect, show him/her the correct photo). Yes, this picture shows rolling down the ramp. Now, which picture shows sliding down the ramp? (Wait for child to point. If incorrect, show him/her the correct photo). Yes, this picture shows sliding down the ramp. I predicted, or guessed, that the ball would roll

down the ramp, so I will put this picture (diagram for rolling) on the chart. This shows my guess; see (point to diagram)?

(After marking the prediction on chart) Now, let's try it on the ramp and see what happens. (Release object at the top of the ramp). What happened; did it roll down or slide (counterbalanced)? It rolled down. Let's put it on the chart. Look at the two pictures again. Which shows what happened when we tried it on the ramp? (Wait for child to point. If incorrect, show correct response). This one shows rolling down. Let's put it on the chart. This will help us remember what really happened when we tried it on the ramp.

After stating (and recording) the observed outcome: Before we tried it on the ramp, what was my prediction, or guess? Did I guess that it would roll down, or slide down? The chart can help me remember. Here is my guess (point to prediction diagram on chart). I predicted, or guessed, that it would roll; see? Then we tried it on the ramp. Here's what really happened (point to observation diagram on chart); it rolled down the ramp; see? So was my prediction right, or do I need to change it? I predicted, or guessed, that it would roll down, and it really did roll down, so my prediction was right. I don't need to change it.

Second Demo Item (block)- The same script as above is followed for the second item, but the experimenter makes an incorrect prediction. At the end, she explains that her prediction was wrong, so it needs to be changed. She also states that it's okay if the

prediction is wrong, because it can be changed (i.e., she demonstrates how to remove the incorrect diagram from the chart and replace it with the correct one).

<u>Ramp Task Script</u>:

Now it's your turn to try. Here, look at this (TinkerToy™ item); (child is able to hold and explore the item briefly. Experimenter then takes object and positions it at the top of the ramp). First, make a prediction, or guess. If I let go, will it roll down, or slide down (counterbalanced)? What is your guess? (Wait for child to respond). Okay. Now let's put your guess on the chart. Which picture shows (rolling/sliding for prediction)? (Wait for child to respond). Okay, put it (diagram) here (on the chart, under the word *prediction*).

Now it's time to try it on the ramp. Let's see what happens. (Experimenter helps child release object on ramp). What happened? Did it roll down or slide down (counterbalanced)? (Wait for child to respond). Okay; let's put that on the chart. Which picture shows (rolling/sliding for outcome). Okay; put it (diagram) here (on the chart, under the word *observation*).

Before we tried it on the ramp, what was your guess, or prediction? Did you guess that it would roll or slide down (counterbalanced)? You can look at the chart to help you remember. (Wait for child to respond).

What really happened; did it roll down, or slide down? You can look at the chart to help you remember. (Wait for child to respond).

So was your prediction right, or do you need to change it? You can look at the chart to help you remember. (Wait for child to respond).

The same process is carried out for each object, twice in a row.

Control Condition

<u>**Introduction**</u>: (Same as above).

<u>**Demonstration Items**</u>: (Same as above, but without use of the recording chart).

(Randomly choose one of the two demo items [ball or block]). Note: for the first item, the experimenter demonstrates the procedure while making a correct prediction. In other words, she correctly predicts whether the item will roll or slide down the ramp. For the second item, the experimenter demonstrates the procedure while making an incorrect prediction. In other words, she incorrectly predicts whether the item will roll or slide down.

<u>First Demo Item (ball)</u>- This is a ball. If I put it on the ramp like this (demonstrate by setting ball on top of ramp, but without releasing it), I predict, or guess, that it will roll down. My guess is that the ball will roll down the ramp.

Now, let's try it on the ramp and see what happens. (Release object at the top of the ramp). What happened; did it roll down or slide (counterbalanced)? It rolled down.

Before we tried it on the ramp, what was my guess? Did I guess that it would roll down, or slide down? I predicted that it would roll. Then we tried it on the ramp. What really happened? It rolled down the ramp. So was my prediction right, or do I need to change it? I predicted, or guessed, that it would roll down, and it really did roll down, so my prediction was right. I don't need to change it.

Second Demo Item (block)- The same script as above is followed for the second item, but the experimenter makes an incorrect prediction. At the end, she explains that her prediction was wrong, so it needs to be changed. She also states that it's okay if the prediction is wrong, because it can be changed at the end.

Ramp Task Script: (Same as above, but without use of the recording chart).

Now it's your turn to try. Here, look at this (TinkerToy™ item); (child is able to hold and explore the item briefly. Experimenter then takes object and positions it at the top of the ramp). First, make a prediction, or guess. If I let go, will it roll down, or slide down (counterbalanced)? What is your guess? (Wait for child to respond). Okay.

Now it's time to try it on the ramp. Let's see what happens. (Experimenter helps child release object on ramp). What happened? Did it roll down or slide down (counterbalanced)? (Wait for child to respond). Okay.

Before we tried it on the ramp, what was your guess, or prediction? Did you guess that it would roll or slide down (counterbalanced)? (Wait for child to respond).

