

STANDARDS-BASED PRACTICES AND MATHEMATICS ACHIEVEMENT:
A HIERARCHICAL LINEAR MODELING ANALYSIS

By

JACK ROBBINS DEMPSEY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

© 2010 Jack Robbins Dempsey

To my parents

ACKNOWLEDGMENTS

I thank my mother for providing constant support and encouragement in all of my endeavors. I also thank my wife, Allison, for helping me to maintain a positive outlook and keeping me grounded. I am also grateful to Dr. Richmond Thompson for originally setting my feet on this path and to Jason Gallant, Bill Cumby, Allan Daniels, and my wife for helping me to laugh as I walked it.

Finally, I thank my exception doctoral committee for all of their efforts on my behalf. In particular, I wish to thank my co-chairs, Dr. John Kranzler and Dr. Stephen Pape, for their commitment to developing my skills in conceptualizing and presenting research.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	7
LIST OF FIGURES.....	8
ABSTRACT	9
CHAPTER	
1 REVIEW OF THE LITERATURE	11
Mathematics Education in the U.S. and Abroad	13
Calls for Reform.....	17
Standards-based Instructional Practices.....	20
Research on Standards-based Instructional Practices	22
Alternatives to Qualitative Research	34
Experimental Studies of Standards-based Instructional Practices	38
The Present Study	49
The Purpose of the Present Study.....	52
2 METHODS.....	53
Overview.....	53
Data Source.....	53
Teacher Questionnaire	55
School Administrator Questionnaire	55
Direct Child Assessment	55
Measurement of mathematics achievement.....	56
Score format	58
Statistical Analyses.....	59
Analytic Samples.....	59
Assessments and Measures	61
Criterion measure	61
Control variables.....	61
Instructional scales	63
Hierarchical Linear Modeling	69
Sample Weights	72
3 RESULTS	78
Kindergarten	78
First Grade.....	81
Third Grade.....	83

Fifth Grade.....	86
4 DISCUSSION	93
Third-and Fifth-Grade Findings.....	96
Delayed Effects of Standards-based Instructional Practices.....	98
Limitations.....	98
Future Studies	100
Conclusions	101
 APPENDIX	
A MATHEMATICS INSTRUCTIONAL ACTIVITY ITEMS FROM TEACHER QUESTIONNAIRES.....	102
B HIERARCHICAL LINEAR MODELING EQUATIONS (KINDERGARTEN)	106
C HIERARCHICAL LINEAR MODELING EQUATIONS (FIRST GRADE).....	110
D HIERARCHICAL LINEAR MODELING EQUATIONS (THIRD GRADE)	117
E HIERARCHICAL LINEAR MODELING EQUATIONS (FIFTH GRADE).....	124
LIST OF REFERENCES	126
BIOGRAPHICAL SKETCH.....	133

LIST OF TABLES

<u>Table</u>	<u>page</u>
2-1	Item means, standard deviations and correlations: Kindergarten sample 74
2-2	Item means, standard deviations and correlations: First-grade sample..... 74
2-3	Item means, standard deviations and correlations: Third-grade sample 75
2-4	Item means, standard deviations and correlations: Fifth-grade sample..... 75
2-5	Goodness of fit indices for exploratory factor analyses 76
2-6	Factor loadings for the two-factor solution at the kindergarten level..... 76
2-7	Factor loadings for the two-factor solution at the first-grade level..... 76
2-8	Factor loadings for the one-factor solution at the third-grade level..... 77
2-9	Factor loadings for one-factor solution at the fifth-grade level 77
3-1	Kindergarten hierarchical linear modeling fixed effects 87
3-2	Kindergarten hierarchical linear modeling random effects..... 88
3-3	First grade hierarchical linear modeling fixed effects..... 89
3-4	First grade hierarchical linear modeling random effects 90
3-5	Third grade hierarchical linear modeling fixed effects..... 91
3-6	Third grade hierarchical linear modeling random effects..... 92
3-7	Fifth grade hierarchical linear modeling fixed effects..... 92
3-8	Fifth grade hierarchical linear modeling random effects 92

LIST OF FIGURES

<u>Figure</u>		<u>page</u>
A-1	Mathematics instructional activities items at kindergarten	102
A-2	Mathematics instructional activities items at first grade	103
A-3	Mathematics instructional activities items at third grade	104
A-4	Mathematics instructional activities items at fifth grade	105

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

STANDARDS-BASED PRACTICES AND MATHEMATICS ACHIEVEMENT:
A HIERARCHICAL LINEAR MODELING ANALYSIS

By

Jack Robbins Dempsey

August 2010

Chair: John H. Kranzler
Cochair: Stephen J. Pape
Major: School Psychology

The instructional practices used to teach mathematics in U.S. classrooms have undergone few changes during the past century. Decades of poor performances in international tests of mathematics achievement, however, have caused a substantial portion of the mathematics education community to begin advocating for the reform of mathematics education. Changes to the traditional instructional practices—largely reliant on passive learning processes such as memorization—constitute the heart of the reform agenda. Specifically, the reformers advocate for teaching mathematics as a discussion rather than a drill, emphasizing the process of arriving at a solution over the actual solution. In reform-based classrooms, students are expected to invent their own problem-solving processes to defend or modify their proposals in response to peer-inquiry. The teacher's role in the classroom is to facilitate such discussion rather than supplying algorithms and evaluating answers. In summary, the aim of the reform movement is the establishment of classroom environments in which mathematics is actively rather than passively learned.

Research supporting the reform movement has used predominantly qualitative research methods and thus has been criticized for its inability to draw causal conclusions regarding the effectiveness of reform approaches to mathematics education. The reform community has rebutted such criticisms, claiming that the rigorous procedural standardization required by the randomized control trials commonly used to establish efficacy would violate the free-flowing, conversational nature of reform instruction.

The purpose of the present study is to investigate the efficacy of reform-based instructional practices by applying a quasi-experimental methodology to a large-scale, longitudinal research database. This approach examines the influence of natural variations in teacher application of reform-based teaching techniques on student achievement using quantitative methods, thereby avoiding the problem of procedural standardization plaguing empirical studies. Results of the study revealed weak relationships between achievement at the elementary school level and all of the instructional practices examined. However, only items assessing teacher encouragement of process-focused mathematical discourse among students demonstrated positive associations with student achievement. Thus, the present study offers support for the teaching practices of the reform movement.

CHAPTER 1 REVIEW OF THE LITERATURE

Over the past 100 years, the presentation of mathematics to students in K-12 schools has remained largely unchanged in the United States (Cajori, as cited in Ellis & Berry, 2005; Fey, 1979; Stigler & Hiebert, 1997). Traditional mathematics education is composed of two phases: acquisition and application (Stigler & Hiebert, 1997). During the acquisition phase, teachers demonstrate a single computational strategy for solving a specific problem type. During the subsequent application stage, students are provided with opportunities (i.e., worksheets) to repeatedly practice applying that computational strategy. For this reason, the traditional approach to mathematics instruction is often described as “drill and practice” (Bosse, 1995; Putnam, Heaton, Prawat, & Remillard, 1992).

The drill and practice methodology is often linked to Edward L. Thorndike, a prominent proponent of this form of instruction during the early 1900s (Brown, 1994; Putnam et al., 1992). For Thorndike, the goal of mathematics education was having students achieve perfect arithmetical accuracy. Economic interests underlay this concern with accuracy. In *The Psychology of Arithmetic*, Thorndike (1922) makes frequent mention of the serious economic consequences likely to accompany the errors of a business clerk when performing a variety of tasks currently performed by computer programs (e.g., adding or subtracting columns of digits, calculating the compound interest of a loan).

Thorndike’s proposed methods for developing arithmetical accuracy in students were rooted in his belief that learning was a passive process. Specifically, Thorndike believed that learning consisted of the formation of a mental association or bond in the

mind of an individual between a specific stimulus (S) and a specific response to that stimulus (R). Placed in an arithmetical context, the stimulus is the arithmetic problem and the response is the student's answer to that problem. The strength of this S-R bond is increased if the response causes a satisfying consequence and decreased if an unsatisfactory consequence results (Thorndike, 1913). Thorndike referred to the consequence immediately following the response as the effect.

In a classroom context, the effect was the feedback the child received regarding the correctness of the answer. With regard to the effect, Thorndike believed that children naturally desired to answer correctly (Thorndike, 1922). Therefore, notification of a correct answer strengthened the S-R bond, whereas an incorrect answer weakened it. Importantly, the "effect was not part of the association that was established during learning" (Hearst, 1999, pp. 441-442). In other words, Thorndike did not believe that individuals formed stimulus-response-effect bonds. In this conception of learning as a passive process, the individual would be cognitively unaware of why he was performing a frequently reinforced response upon the presentation of the specific stimulus.

Thorndike believed that practice was the key to strengthening S-R bonds. This aspect of his learning theory was known as the Law of Exercise (Hergenhahn, 1992). Together, the Law of Effect and the Law of Exercise dictated that repeatedly practicing a particular arithmetical S-R bond would strengthen the bond. In a strong S-R bond, the response would appear following nearly every presentation of the stimulus. Thus, in Thorndike's conception of mathematics education, teachers were of less importance than textbooks or worksheets because of the practice opportunities associated with the latter. Notably, none of Thorndike's research in the area of mathematical learning

investigated the impact of modifying the instructional delivery of teachers (Donovan & Thorndike, 1913; Thorndike, 1910, 1915; Thorndike, 1922; Thorndike, 1925). Instead, the effects of extra practice via worksheets were examined. Thorndike eventually modified the Law of Effect and discarded entirely the Law of Exercise (Hergenhahn, 1992). These decisions, however, were made after the publication of *The Psychology of Arithmetic* (1922), and he never sought to redefine his philosophy on mathematics in light of these changes.

Mathematics Education in the U.S. and Abroad

Although conducted nearly 20 years apart, the studies of Porter (1989) and Hiebert et al. (2005) both reveal the continuing influence of Thorndike's theories on mathematics education. In 1989, Porter examined a year's worth of daily activity logs from 41 elementary school mathematics teachers and found that most teachers devoted between 70-75% of their instructional time to skills-based topics, such as how to add, subtract, multiply, and divide. This finding corresponded to the results of a content analysis of commonly used fourth grade textbooks also reported in the article. Here, Porter (1989) again found an emphasis on lower-level procedural learning as 65-80% of the exercises in these books focused on procedural skills practice (e.g., $20 \div 4 = ?$).

In a cross-national comparison of mathematics instruction, Hiebert et al. (2005) also found a heavy emphasis on procedural mathematics learning within the United States. These researchers used data from the 1999 Trends in International Mathematics and Science Study (TIMSS) to compare mathematics instruction in the United States with instruction in countries with a history of outperforming American students in international tests of mathematics (e.g., Japan). Their analysis revealed that mathematics education in higher-achieving countries was characterized by instruction

emphasizing conceptual understanding along with procedural skills. In contrast, teachers in the United States continued to display the relentless focus on lower-level procedural skills described by Porter (1989) nearly 20 years earlier.

This style of teaching represents the influence of Thorndike's drill and practice philosophy on mathematics instruction in America (Resnick & Hall, 1998). When the goal of mathematics instruction is the strengthening of S-R bonds, an instructional focus on procedural skill-building over conceptual understanding is natural. From this perspective, developing a deeper conceptual understanding of addition or subtraction is unnecessary and inefficient for student learning. In other words, students do not have to understand the conceptual underpinnings for adding fractions, they only need to memorize rules, such as $\frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ (Thorndike, 1922).

The extent to which an instructional emphasis on procedural or conceptual learning impacts the teaching techniques of mathematics instructors has been explored through cross-cultural discourse analysis. Both Japanese and American teachers spend a substantial portion of class time engaged in teacher inquiry, student response, and teacher feedback (IRF) discourse patterns (Inagaki, Morita, & Hatano, 1999; Wells, 1993). In the U.S., the teacher feedback in over half of the IRF discourse patterns consisted of directly evaluating the correctness of the student response. Japanese teachers, however, rarely engage in such direct evaluation, preferring to invite other students to evaluate the response. Additionally, the feedback of American teachers was more likely to be directed to the preceding speaker, whereas Japanese teachers were more likely to address feedback to the entire group of students, such as asking other students if they understood what the preceding speaker had said.

Because of these differences, the typical IRF unit was of shorter duration in American lessons despite American and Japanese teachers spending comparable amounts of class time in IRF discourse. Specifically, 57 IRF units occurred over 26.7 minutes for the median American teacher and 17 IRF units occurred over 26.3 minutes for the median Japanese teacher (Inagaki et al., 1999). This difference resulted from the tendency of American teachers to ask simple procedural questions more frequently than their Japanese counterparts. The following is an example of the smaller American IRF units:

Teacher: All right, look at the next one. Here, they've done it for you. They're telling you $\frac{5}{6}$ is equal to $\frac{15}{18}$. You have to decipher it, be the detective. What did they do in the process, Peter?
Peter: They multiplied it by 3.
Teacher: Right, times 3. (p. 107)

The above exchange exemplifies the short IRF units, simple procedural questions, and the evaluative (i.e., right or wrong), individually focused feedback characteristic of traditional American mathematics instruction. These practices appear to represent a continuation of Thorndike's belief that mathematical learning consists of forming S-R bonds, which can be strengthened through a high frequency of practice opportunities. The use of these practices in present day mathematics classrooms suggests that the learning theory outlined in *The Psychology of Arithmetic* continues to influence American education close to a century after its publication (Resnick & Hall, 1998).

Compared to their American counterparts, Japanese teachers pose fewer, but more complicated questions to students. They spend more time working through these questions, often eliciting opinions on problem-solving strategies from several students to illustrate different methods of finding the solution. The following example depicts a teacher in the middle of an IRF unit; the teacher has already asked two students to

describe how they reached their somewhat different answers to a previously posed question and is now asking a third student for his opinion on the relative merits of each process.

Teacher: Then, can we conclude that Kayo's and Takashi's answers are the same? Are $45/60$ hours and $9/12$ hours the same? Ryoji, do you have any opinion?

Ryoji: I think all the answers are the same. As for Keiko's and Takashi's, we get Takashi's by multiplying both numbers of 4 and 3 of Keiko's answer $\frac{3}{4}$ by 15, and we get Kayo's by multiplying both numbers, numerator and denominator of Keiko's by 3. Also we get Kyoko's by multiplying both numbers of Keiko's by 2---thus, all the answers are equal.

Teacher: Let's put aside Ryoji's opinion for a while-we'll examine it later, but let's first check whether Kayo's and Takashi's, $45/60$ hours and $9/12$ hours are the same. Can we conclude that these are equal? Who believes that they are equal? Who are not yet convinced? (Inagaki et al., 1999, p. 107)

The low frequency/long duration IRF units, complicated problems, and lack of evaluative feedback characterizing mathematics education in Japan indicate that Japanese teachers' conception of the learning process substantially differs from the teachings of Thorndike. In particular, the lack of direct teacher feedback regarding the accuracy of proffered solutions suggests that Japanese educators conceive of mathematical learning as an active process (Schumer, 1999) rather than the passive process of S-R bond formation envisioned by Thorndike (1922). In other words, the teaching practices of Japanese instructors indicate a belief that mathematics learning results from deep conceptual knowledge of problem-solving processes acquired through teacher-facilitated student-to-student mathematical discourse. As will be described in the next section, this theory of learning is referred to as social constructivism (O'Connor, 1998).

Calls for Reform

To many in the mathematics education community, the notable instructional differences between the U.S. and other industrialized nations (Hiebert et al., 2005; Inagaki et al., 1999) and the decades-spanning mediocrity of American students in international tests of mathematics (e.g., Lemke et al., 2004; McKnight et al., 1987; Mullis, Martin, Gonzalez, & Chrostowski, 2004) indicate a need to abandon drill and practice teaching. The National Council of Teachers of Mathematics (NCTM) published the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) to provide a roadmap for reforming mathematics education in the U.S. This publication and its 2000 revision, *Principles and Standards for School Mathematics* (often referred to as the *Standards*), provide a perspective that is in opposition to that of Thorndike, stating that “students who memorize facts or procedures without understanding often are not sure when or how to use what they know, and such learning is often quite fragile” (2000, p. 20). The *Standards* further noted that “learning mathematics without understanding has long been a common outcome of school mathematics instruction [in the U.S.]” (p. 20).

As described by Woodward and Montague (2002), “the NCTM Standards are grounded in cognitive and constructivist approaches to learning, as reflected in their emphasis on developing children’s ability to think about mathematics” (p. 90). With regards to thinking about mathematics, the *Standards* state that “mathematics makes more sense and is easier to remember and to apply when students connect new knowledge to existing knowledge in meaningful ways,” (NCTM, 2000, p. 20). Thus, the *Standards* depict the child as an active participant in the learning process: one who strives to incorporate novel information into a pre-existing schema through the process

of assimilation and accommodation described by Piaget (1954). Whereas Thorndike viewed learning as a passive process on the part of the learner, the *Standards* argue that the learning process requires the learner's active participation.

It is important to note that the *Standards* (NCTM, 2000) are not built upon a purely Piagetian constructivist foundation (O'Connor, 1998; Salomon & Perkins, 1998). This "radical constructivism" (O'Connor, 1998, p. 34) is a developmental learning theory focused on the cognitive development of the individual. From this perspective, a child is constantly in the process of "assimilating data from the world and accommodating to the world by creating new knowledge structures" (O'Connor, 1998, p. 34). When examining learning situated in the classroom, however, many constructivists found that the individual's construction of knowledge was significantly impacted by the myriad social elements operating within this setting (O'Connor, 1998; Salomon & Perkins, 1998). In other words, these researchers realized that the construction of knowledge within the classroom is profoundly affected by factors such as attention, motivation, and social and discourse norms created by joint interactions among members of the classroom community (Cobb & et al., 1991; O'Connor, 1998; Salomon & Perkins, 1998; Yackel, Cobb, & Wood, 1991).

Social constructivism is a broad term that encompasses multiple viewpoints (O'Connor, 1998). Regarding its contributions to the *Standards* (NCTM, 2000), however, social constructivism can be understood as the idea that any type of learning within the school system cannot be considered apart from the social dynamics of the classroom in which the learning is situated. In other words, "mathematical activity can be viewed as

intrinsically social in that what counts as a problem and as a resolution has normative aspects" (Cobb, Wood, & Yackel, 1993, p. 93).

Yackel and Cobb (1996) provided an anecdote from their work demonstrating the influence of classroom norms on mathematical problem solving. They described a situation in which "one child attempted to resolve a dispute about an answer during small-group work by initiating a discussion about who had the best pencil and then about which of them was the smartest" (pp. 467-468). According to the authors, the student's practice of "relying on authority and status to develop [mathematical] rationales" (p. 467) was the result of exposure to drill and practice classrooms where teachers are the only members of the classroom community to provide mathematical explanations.