And what really happened; did it roll down, or slide down? (Wait for child to respond).

So was your prediction right, or do you need to change it? (Wait for child to respond).

The same process is carried out for each object, twice in a row.

Appendix D

Scripts and Procedure for the Balance Scale Predict and Check Activity (Used for Study 2)

Diagram Condition

<u>**Introduction**</u>: We are going to play a game. I brought this balance scale and some other things with me. For the game, you will predict, or guess, what will happen if we try putting different things on each side of the scale. A prediction is like a guess.   You will predict, or guess, whether the scale will balance, like this (demonstrate with scale), or if one side will go down, like this (demonstrate with scale).   Let me show you.

<u>**Demonstration Items**</u>: (Randomly choose one of the two demo items [ball or block]). Note: for the first item, the experimenter demonstrates the procedure while making a correct prediction. In other words, she correctly predicts whether the scale will balance. For the second item, the experimenter demonstrates the procedure while making an incorrect prediction. In other words, she incorrectly predicts whether the scale will balance.

<u>First Demonstration Set (uninflated latex balloon; golf ball)-</u> I have a ball, and a balloon. If I put them on the scale, like this (demonstrate by holding one item above each bucket on the scale, but without releasing them), I predict, or guess, that one side will go down. My guess is that this side will go down, and the scale will look like this (demonstrate "uneven" with hands). Now I am going to put my prediction, or guess, on this chart. See the two pictures? Which one shows one side going down? (Wait for child to point. If incorrect, show him/her the correct photo). Yes, this picture shows one side going down.

Now, which picture shows the scale balanced, or even? (Wait for child to point. If incorrect, show him/her the correct photo). Yes, this picture shows the scale balanced, or even. I predicted, or guessed, that one side of the scale would go down, so I will put this picture (diagram for unbalanced scale) on the chart. This shows my guess; see (point to diagram)?

(After marking the prediction on chart) Now, let's try it on the scale and see what happens. (Place one item on each side of the scale). What happened; is it balanced, or did one side go down (counterbalanced)? One side went down. Let's put it on the chart. Look at the two pictures again. Which shows what happened when we tried it on the ramp? (Wait for child to point. If incorrect, show correct response). This one shows one side going down. Let's put it on the chart. This will help us remember what really happened when we tried it on the scale.

After stating (and recording) the observed outcome: Before we tried it on the scale, what was my prediction, or guess? Did I guess that the scale would balance, or one side would go down? The chart can help me remember. Here is my guess (point to prediction diagram on chart). I predicted, or guessed, that one side would go down; see? Then we tried it on the scale. Here's what really happened (point to observation diagram on chart); one side went down; see? So was my prediction right, or do I need to change it? I predicted, or guessed, that one side would go down, and it really did go down, so my prediction was right. I don't need to change it.

Second Demonstration Set (two identical plastic counting bears)- The same script as above is followed for the second pair of items, but the experimenter makes an incorrect prediction. At the end, she explains that her prediction was wrong, so it needs to be changed. She also states that it is okay if the prediction is wrong, because it can be changed (i.e., she demonstrates how to remove the incorrect diagram from the chart and replace it with the correct one).

Balance Scale Task Script:

Now it's your turn to try. Here, look at these (experimenter hands the child two items, and the child is able to hold and explore them briefly). First, make a prediction, or guess. If you put them on the scale, will it balance like this (demonstrate with hands), or will one side go down, like this (demonstrate with hands)? (Order is counterbalanced). What is your guess? (Wait for child to respond). Okay. Now let's put your guess on the chart. Which picture shows (balanced/unbalanced for prediction)? (Wait for child to respond). Okay, put it (diagram) here (on the chart, under the word *prediction*).

Now it's time to try it on the scale. Let's see what happens. (Experimenter helps child place one item on each side of the scale). What happened? Did it balance, or did one side go down (counterbalanced)? (Wait for child to respond). Okay; let's put that on the chart. Which picture shows (balanced/unbalanced for outcome). Okay; put it (diagram) here (on the chart, under the word *observation*).

Before we tried it on the scale, what was your guess, or prediction? Did you guess the scale would balance, or that one side would go down (counterbalanced)? You can look at the chart to help you remember. (Wait for child to respond).

What really happened; did the scale balance, or did one side go down? You can look at the chart to help you remember. (Wait for child to respond).

So was your prediction right, or do you need to change it? You can look at the chart to help you remember. (Wait for child to respond).

Let's try again with some new things. (The same process is carried out for the remaining pairs of items).

Control Condition

<u>**Introduction**</u>: (Same as above).

<u>**Demonstration Items**</u>: (Same as above, but without use of the recording chart).

(Randomly choose one of the two demonstration item sets above). Note: for the first item pair, the experimenter demonstrates the procedure while making a correct prediction. In other words, she correctly predicts whether placing the items on the scale will result in the scale balancing, or one side going down. For the second pair of items, the experimenter demonstrates the procedure while making an incorrect prediction. In other words, she incorrectly predicts whether the scale will balance, or if one side will go down.

First Demonstration Set (uninflated latex balloon; golf ball)- I have a ball, and a balloon.