This example illustrates a dichotomy between the social norms regarding mathematical discourse (sociomathematical norms) within the academic discipline and within the classroom noted by proponents of reform (e.g., Civil, 2002; Lampert, 1990; Yackel & Cobb, 1996). Professional mathematicians do not resolve debates based on status. Instead, answers are justified by the validity of the mathematical processes used to arrive at them. To align the sociomathematical norms of mathematics classrooms with those governing the professional discipline, NCTM (2000) made the development of classroom learning environments that support doing and talking about mathematics one of the central goals of mathematics reform. In such a classroom environment, students will develop, justify, and defend their own mathematical reasoning and critically assess the reasoning processes of their peers. In short, they will be actively involved in

the learning process. This goal represents the underlying social-constructivist foundation of the *Standards* (NCTM, 2000).

Standards-based Instructional Practices

The instructional methodology arising from the reform movement's social-constructivist underpinnings is referred to as inquiry-based instruction (Wilkins, 2008). Students' active engagement in solving conceptually-rich problems and an emphasis on "the notion that mathematics is a social activity in which discussion, justification, argumentation, and negotiation are central to the mathematical discourse among students, and between students and teachers" (Wilkins, 2008, pp. 140-141) characterize this method of teaching. In other words, rather than teaching students a single problem-solving strategy and subsequently presenting them with multiple opportunities to practice its use, teachers in reform-based classrooms present students with problems requiring the use of one or more specific algorithms prior to formally teaching the procedures (e.g., Cobb et al., 1991; Lampert, 1990). The students must apply background knowledge and problem-solving skills to the task of inventing the necessary algorithm. Because most arithmetical problems can be solved in multiple ways, and because students have varying levels of facility with mathematics, multiple solutions to these problems are expected. Guided by the teacher, the class is expected to engage in the process of mathematical debate to determine which, if any, of the presented solutions are correct. A classroom operating in this fashion is referred to as a community of inquiry (Lipman, 1987).

Multiple forms of inquiry-based instruction exist. For example, the instructional format in Cobb et al. (1991) differs from that used by Lampert (1990), which in turn differs from that of Civil (2002). According to Wilkins (2008), however, "instructional

practices that would support a community of mathematical inquiry...are representative of inquiry-based instruction” (p. 141). In other words, inquiry-based instruction can take multiple forms, but the goal of the instruction is always the same: creating a community of inquiry. Specific techniques used to accomplish this task may include, but are not limited to the following: whole class instruction, student use of manipulatives, collaborative learning, and situating mathematics in realistic contexts.

To aid teachers in creating such communities of inquiry, the *Standards* recommends the use of “interesting problems that ‘go somewhere’” (NCTM, 2000, p. 60) to stimulate problem-solving discourse among the students. In terms of developing interesting tasks, the *Standards* (NCTM, 2000) propose the use of problems featuring multiple pathways to a correct solution. For younger students, the use of arithmetical tasks is recommended. It is the NCTM's intention for even the youngest learners to receive exposure to inquiry-based instruction despite a determined opposition stating that arithmetic must be learned through drill and practice (e.g., Vukmir, 2001).

Specifically, the *Standards* (NCTM, 2000) state:

A good setting in which young students can share and analyze one another's strategies is in solving arithmetic problems, where students' invented strategies can become objects of discussion and critique. Students must also learn to question and probe one another's thinking in order to clarify underdeveloped ideas. Moreover, since not all methods have equal merit, students must learn to examine the methods and ideas of others in order to determine their strengths and limitations. By carefully listening to, and thinking about, the claims made by others, students learn to become critical thinkers about mathematics. (p. 63)

The *Standards* (NCTM, 2000) also encourage teachers to assign students to work in small groups and to situate mathematics within realistic contexts. These two instructional techniques are often used in conjunction with inquiry-based mathematics, but are not synonymous with the term. Importantly, small group work is expected to

facilitate mathematical discourse within the groups in the style of whole-class, inquiry-based instruction, but on a smaller scale, thereby allowing more students to participate. With respect to situating mathematics in realistic contexts, the *Standards* (NCTM, 2000) mandates that students be able to “recognize and apply mathematics in contexts outside of mathematics” (p. 65). To develop this ability, teachers are expected to build connections between mathematics and “other subject areas and disciplines as well as to students’ daily lives” (NCTM, 2000, p. 66).

Research on Standards-based Instructional Practices

To align with the recommendations of the *Standards* (NCTM, 2000), current teacher practices must radically change. To examine the processes through which to facilitate this change, educational researchers conducted teaching experiments (Cobb, 1995, p. 25) characterized by the use of qualitative methods to depict researchers’ attempts to align classroom practices with the *Standards* (NCTM, 2000). In these attempts at alignment, researchers either directly assumed classroom instruction duties or served as an instructional consultant to the regular classroom teacher. It is important to note, however, that the methodologies of the studies below defy strict standardization due to the student-centered nature of the instruction.

In one of the most highly regarded teaching experiments (Schoenfeld, 2007), Lampert (1990) chronicled her efforts to move a fifth-grade class from conventional mathematics beliefs in which “students believe that the teacher knows which answers are right, and teachers believe that the paths to these answers can be found in rules in books” (p. 32) to a form of classroom interaction in which “truth came to be determined by logical argument among scholars” (p. 35). Lampert sought to use instructional modifications—inquiry-based instruction—to bring the sociomathematical norms of the

classroom closer to those governing the professional discipline. Her goal was to enable students to engage in and learn from the type of mathematical discourse used by professional mathematicians. Lampert described her research methodology as follows:

My role in this project has been to develop and implement new forms of teacher-student interaction as well as to experiment with new forms of content as a teacher of fifth-grade mathematics. I have taught fourth- and fifth-grade mathematics during the past 6 years, collecting data on both teaching and learning during 3 of those years. The teaching practice that produced the data was constructed to be congruent with ideas about what it means to do mathematics in the discipline.

To convince students to abandon traditional sociomathematical norms (i.e., status-based determination of correctness), Lampert refused to evaluate student responses herself. As a result, students had to justify the processes by which they reached their own answers and compare these processes to those of peers who arrived at different answers. To facilitate mathematical debate, Lampert used conceptually difficult questions with multiple solution pathways. For example, students were asked to determine what the last digit of 5^4 would be without multiplying $5 \times 5 \times 5 \times 5$. Suggested student solutions were put on the board with the student's name by his or her solution. Students were then asked to justify their solutions and classmates were encouraged to question or disagree with these justifications. When disagreeing with another student's hypothesis, however, students were required to use language such as "I want to question so-and-so's hypothesis" (p. 40) and to give reasons for the questioning so that "their challenge took the form of a logical refutation rather than a judgment" (p. 40). Lampert expected content learning to occur through the resulting mathematical discourse and directly conveyed few problem-solving procedures (i.e., algorithms).

To provide evidence for the occurrence of learning, Lampert described the following anecdote:

Sam asserted, about the last digit in 5^4 , "It has to end in a 5." I invited everyone in the class to consider the validity of Sam's decisive assertion and to see if they could explain why he seemed to be so sure. The question I was asking was, How does he know that is true? Harriet said, "Well, anything multiplied by 5 has to end in a 5 or a zero," and Theresa quickly added, "but it has to be a 5 because when you multiply 5 times 5 you get a 5 [for a last digit]." Martha observed, "You times the square number, you square it again and you get 625." And Carl responded, moving to the level of a mathematical generalization, "You don't have to do that. It's easy, the last digit is always going to be 5 because you are always multiplying last digits of 5, and 5 times 5 ends in a 5." (p. 48)

According to Lampert, Carl's response indicated both a conceptual knowledge of exponents and of what it meant to engage in mathematical discourse. She noted that, after listening to and scrutinizing the various solutions of his classmates, Carl raised the discussion to the level of a generalized strategy that works for five raised to any power and provided mathematical evidence supporting his argument. This generalized strategy represented the type of "well-connected, conceptually grounded ideas...readily accessed for use in new situations" (NCTM, 2000, p. 20) seen as the desired outcome of *Standards*-based instruction.

Another such outcome was described as follows:

By the end of the lesson, 14 of the 18 students present in the class had had something mathematically substantial to say about exponents: an interpretation of language or symbols, an assertion about a pattern, a proof that a pattern would continue beyond the observed data, or an interpretation of another student's assertion. (p. 52)

Thus, Lampert's style of instruction enabled the vast majority of students to actively participate in mathematical discourse. Through this process, her students fulfilled the NCTM's goal of learning to "communicate their mathematical thinking coherently and clearly to peers, teachers, and others," (NCTM, 2000, p. 61). In summary, Lampert's findings appear to fulfill the goal of reform mathematics elucidated by Cobb et al. (1991) as students coming "to view mathematics as an activity in which they are obliged to

resolve problematic situations by constructing personally meaningful, justifiable solutions as they actively contribute to the interactive constitution of an inquiry mathematics tradition” (p. 8).

Lampert (1990) used rich problems and activities to create a mathematical discourse community within the classroom; however, this process can also be accomplished using simplistic, procedural problems. Using such activities, Yackel, Cobb, and Wood’s (1998) teaching experiment depicts the use of inquiry-based instruction to stimulate mathematical discourse, active learning, and conceptual understanding amongst students. In this teaching experiment, a second-grade teacher was provided with extensive external supports (via weekly meetings with researchers) in the use of classroom activities to enable students to “develop the ability to engage in mathematical discussion including giving explanations and justifications as appropriate” (p.472). Through observations conducted over the course of a year, the researchers chronicled the instructor’s use of number sentence activities (i.e., $5 + 6 = ?$) to teach addition within an inquiry mathematics tradition.

Activities such as number sentences form the backbone of traditional school mathematics. Instead of having his students simply answer the question, however, the teacher asked them to describe their problem-solving processes and find multiple solution pathways. His goal for this approach was to stimulate classroom discourse and deepen conceptual understanding. Importantly, the “teacher legitimized solutions that consisted of decomposing the summands in differing ways and combining the results of the decomposition in various orders but offered sanctions against those that were little more than restatements of previously given solutions” (p. 476). Thus, listening to and

critically examining the problem-solving procedures used by one's peers and inventing a different method from those already used became necessary components for classroom participation. This tactic was highly successful. For example, 10 students produced 13 different solution pathways to the problem of $16 + 14 + 8$. In this way, mathematics became an active rather than passive process for students.

This manner of inquiry-based instruction contributed to students' conceptualizations of tens and ones units and led to the development of generalized strategies for addition and subtraction. Specifically, as the school year proceeded, many of the students came to understand that the most efficient way to solve two and three digit addition and subtraction problems was to decompose the numbers into tens and ones and then rearrange them to facilitate problem solving. For example, one student answered the question of $37 + 24 + 13 = ?$, saying "7 + 3 = 10, plus 10 from 14 is 20, plus 20 is 40, plus 30 is 70, plus 4 is 74" (p. 477), whereas another student provided the following solution pathway, "30 + 20 equals 50, plus 7 is 57, plus 3 is 60, plus 4 is 64, plus 10" (p. 477). These types of generalized strategies represent a substantial improvement over the counting and rote memorization strategies used by students at the beginning of the school year. Yackel et al. (1998) offered the following assessment of the benefits of inquiry-based instruction:

Interviews conducted at the beginning of the school year indicated that a majority of the children could not count on to solve a missing addend task, could not decompose numbers into component parts, and were unable to coordinate units of ten and one. By the end of the year almost all of the children could do these things. The number sentence activity played a crucial role in the development of the children's increasingly powerful mathematical conceptions, especially their concepts of tens and ones. (p. 482)

Despite differences in the mathematical activities used to create a culture of inquiry, students in the classrooms of Lampert (1990) and Yackel, Cobb, and Wood (1998) developed increasingly sophisticated mathematical reasoning skills through exposure to inquiry-based instruction. Expanding on these findings, Civil (2002) examined the impact of inquiry-based instruction using mathematical activities situated within realistic contexts on students' learning and conception of mathematics during a year-long experiment in which she co-taught a fifth-grade mathematics course. During this time, Civil and her research staff spent approximately 6 hours per week in the classroom and engaged in over 60 hours of planning and debriefing with the regular classroom teacher. Sources of data included field notes and the students' permanent products (i.e., homework).

Similar to Lampert (1990), inquiry-based instruction in this study consisted of using conceptually difficult, open-ended questions with several possible solution pathways to facilitate mathematical discourse. The role of the teacher was to facilitate and guide the discussion while refusing to validate answers. Students were expected to learn mathematical content through generating and refining their own solutions to these problems rather than receiving preformed knowledge from a teacher.

Civil's (2002) version of inquiry-based instruction differed slightly from Lampert's (1990) in that students were exposed to mathematics situated within realistic contexts prior to moving to more abstract mathematics involving mathematical notation. For example, during the geometry unit taught by the researcher, classroom activities progressed from identifying patterns and tessellations in local southwestern art and clothing patterns to identifying the sum of the interior angles within a hexagon. In her

own words, Civil attempted “to develop a teaching innovation that would promote the mathematical values of mathematicians' mathematics by connecting these values to students' interests and everyday experiences” (p. 60).

The author reported that this style of instruction produced unprecedented levels of student participation when the open-ended activities were situated in realistic/everyday contexts. For example, a homework assignment requiring students to identify and draw three different symmetrical patterns occurring within their home or neighborhood was almost unanimously completed—a rarity for the class. Furthermore, during an in-class discussion of these patterns, students who rarely participated in class and considered themselves to be bad at mathematics actively contributed to the discussion.

These enhanced levels of participation, however, exposed substantial within-class variation regarding students' conceptual understanding of the lessons. Regarding the homework assignment described above, Civil wrote “their written work left me wondering about their understanding of patterns (for some students, almost anything was a pattern)” (p. 53). Furthermore, the number of students contributing to the classroom discourse fell dramatically as the lesson moved towards abstracting principles from the aforementioned everyday mathematics tasks. As described by Civil:

During our work on finding the basic repeating tile on tessellations, we had succeeded in attracting as participants students who hardly ever added their voices to mathematics discussions. Yet, as soon as we became more involved in the exploration of angles and the question of why some shapes tessellate and other do not, many students withdrew from the conversation. By the time we were working on the task of finding the measure of angle A, only a few students seemed to be participating. (p. 58)

The author suggested that the decline in participation described above may have resulted from students' perceptions that the latter two tasks were “more mathematical” (p. 58), which activated their feelings of low self-efficacy regarding mathematics and

caused them to cede the floor to the students perceived to be good at mathematics. In summary, Civil (2002) demonstrated that situating mathematics in realistic contexts within inquiry-based instruction resulted in a substantial increase in the variety of students participating in mathematical discourse. It is unknown, however, if the increased participation resulted in increased content knowledge.

Steencken and Maher (2003), however, were more successful in demonstrating that inquiry-based instruction using realistically-situated activities enhances students' conceptual knowledge of mathematics. This study reports on the first seven sessions of a year-long teaching experiment with fourth-grade students in which the authors served as the instructors. The authors' sources of data included videotaped classroom sessions, students' written work, and the authors' field notes.

As teachers the authors frequently attempted to situate student's initial exposure to mathematical concepts in familiar everyday contexts (i.e., using candy bars to explain fractions) in a manner similar to that of Civil (2002) and used rich questions with several possible solution paths to stimulate mathematical discourse. As usual, the role of the teacher was to refrain from validating any suggested solution and to facilitate classroom discussion by encouraging students to justify their answers and to respectfully question the problem-solving processes of their peers.

Collaborative learning was also a major component of Steencken and Maher's teaching experiment. The students worked in pairs on the problems posed by the teacher and the researchers rearranged the pairs at multiple times in the experiment if they felt that an existing pair was not working well together. Additionally, the students were encouraged to write about their solutions—not just in symbols, but in words, so as

not to forget. Finally, the teaching experiment featured the extensive use of manipulative materials (rods of a variety of lengths). In many of the earlier sessions, students were expected to model their problem-solving processes using these manipulative materials. Although, this study emphasizes the use of manipulative materials more than some of the others reviewed in this section, it is important to note that the manipulative materials were used within a student-centered context. That is, students were responsible for modeling their thinking using the manipulative materials rather than the teacher using them to illustrate ideas for the students.

The results of the teaching experiment demonstrated that the use of reform-based mathematics both changed students' conceptions of the nature of mathematics and resulted in a conceptual understanding of the material deeper than that provided by rote memorization. With regards to the former, this American mathematics classroom—characterized by short, product-based answers and correct/incorrect teacher product evaluations—was transformed into a discourse community in which ideas were exchanged, debated, justified, and refined. In short, mathematical discourse within the classroom became more similar to the discourse of professional mathematics. By the end of the experiment, students were accustomed to justifying their answers using manipulative materials and to critiquing the reasoning processes of their peers.

To demonstrate the positive impact of the instruction on student learning, the authors provide several anecdotes. In one, a student provided an incorrect solution and later refined it based on criticism offered by his peers. In another, a student demonstrated two possible solutions to a problem (i.e., $1/6$ and $2/12$) and justified both.

Learning was also demonstrated as the students began to express fractional ideas in more precise language. The authors provide an example of this process:

In later sessions, the children interchanged the color names for rods with number names for fractions. In Session 5, Jessica and Laura describe the difference between one half and one third as “a red bigger.” Later, the children ultimately named the difference “one sixth.” (p. 130)

Another example related to the development of students’ language for expressing fractional ideas suggested that the students developed deeper conceptual knowledge of the subject matter than could be attained through the memorization of S-R bonds. Specifically, when two students (Jessica and Allan) presented differently-sized models demonstrating that $1/2 - 1/3 = 1/6$, Jessica objected to Allan’s model stating that “it can only be one size candy bar and that’s it” (p. 124). She was referring to a prior demonstration in which the teacher broke two differently sized candy bars in half to show the students that one-half of the small candy bar did not equal one-half of the larger candy-bar, although they were both halves.

Taken by itself, Jessica’s comment suggests that the teacher’s demonstration of a mathematical principle using everyday objects familiar to the students was a failure. This demonstration should be regarded as a success, however, because two other students were able to demonstrate to Jessica the flaws in her reasoning using the candy bar analogy. This example demonstrates the capacity of *Standards*-based instructional practices to develop the ability to apply mathematical principles to realistic situations in students, one of the stated goals of mathematics reform (NCTM, 2000).

In addition to the use of realistically-situated mathematical activities, Steencken and Maher’s (2003) study is also notable for its use of collaborative learning activities within an inquiry-based instruction framework. Although many of the above authors

used a whole-class instructional format, collaborative learning can be a valuable supplement to this type of instruction as it provides students with a greater number of opportunities to contribute to a mathematical discourse.

Yackel, Cobb, and Wood (1991) sought to investigate the collaborative learning processes occurring within teacher-assigned dyads in an inquiry-based classroom. The methodology of this experiment is described below. To provide detailed information regarding collaborative learning processes, the researchers focused their analysis on four pairs of students within a single second-grade classroom which alternated between whole-class and collaborative learning formats. Within this classroom, the teacher's main responsibility during the collaborative learning portion of the class was helping the pairs "learn how to engage in a collaborative dialogue about mathematics" (p. 392). Initially, more mathematically-advanced members of the pairings would simply work ahead, failing to consult their partners or explain their solution processes. At other times, the more advanced member of the pair would explain his or her problem-solving processes but would refuse to acknowledge the incorrect answers provided by the other member. Thus, the teacher was required to intervene with students to model her expectations through participating and guiding the dialogue between the pairs.