If I put them on the scale, like this (demonstrate by holding one item above each bucket

on the scale, but without releasing them), I predict, or guess, that one side will go down.

My guess is that this side will go down, and the scale will look like this (demonstrate

"uneven" with hands).

Now, let's try it on the scale and see what happens. (Place one item on each side of the

scale). What happened; is it balanced, or did one side go down (counterbalanced)? One

side went down.

After stating the observed outcome: Before we tried it on the scale, what was my

prediction, or guess? Did I guess that the scale would balance, or one side would go

down? I predicted, or guessed, that one side would go down. Then we tried it on the

scale. What really happened? One side went down. So was my prediction right, or do I

need to change it? I predicted, or guessed, that one side would go down, and it really did

go down, so my prediction was right. I don't need to change it.

Second Demonstration Set (two identical plastic counting bears)- The same script as

above is followed for the second pair of items, but the experimenter makes an incorrect

prediction. At the end, she explains that her prediction was wrong, so it needs to be

changed. She also states that it is okay if the prediction is wrong, because it can be

changed.

Balance Scale Task Script: (Same as above, but without use of the recording chart).

Now it's your turn to try. Here, look at these (experimenter hands the child two items,

and the child is able to hold and explore them briefly). If you put them on the scale, will

it balance like this (demonstrate with hands), or will one side go down, like this

(demonstrate with hands)? (Order is counterbalanced). What is your guess? (Wait for

child to respond). Okay.

Now it's time to try it on the scale. Let's see what happens. (Experimenter helps child

place one item on each side of the scale). What happened? Did it balance, or did one side

go down (counterbalanced)? (Wait for child to respond). Okay.

Before we tried it on the scale, what was your guess, or prediction? Did you guess the

scale would balance, or that one side would go down (counterbalanced)? (Wait for child

to respond).

What really happened; did the scale balance, or did one side go down? (Wait for child to

respond).

So was your prediction right, or do you need to change it? (Wait for child to respond).

Let's try again with new things. (The same process is carried out for the remaining pairs

of items).

Appendix E

Scripts and Procedure for the Collaborative Science Activity (Used for Study 3)

Introduction: We read this book together the other day. (Show book, *Swimmy,* by Leo Lionni). What was it about? It was about collaboration. What does it mean to collaborate? What are some things that you need to do when you are collaborating, or working on something together? Today we are going to play a game. For this game, it is really important that you collaborate, or work together. Can you do that?

Construct Object 1: I brought Tinker Toys™ and this ramp with me. The ramp can move up and down like this, see? (Demonstrate how the angle of the incline can be increased or decreased by moving the ramp board up and down). Your job is to work together to build something that can *roll* down the ramp. (*Roll* and *slide* were counterbalanced). You don't want it to slide down, you want it to roll. You can use any of the pieces, but you need to work together as a team. Let me know when it is finished, and then we will test it on the ramp.

Test Object 1 on Ramp: Are you ready to try it on the ramp? Let's find out what happens. Did it roll? Why/Why not? (If the object does not move as expected, allow children to revise their design until they successfully create an object that rolls).

Increase Speed Question: How can we make it roll down faster? (Allow children to test their ideas and predictions. If neither child suggests increasing the angle of the incline on the ramp, ask, "What would happen if I move the ramp up, like this?").

Construct Object 2: Are you ready for your next job? This time you need to work together to build something that can *slide* down the ramp. You don't want it to roll down, you want it to slide. Remember that you must collaborate, or work together to build. Let me know when you're finished and then we will test it on the ramp.

Test Object 2 on Ramp: Are you ready to try it on the ramp? Let's find out what happens. Did it slide? Why/ why not? (If the object does not move as expected, allow children to revise their design until they successfully create an object that slides).

Increase Speed Question: How can we make it slide down faster? (Allow children to test their ideas and predictions. If neither child suggests increasing the angle of the incline, ask "What would happen if I move the ramp up, like this?").

Comparing Speed Question: Are you ready for my last question? You worked together to make two objects. One rolled down the ramp, and one slid down. Which one went faster down the ramp? The one that rolled down, or the one that slid (counterbalanced). Why?

Footnotes

[1] Analyses for the Studies 1 and 2 were carried out using parametric statistics, as well as the equivalent non-parametric tests. Results were nearly identical. However, since distributions for most of the data violated assumptions of normality required for most parametric tests, we decided to use the nonparametric equivalents throughout.

[2] Error bars on all graphs represent one Standard Error of the mean.

[3] In Study 3, the analyses exploring group differences in mean frequencies for children's collaborative behaviors (i.e., differences in the mean number of coding categories applied per dyad) were conducted both on mean frequencies, as well as mean frequencies controlling for time spent on the task. Results were nearly identical. When controlling for time, the only different finding was that older dyads were also found to engage in significantly more explaining, instructing and modeling (E/I/M) to peers compared to younger dyads, $U = 18.00$, $p = .047$. All other results were the same. Since analyses found no group differences in the amount of time dyads spent engaged on the task, we decided to use the analyses which did not control for differences in time.