Gradually, social norms governing group work were negotiated between the teacher and students. Numerous learning opportunities arose when the pairs engaged in constructive mathematical dialogue. Specifically:

As the children work together and strive to communicate, opportunities arise naturally for them to verbalize their thinking, explain or justify their solutions, and ask for clarifications. Further, attempts to resolve conflicts lead to both the opportunity to reconceptualize a problem and thus construct a framework for another solution method, and the opportunity to

analyze an erroneous solution method and provide a clarifying explanation. (Yackel et al., 1991, p. 406)

The authors also described that these types of discussion-based learning opportunities do not arise within classrooms emphasizing the drill and practice method.

This qualitative analysis of the collaborative learning processes within a single classroom made use of a subset of data generated from a larger mixed-methods study conducted by Cobb et al. (1991). In this study, the authors used qualitative and quantitative data to compare the arithmetical learning and attitudes towards mathematics of 187 second-grade students in 10 inquiry-based classrooms to 151 students from 8 traditional classrooms over the course of a year. Inquiry-based classrooms alternated between small-group and whole-class instructional formats. At the start of each class, students worked in dyads to answer one or more mathematical problems. After 20 minutes, the students would reconvene for a whole-class discussion in which the problem-solving processes and products of the various groups were presented and subsequently debated.

Within these classrooms, conceptually-difficult problems and teacher refusal to validate answers were used to facilitate mathematical discourse among the class. Students sought to justify their answers and to refute the mathematical reasoning of others when they perceived it to be incorrect. As with the forms of inquiry-based instruction described above (Civil, 2002; Lampert, 1990; Steencken & Maher, 2003), the expectation was for students to learn mathematical content through generating and refining their own solutions to problems rather than receiving this knowledge directly from a teacher. The quantitative findings of this study are presented below.

Alternatives to Qualitative Research

The explicit goal of the above studies was to demonstrate that the teaching techniques recommended by the *Standards* (NCTM, 2000)—hereafter to be referred to as Standards-based instructional practices—could create a “shift in the social norms away from conventional classroom discourse patterns” (Lampert, 1990, p.33). Although the majority of these studies (Lampert, 1990; Steencken & Maher, 2003; Yackel, Cobb, & Wood, 1998) provided anecdotal evidence of deep conceptual learning occurring in response to these practices, these and other qualitative studies of reform mathematics have been criticized for failing to describe the extent to which learning outcomes differed between traditional and reform instruction (Benbow & Faulkner, 2008).

Based on these concerns, the research of Cobb et al. (1991) represents an important contribution to the literature on reform mathematics as it avoids several of the limitations associated with the above research (e.g., potential selection bias and subjective outcome measures). In this study, students completed two arithmetical tests (ISTEP and Project Arithmetic Test) and a questionnaire regarding beliefs about mathematics at the end of the school year. The ISTEP—a standardized assessment—was composed of two subtests: Computation and Concepts and Applications. The Computation subtest consisted of items requiring direct computation of two-or-three digit addition or subtraction problems presented in a vertical column format. The Concepts and Applications subtest required students to solve addition and subtraction problems using graphic representations and to deconstruct large numbers into smaller units (i.e., 75 into 7 tens and 5 ones units). The Project Arithmetic Test—a researcher-designed

measure—was also composed of two subtests: Instrumental and Relational. In describing the two scaled scores, the authors reported:

The scale was labeled Instrumental in that it was possible to perform well on it by using computational algorithms without conceptual understanding. In contrast, items on the Relational scale were designed to assess students' conceptual understanding of place-value numeration and computation in nontextbook formats. (p. 15)

Students in the treatment and control groups performed comparably on the Computation and Instrumental portions of the ISTEP and Project Arithmetic Test, respectively. Treatment students, however, significantly outperformed controls on the Concepts and Applications and Relational portions of these tests. This result suggests that classrooms using inquiry-based instruction in conjunction with collaborative learning provided students with a deeper conceptual knowledge of arithmetic than traditional classrooms. Additionally, student responses to the beliefs questionnaire indicated that students from treatment classrooms believed that successful mathematics students were more likely to collaborate with others and share ideas with their peers, whereas students from control classrooms were more likely to associate success with conforming to the solutions of others. In other words, children were less likely to view mathematics as a process of solving problems using known algorithms and procedures following a year of inquiry-based instruction.

Although this study appears to supplement qualitative examinations of reform-based instructional practices by providing objective, generalizable results depicting a cause-and-effect relationship between this form of instruction and student learning, several sources argue to the contrary (Benbow & Faulkner, 2008; Slavin & Lake, 2008). Randomized controlled trials represent the gold standard in experimental research, because the process of random assignment balances the distribution of confounding

variables between treatment groups (Slavin, 2008). The results of experiments in which comparison groups are matched on pretest scores are also considered capable of revealing cause-and-effect relationships (Benbow & Faulkner, 2008; Slavin & Lake, 2008). Because Cobb et al. (1991) did not use either of the above designs, their finding of significantly different post-test differences cannot be conclusively determined to have resulted from the experimental treatment rather than from a prior achievement discrepancy between the groups.

Other researchers, however, have used randomized controlled trials and matched comparison group designs to investigate the impact of *Standards*-based instruction on student learning at the elementary school level. Such studies, however, have been met with criticism from reform-oriented researchers (Boaler, 2008; Borko & Whitcomb, 2008; Cobb & Jackson, 2008; Confrey, Maloney, & Nguyen, 2008; Lobato, 2008). According to these critics, such experimental designs are nearly impossible to implement within the educational system because researchers “cannot persuade schools to treat children as experimental subjects and [randomly] assign them to different conditions” (Boaler, 2008, p. 590). On the rare occasions school administrators do allow such designs, they generally insist on study durations of days or weeks rather than months (Boaler, 2008), thereby limiting the generalizability of study findings to the classroom-based learning occurring over the course of the school year.

Other criticisms concern the ecological validity of studies produced using randomized controlled trials and matched comparison research designs (Boaler, 2008; Borko & Whitcomb, 2008; Cobb & Jackson, 2008; Confrey, Maloney, & Nguyen, 2008; Lobato, 2008). Specifically, reform-oriented researchers argue that the treatment

standardization required by such designs (e.g., Hopkins, McGillicuddy, De Lisi, & De Lisi, 1997; Rittle-Johnston, 2006) fails to realistically represent a classroom learning environment and is incompatible with the flexible, free-flowing nature of the inquiry-based instruction recommended by the *Standards* (NCTM, 2000) demonstrated in the teaching experiments described above (e.g., Lampert, 1990). In summary, many researchers in the mathematics education community consider randomized controlled trials and matched comparison designs within the educational system to produce invalid results as such experiments are characterized by short treatment durations and artificial learning conditions unlikely to occur within an actual classroom (Boaler, 2008; Confrey, Maloney, & Nguyen, 2008; Lobato, 2008).

Interpreted in isolation, neither type of research (experimental and qualitative/quasi-experimental) is capable of conclusively determining the efficacy of *Standards*-based instructional practices. The qualitative and quasi-experimental methods of the teaching experiments described above do not permit generalization of their results and the extent to which learning outcomes differ from drill and practice cannot be determined. Conversely, the results of the experimental studies described below are replicable and can be used to compare students exposed to reform or traditional instruction; however, the questionable ecological validity of such studies limits the ability to generalize their results to actual classrooms. To develop a detailed picture regarding the efficacy of *Standards*-based teaching practices, therefore, requires a thorough examination of both literatures.

In the following section, I will describe the results of experimental studies examining the use of *Standards*-based instructional practices such as inquiry-based

instruction, collaborative learning, and situating mathematics learning in realistic contexts. Findings from these studies will be contrasted with those from the teaching experiments. Inclusion criteria for the studies reviewed below are (a) the use of random assignment or matched comparison designs, (b) examination of realistically situated learning, collaborative learning, or at least one component of inquiry-based instruction, and (c) curriculum neutrality (i.e., treatment and control groups do not receive different curricula). All selected studies were of sufficient methodological rigor to also meet inclusion standards for the recent survey of the mathematics education literature conducted by the National Mathematics Advisory Panel (2008). Due to the limitations imposed by the selected research designs, however, some of the treatment conditions in the studies below may differ substantially from the descriptions of reform mathematics presented in qualitative research and in the *Standards* (NCTM, 2000). The impact of these differences on the obtained results will be discussed.

Experimental Studies of Standards-based Instructional Practices

No study included within this section explicitly purports to examine inquiry-based instruction. As discussed earlier, however, inquiry-based instruction is an umbrella term (Wilkins, 2008), even within the qualitative literature on the subject. Thus, although experimental conditions in this section are referred to as discovery learning, constructivist, and guided discovery among other names, they all display aspects of inquiry-based instruction. Importantly, however, the treatment conditions in some studies contain only superficial aspects (e.g., Rittle-Johnson, 2006) of this form of instruction. A comparison of the effects of superficial and central aspects of *Standards*-based instruction on student learning will be included in this review.

Rittle-Johnson (2006) used an experimental design to determine whether inquiry-based instruction (here referred to as discovery learning) produced significantly higher rates of learning compared to traditional teacher-directed instruction. Third-through-fifth grade students ($N = 85$) participated in a pretest, intervention, posttest, and delayed posttest (two weeks later) related to knowledge of mathematical equivalence problems.

Students in all conditions completed a series of eight mathematical equivalence problems with a repeated addend on both sides of the equation. The problems differed in the placement of the unknown value following the equal sign (e.g., $5+6+6=3+ _$ and $2+6+8= _+8$). Questions were administered via computer during a one-on-one session with one of two experimenters. Students were asked to describe their problem-solving processes for each question. As accuracy feedback, the computer program presented the correct answers following student responses.

Participants were assigned to one of four intervention conditions. The four conditions resulted from crossing two factors: (a) instruction versus no instruction and (b) self-explanation versus no explanation. For students receiving the instruction treatment, experimenters provided the participants with a computational strategy for solving the eight equivalence problems presented during the intervention. For students receiving the self-explanation treatment specialized accuracy feedback was provided by the computer program. Specifically, an additional computer screen appeared describing the answers provided by two students at a different school (one correct, one incorrect) following the completion of each problem. The student was then asked to describe the procedures each of the two students used to reach his or her answer.

Posttests were administered both immediately and two weeks after completing the intervention. Students' abilities to use of the same procedures learned in the intervention (procedural learning) and to adapt the procedures for use with novel question types (procedural transfer) were assessed. Additionally, experimenters measured conceptual knowledge of equivalence through a five-question assessment that asked students to perform tasks such as defining an equal sign.

Students in the teacher-directed instruction condition (instruction treatment) outperformed the no instruction condition on procedural learning items. Because the experiment required each participant to be alone in a room with the experimenter, however, students in the no instruction + no explanation condition were unable to benefit from the classroom discourse considered so essential in inquiry-based instruction. Therefore, little meaning with regards to reform versus traditional mathematics can be drawn from this finding.

The self-explanation condition was more representative of the interaction-based learning expected in inquiry-based instruction because—despite the computer-generated nature of the other students in this condition—the participants were asked to engage in discourse regarding the problem-solving processes of others. For students receiving this treatment, both procedural learning *and* the ability to adapt these procedures to solve novel problems increased regardless of instructional condition. This finding suggests that teachers who encourage student-to-student problem-solving discourse (a major component of inquiry-based instruction) will produce higher achieving students compared to their more traditionally-oriented counterparts.

Evidence from studies in which students were allowed to interact with actual rather than computer-generated peers also supports this conclusion. Muthukrishna and Borkowski (1995) randomly assigned 106 third-grade students from three different schools to one of four instructional conditions: guided discovery (containing aspects of inquiry-based instruction), direct instruction, a combination of guided discovery and direct instruction, and control. Over the course of the 14-day intervention, students received instruction in use of the part-whole/number family strategy, a technique previously found to be useful in solving addition and subtraction word problems. All treatment conditions were taught by the experimenter and a research assistant. Control students received instruction from their regular teachers, which represents an experimental confound.

The direct instruction condition received explicit and systematic teacher-direct instruction in the use of this strategy through a whole-class format. That is, instructors taught and modeled the use of this technique with certain problem types and provided opportunities for students to practice the application of these techniques through individual seatwork. In the combination condition, the instructor directly taught the part-whole strategy for the first two days but did little modeling of how to use it with specific problem types. Beginning on Day 3, students in this condition received the same style of instruction as the guided discovery group.

The style of instruction used in the guided discovery condition was similar to that used in the teaching experiments of Cobb, Yackel, and Wood (Cobb et al., 1991; Yackel, et al., 1991). Students worked in pairs on addition and subtraction word problems for the first half of class and engaged in whole-class discussions regarding

their solutions to these problems during the second half of the instructional period. The instructors worked to facilitate student interactions in both the group and whole-class formats “by asking questions such as ‘why?’, ‘explain that’, and ‘why did you reach that conclusion?’” (p. 432) to promote process-oriented answers.

Posttests used addition and subtraction word problems to measure near transfer (problem types similar to those learned) and far transfer (problem types different in either form or context from those previously learned). Posttests were administered one day, four weeks, and nine weeks post-intervention. Students in all treatment conditions significantly outperformed control students on these measures. The validity of any conclusions drawn from this finding is suspect, however, because significant curricular differences existed between the treatment and control conditions.

Of greater significance, students in the guided discovery and combination conditions significantly outperformed students from the direct instruction condition on measures of far transfer involving novel problem forms on the nine-week posttest. Both of these conditions contained elements of *Standards*-based teaching practices, such as inquiry-based instruction and collaborative learning.

The empirical findings reviewed above (Muthukrishna & Borkowski, 1995; Rittle-Johnson, 2006) suggest that exposure to core principles of inquiry-based instruction (e.g., student-to-student mathematical discourse) enables students to engage in more adaptive use of previously-learned mathematical principles when confronting novel problem types. Ginsburg-Block and Fantuzzo (1998) extended these findings by examining whether the use of collaborative learning strategies in conjunction with

aspects of inquiry-based instruction created greater gains in learning than the use of either of these *Standards*-based instructional practices in isolation.

The authors randomly assigned 104 low-achieving third and fourth-grade students to one of four researcher-taught instructional conditions: problem-solving instruction, peer-mediated instruction, problem-solving and peer-mediated instruction (PLUS), and traditional instruction (control). Students in all conditions received two 30-minute supplementary mathematics sessions per week for seven weeks. These sessions began with a 5-minute warmup exercise. Next, students spent 15 minutes working on computation and word problems. Each session concluded with the administration of a 10-item quiz. Students earned rewards based on their quiz scores.

After completion of the warm-up exercise, students in the problem-solving and PLUS conditions received exposure to several aspects of inquiry-based instruction. Specifically, they spent several minutes discussing their problem-solving processes for the warm-up exercise with the rest of the class and the problems provided subsequent to the warm-up often required the use of several alternative methods to solve and/or were presented in the form of a game or experiment (providing a more realistic context). Students in the PLUS condition also engaged in collaborative learning, as did students in the peer collaboration condition. Students in these conditions worked in dyads and alternated student and teacher roles while completing the 15-minute computation and word problem portion of the session. Using experimenter-provided sample solutions, the teaching member would provide instructional prompts and check answers as his or her partner worked on the problems. Students in the problem-solving and control conditions worked alone during this portion of the lesson.

Student learning was assessed using 20 minute pre-and post-tests constructed from computation and word problems taken directly from district textbooks. The authors found that students in all experimental conditions outperformed controls. These results indicate that students exposed to *Standards*-based instructional practices (collaborative learning or inquiry-based instruction) outperform peers taught using the drill and practice methodology. Importantly, all treatment conditions provided increased opportunities for student-to-student mathematical discourse (although level of discourse was not explicitly measured). This study provides evidence for the social-constructivist claim that students benefit from sociomathematical norms encouraging high levels of classroom discourse (Cobb et al., 1993; Cobb et al., 1991).

One of the most interesting aspects of Ginsburg-Block and Fantuzzo's (1998) study was the experimenters' use of reward structures. Specifically, students received rewards based on their performance on the quizzes administered at the end of every session. In the problem-solving and control conditions individuals received rewards based on their own quiz scores. Students in the peer collaboration and PLUS conditions, however, earned rewards based on the individual performances of both students in the dyad. The reward structure for the collaborative learning conditions encouraged higher-achieving group members to help lower-achieving group members learn the material. In other words, group-performance based rewards likely facilitated mathematical discourse within the dyads and disallowed use of the maladaptive interactional strategies described by Yackel et al. (1991).

Due to its focus on the conditions that facilitate mathematical dialogue within groups, Hurley, Boykin, and Allen (2005) provides an interesting follow-up to Ginsburg-

Block and Fantuzzo (1998), despite its use of highly artificial learning environments. The authors of this study drew a sample of fifth-grade students ($N = 78$) from two urban public schools. Participants were assigned to one of two instructional conditions: high communal or low communal.

All participants received a 20-minute researcher-administered intervention. During the intervention, students were provided with an 11-page experimenter-prepared workbook describing the use of estimation-in-multiplication strategies (i.e., treating 9×20 as 10×20). The workbook also included practice problems. The experimenter provided a brief introduction to the materials, but did not otherwise participate in the intervention, which was administered to three students at a time. Students in the high communal condition sat at a single table and shared a single workbook. After asking the students to hold hands, the researchers reminded them that they were members of the same school and community and encouraged them to work together in mastering the material. No performance-based rewards were offered to students in this condition. Students in low-communal groups were seated at individual desks and given their own sets of materials. The researchers prompted students in this condition to work individually on the material and explained that improvements based on pre-test scores would result in a reward. Student achievement was measured pre-and post-intervention using equivalent forms of a researcher-designed test. Students from high-communal groups significantly outscored peers from the low-communal condition on the posttest measure.

When providing the same curriculum to students in collaborative learning and traditional drill-and-practice instructional conditions, Ginsburg-Block and Fantuzzo

(1998) reported that the use of collaborative learning structures enhanced academic achievement. Consistent with Yackel's (1991) findings, however, the results of Hurley et al. (2005) suggest that collaborative learning is not an effective instructional technique when external stimuli (i.e., hand holding, shared materials, group-performance based rewards) facilitating within-group mathematical discourse are not present. The findings of Hurley et al. (2005), suggest that without the use of group-performance based rewards, Ginsburg-Block and Fantuzzo (1998) may have obtained different results.

The studies of Janicki and Peterson (1981) and Madden and Slavin (1983) provide an interesting test to this hypothesis as neither study provided reward-incentives for within-group collaboration (Ginsburg-Block & Fantuzzo, 1998) or environmental conditions necessitating it (Hurley et al., 2005). Not surprisingly, neither study reported significant differences in the learning rates of the third, fourth, or fifth-grade students assigned to traditional instruction or group learning conditions. For both of these studies, instruction in each instructional condition matched Stigler and Hiebert's (1997) description of traditional mathematics instruction as practiced in the United States. That is, the instruction consisted of an acquisition phase during which teachers demonstrated problem-solving procedures to students in a whole-class format and an application phase during which the students applied the newly-taught procedures to practice problems. In both studies, the only difference between the two conditions was that students in the group learning condition completed the application phase working in mixed ability groups. Thus, Janicki and Peterson's (1981) and Madden and Slavin's (1983) use of collaborative learning bears little resemblance to the recommendations of

the *Standards* (NCTM, 2000) or to the teaching experiments described above (Cobb & et al., 1991; Steencken & Maher, 2003; Yackel et al., 1991).

The majority of the experimental studies reviewed above have featured the use of collaborative learning structures in the classroom (Ginsburg-Block & Fantuzzo, 1998; Hurley et al., 2005; Muthukrishna & Borkowski, 1995). A substantial minority of these studies indicate that this instructional practice does not result in significant improvement to student achievement (Janicki & Peterson, 1981; Madden & Slavin, 1983). The collaborative learning conditions of studies demonstrating positive effects, however, featured environmental stimuli (i.e., rewards, teacher prompts, need to share materials) encouraging higher levels of within-group discourse (Ginsburg-Block & Fantuzzo, 1998; Hurley et al., 2005; Muthukrishna & Borkowski, 1995). Combined with descriptions of the interactional processes occurring within collaborative learning structures (Yackel et al., 1991), these findings tentatively suggest that collaborative learning is effective when group members receive external encouragement or training to actively collaborate in the construction of knowledge. In summary, the findings from both qualitative and quantitative research on collaborative learning support social-constructivist principles espoused by proponents of mathematics reform (i.e., encouragement of student-to-student mathematical discourse enhances learning outcomes).

As with inquiry-based instruction and collaborative learning, a hallmark of Standards-based instruction at all grade levels is the practice of situating mathematics in real-world contexts (See Hiebert, 1999; Schoenfeld, 2006; Senk & Thompson, 2003). Unfortunately, most studies comparing “realistic” mathematics to traditional instructional practices within the discipline consider the practice of situating mathematics in real-

world contexts to be a curricular issue and use experimenter-designed curricula for treatment conditions (e.g., Verschaffel & De Corte, 1997). Such studies are of obvious importance to the school administrators and curriculum specialists who are responsible for choosing curricula. The findings of these studies are less applicable to teachers whose job it is to teach the assigned curriculum to the best of their abilities.

Of greater use to teachers is Anand and Ross's (1987) curriculum-neutral study comparing the differential effectiveness of situating example problems in concrete and abstract contexts. The authors randomly assigned 96 fifth- and sixth-grade students to receive one of the three versions of the computerized lesson: concrete, personalized, and abstract. All versions of the program presented a four-step rule for dividing fractions and demonstrated the use of this rule on five example word problems. The context of the example problems differed by condition (numerical values remained constant). In the abstract condition, these problems were abstract and, in the concrete condition, the problems were posed in a real-world context (i.e., using the word "candy bar" instead of "object"). In the personalized condition, information from a biographical questionnaire was used to generate example problems featuring names and items of personal importance to the student (e.g., family members, friends, favorite food items).

Anand and Ross used a two-section posttest to assess student learning. In the first section, students completed six questions both numerically and operationally similar to the example exercises demonstrated by the program. The context of the first 6 posttest questions varied. Two questions used abstract contexts, two used concrete contexts, and two used the personal context. The second section of the posttest assessed student ability to adapt or transfer the skills presented in the intervention to

different problem-solving situations (e.g., presenting the students with $7/3 \div 5/8$ numerically rather than in the context of a word problem).

Students in the personalized condition significantly outperformed peers in the abstract condition on the first section of the test (called the context subtest), although neither group significantly differed from the concrete condition. With regards to transferring the information presented in the intervention to new settings, students in the personalized condition significantly outscored students from the other two conditions on the second section of the posttest (called the transfer subtest). Thus, this study provides confirmatory evidence for the NCTM's recommendation of situating mathematics in real-world contexts when instructing younger students.

Despite the highly artificial learning conditions (i.e., no teacher, no classmates) of the study, these findings are applicable to elementary school mathematics classrooms. Specifically, the study suggests that adding personalized information from students in the classroom to abstract example problems used to illustrate mathematical principles can benefit student learning. Such a technique can be applied to any curriculum, making the practice of situating learning within real-world contexts feasible for teachers.

The Present Study

At present, researchers have examined the efficacy of *Standards*-based instructional practices (i.e., inquiry-based instruction, collaborative learning, situating mathematics in realistic contexts) using either qualitative/quasi-experimental or random assignment/matched comparison research designs. Each design has limitations. In the former, potential selection bias and subjective outcome measures limit both the generalizability of the results and the ability to determine cause-and-effect relationships.

In the latter, the artificial learning conditions imposed on subjects limit the generalizability of the results to actual classrooms.

Interpreted in the context of the rich descriptions of Standards-based instructional practices provided by the qualitative teaching experiments, the results from the experimental literature reveal that the conditions containing central (e.g., Anand & Ross, 1987; Ginsburg-Block & Fantuzzo, 1998; Hurley et al., 2005; Muthukrishna & Borkowski, 1995) rather than superficial (Janicki & Peterson, 1981; Madden & Slavin, 1983) aspects of reform mathematics positively influenced student learning. In other words, experimental studies reported positive learning outcomes only when the instruction conditions provided contexts which supported the type of discourse community-based learning described in the qualitative literature and the *Standards* (NCTM, 2000). To qualify the above statement, the quantitative studies reporting positive-findings mentioned above did not replicate the multi-faceted instructional approaches reported in the qualitative literature. Instead, they empirically tested highly specific components of instructional approaches consistent with a central goal of reform mathematics: the facilitation of process-focussed mathematical discourse focused on the processes rather than the products of mathematical operations. For example, the approach used by Hurley et al. (2005)—holding hands, sharing one set of materials—is not specifically recommended by the *Standards* (NCTM, 2000); however, the experiment demonstrated that incorporating elements facilitating process-focused mathematical discourse into collaborative learning assignments enhanced student outcomes. Thus, as a whole, the literature suggests that the use of Standards-based instructional practices results in improved learning outcomes for students.

Large gaps, however, still exist in this literature. Studies vary by instructional intervention, mathematical topic (e.g., fractions), and age group, and are too few in number. To address the problems of the existing research base, Boaler (2008) advocated for the use of a longitudinal, regression-based quasi-experimental approach to investigating the efficacy of *Standards*-based instructional practices. Specifically, she recommended for researchers “to compare teaching approaches, not by sorting children into control and experimental groups and applying treatments, but by finding schools [or teachers] that use different approaches and studying their effectiveness” (p. 590). In applying such a design, Boaler proposed the use of regression analyses, specifically for the purpose of controlling student characteristics that influence achievement. Accordingly, controlling such variables through regression would eliminate the necessity of attempting to use random assignment within the school system.

Boaler proposed that the use of these methods will enable the type of large-scale, long-term studies needed to produce generalizable results regarding the efficacy of *Standards*-based instructional practices. She further noted that experimental studies of similar depth and breadth simply cannot be conducted within the school system. She concluded her argument for the use of these quasi-experimental methods in mathematics educational research with the statement:

The external validity of a study that does not assign students to groups may be weaker than one that does, but this is compensated by the increased ecological validity of a study that examines the natural operating of a school. Thus, quasi-experimentalists can study schools and students working in ways that are realistic and achievable by other schools, rather than ways that have been artificially created by researchers. (p. 590)

The Purpose of the Present Study

The aim of the present study is to examine the efficacy of *Standards*-based instructional practices using Boaler's (2008) proposed quasi-experimental methodology. Specifically, a hierarchical linear modeling analysis will be conducted upon a large-scale, longitudinal sample of students and their teachers drawn from the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K; National Center for Education Statistics 2001, 2002, 2004, 2006). In greater detail, these analyses will be used to examine the relationship between students' mathematical performance on a standardized instrument and their teachers' self-reported use of teaching methods consistent with the above described Standards-based instructional practices.

This study will address the following research questions:

- Does teacher use of Standards-based instructional practices positively influence classroom mathematics achievement across kindergarten, first grade, third grade, and fifth grade?
- Does the extent of prior teachers' engagement in Standards-based instructional practices influence current classroom mathematics achievement?
- What is the relative strength of teacher engagement in Standards-based instructional practices as a predictor of class-level mathematics achievement compared to the classroom demographics, such as socioeconomic status (SES), ethnicity, and prior achievement?

CHAPTER 2 METHODS

Overview

The present study consists of an analysis of pre-existing data, the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K). Therefore, the data collection methods of the ECLS-K and the data selection criteria and statistical analyses and the present study are reviewed in this chapter. The first half of the chapter presents information on the design and data collection procedures of the ECLS-K with a focus on the variables of interest to the present study. This information was abstracted from the user's manuals pertaining to the base-year, first grade, third grade, and fifth grade public use data files (NCES, 2001, 2002a, 2004, 2006). The second half of this chapter—beginning at the Statistical Analyses section—details the data selection criteria and statistical procedures used in the present study.

Data Source

The ECLS-K, sponsored by the National Center for Education Statistics (NCES) of the United States Department of Education, was designed to provide information regarding children's educational experiences between kindergarten and fifth grade. The study followed a nationally representative sample of 22,782 first-time kindergarten students from the fall of 1998 through the spring of 2004, when most of the students were in the fifth grade. Data collection occurred during the fall and spring of kindergarten (1998-1999), spring of first grade (2000), spring of third grade (2002), and spring of fifth grade (2004). Data concerning students' academic and social-emotional competencies and the characteristics of their home, classroom, and school

environments was gathered from multiple informants, including parents, teachers, school administrators, and the students themselves.

To select a nationally representative sample of kindergarten students during the 1998-1999 school year, the ECLS-K used a multistage probability sampling design. Counties or groups of counties served as the primary sampling units (PSUs). Schools within sampled PSUs and students within sampled schools served as the second and third stage units, respectively. Approximately 24 children were targeted for participation at each school. Students of Asian/Pacific Islander descent were systematically over-sampled to provide a sufficient sample size for researchers interested in this population. A more comprehensive and detailed discussion of the sampling procedure may be found in the ECLS-K Base Year Public-Use Data File User's Manual (NCES, 2001).

Due to attrition, the sample had to be "freshened" (i.e., new students were added) during the spring 2000 data-collection period to ensure that the sample was nationally representative of first grade students in the United States. The sample was not freshened during the third or fifth grade data collection periods. Therefore, the third and fifth grade samples are not representative of the general population of third and fifth grade students in the United States (NCES, 2004, 2006). Due to the aforementioned budgetary constraints, kindergarten was the only grade-level at which a nationally representative teacher sample was collected. That is, at other grade-levels teachers were included in the sample solely on the basis of whether a sampled student had been placed in their classrooms—a selection criteria incapable of providing a nationally representative teacher sample.

Teacher Questionnaire

Self-report questionnaires were distributed to the teachers of sampled students. Through these questionnaires, teachers provided information about their backgrounds (i.e., age, gender, ethnicity), classroom demographics, and the frequency with which they used certain teaching practices. In addition to this class-wide information, the questionnaires also required teachers to rate the academic and social skills of each sampled child in the classroom. These questionnaires were distributed to the current teachers of sampled students during the spring kindergarten, first-grade, third-grade, and fifth-grade data collection periods.

School Administrator Questionnaire

School administrators were also asked to complete self-report questionnaires. These questionnaires solicited information about school policies, the physical and fiscal condition of the school, available learning programs, and student and staff characteristics. These questionnaires were distributed to the current school administrators of sampled students during the spring kindergarten, first-grade, third-grade, and fifth-grade data collection periods.

Direct Child Assessment

Direct cognitive assessment of students occurred at each point of data collection. These direct cognitive assessments consisted of three sections: reading, mathematics, and science. The assessment was untimed and individually administered to each student by trained research staff. Administration of the kindergarten and first grade versions of the cognitive assessment required 50-70 minutes. Administration of the third and fifth grade versions required averages of 94 and 97 minutes, respectively. During the kindergarten year (the only year in which students were assessed on two

occasions), the fall and spring direct child assessments were scheduled in such a way as to maximize uniformity in exposure to instruction among students. For example, students assessed late in the fall were also assessed late in the spring.

Measurement of mathematics achievement

The direct cognitive assessment measured mathematics achievement using a concurrently administered two-stage approach. First, each student was given a routing test with a diverse array of item difficulty levels. Performance on the routing test determined the difficulty level of the second-stage form of the assessment: high, medium, or low. Item difficulty level was more homogenous within the second-stage forms. The two-stage assessment procedure allowed the administration of questions most appropriate to the child's current ability level, which ensured maximum measurement accuracy.

The tests used multiple choice and open-ended items to measure skill in the following content areas: (a) number sense, properties, and operations; (b) measurement; (c) geometry and spatial sense; (d) data analysis, statistics, and probability; and (e) patterns, algebra, and functions. The National Assessment Governing Board's (NAGB) publication *Mathematics Frameworks of the 1996 National Assessment of Educational Progress (1996)*—derived from the 1989 *Standards of the NCTM*—was used to determine the five content strands and the emphasis put on the differing strands at each assessment occasion. The NAEP-defined cognitive processes, conceptual understanding (i.e., knowing what) and procedural knowledge (i.e., knowing how), were assessed in each of the strands (NCES, 2001, 2002, 2004, 2006). Paper and pencil were used by students during certain portions of the assessment and several items featured manipulative materials to aid in problem-solving.

In addition to the content strands, items from the kindergarten and first grade versions of the test can also be grouped into five proficiency level clusters. The NCES (2002a) described these proficiency levels as follows:

(1) identifying some one-digit numerals, recognizing geometric shapes, and one-to-one counting up to ten objects; (2) reading all one-digit numerals, counting beyond ten, recognizing a sequence of patterns, and using nonstandard units of length to compare the size of objects; (3) reading two-digit numerals, recognizing the next number in a sequence, identifying the ordinal position of an object, and solving a simple word problem; (4) solving simple addition and subtraction problems; and (5) solving simple multiplication and division problems and recognizing more complex number patterns. (p. 39)

The third-grade assessment included four proficiency level clusters. Levels 4 and 5 from the kindergarten and first grade assessment were retained, and two new levels were added. The NCES (2004) described the proficiency levels measured in this test as follows:

(1) solving simple addition and subtraction problems; (2) solving simple multiplication and division problems and recognizing more complex number patterns; (3) demonstrating understanding of place value in integers to hundreds place; and (4) using knowledge of measurement and rate to solve word problems. (p. 49)

The fifth-grade assessment included five proficiency level clusters. Levels 2, 3, and 4 from the third grade assessment were retained and two new levels were added: fractions and word problems requiring knowledge of volume and area (NCES, 2006).

Test items were selected by elementary school teachers and curriculum specialists and item writers from the Educational Testing Service (ETS). Items were selected from commercially available materials, such as the Peabody Individual Achievement Test-Revised (Markwardt, 1989), the Primary Test of Cognitive Skills (Huttenlocher & Levine, 1990), The Test of Early Mathematics Ability (Ginsburg & Baroody, 1990), and the Woodcock-Johnson Tests of Achievement-Revised (Woodcock

& Bonner, 1989). Next, the selected items were field tested to determine their psychometric characteristics. The final form of the mathematics assessment was constructed using items with psychometrically-appropriate characteristics as determined through this process.

Score format

The ECLS-K presents test scores in multiple formats: number-right scores (i.e., the number of correct answers produced), item response theory (IRT) scale scores, and standardized scores (T scores). The latter two formats are both derived using IRT. Specifically, test item characteristics (i.e., difficulty, discriminating ability) and the student's own pattern of correct and incorrect responses are used to define an ability estimate, theta. Theta represents a point estimate of the student's performance on an ability continuum and serves as the basis for the criterion-referenced (IRT scale scores) and norm-referenced (standardized scores) scores. The criterion-referenced scores are less distorted by omitted answers than number-right scores and better compensate for the influence of guessing (e.g., a low-ability student guessing several difficult questions correctly). Also, because the IRT-scoring process creates a continuous ability scale, longitudinal measurement of achievement gains are possible although test items differ across data collection occasions.

While the IRT scale scores and standardized scores share the above advantages over the number-right score format, they provide very different information. As described by the NCES (2002a), T scores "provide information on status compared to children's peers, whereas the IRT scale scores...represent status with respect to achievement on a particular criterion set of test items" (p. 59). Thus, IRT scale scores were selected as the unit of analysis for the current study because their criterion-

referenced nature allows for longitudinal analysis. Longitudinal use of T scores would indicate only the extent to which an individual ranking relative to others changes over time.

Reliability estimates for the mathematics portion of the direct cognitive assessment were determined at each data collection occasion. The NCES (2006) states that “for the IRT-based scores, the reliability of the overall ability estimate, theta, is based on the variance of repeated estimates of theta compared with total sample variance” (p. 103). These reliabilities were .89, .91, .92, .94, and .94 for the fall-kindergarten, spring-kindergarten, first grade, third grade, and fifth grade collections, respectively.

Statistical Analyses

In the current study, data from the spring of kindergarten, Grade 1, Grade 3, and Grade 5 data-collection periods were examined. Because the goal of the study was to assess the influence of teacher instructional practices on student achievement, data from the fall of kindergarten was not examined due to students’ limited exposure to instruction at this time.

Analytic Samples

Because instructional variables may differentially impact student achievement at different grades-levels, retained and accelerated students were omitted from each analytic sample. Thus, data from each collection period were limited to students at a specific grade level (e.g., the spring 2004 sample was limited to students in the fifth grade). Additionally, students not linked to a teacher were also excluded at each grade level. Finally, at the kindergarten level, students who changed teachers between their fall and spring assessments were also eliminated to more accurately estimate teacher effects as were students enrolled in half-day kindergarten programs. After applying

these initial selection criteria, the kindergarten, first-grade, third-grade, and fifth-grade samples consisted of 9,854; 15,878; 13,040; and 10,293 students, respectively.

Final analytic samples at each grade level were derived using a further series of filters. First, students missing data on mathematics achievement scores from the current grade and from the immediately preceding data collection occasion (e.g., the third-grade mathematics score for a fifth-grade student) were omitted. Omissions at the first, third, and fifth-grade levels also resulted from missing data on any variables representing prior exposure to *Standards*-based practices. Next, students missing classroom-level data (i.e., instructional scales or demographic variables) were excluded. If that information was available from one of the other sampled students in the classroom, however, the student was retained in the sample.

Finally, to accurately estimate teacher effects, each grade-level sample was restricted to students in classrooms with three or more sampled students (Jennings & DiPrete, 2008). Other investigations of teacher-level effects on student achievement in the ECLS-K have limited the sample to teachers with only two sampled students in the classroom (Croninger, Rice, Rathbun, & Nishio, 2007) or have not applied any such restrictions (Milesi & Gamoran, 2006). The latter approach is problematic because, in many cases, estimates of class achievement are based on the score of only one student. Jennings and DiPrete applied the most conservative qualifications for estimating teacher-level effects using the ECLS-K data in limiting their sample to teachers with at least three sampled students in their classrooms. At the kindergarten, first-grade, third-grade, and fifth-grade levels, teachers have an average of 5.37, 4.74, 4.50, and 3.69 sampled students per classroom, respectively.

The final sample at the spring-kindergarten level consisted of 4,841 students, 902 teachers, and 335 schools. The first-grade sample included 5,999 students, 1,265 teachers, and 583 schools. The third and fifth-grade samples consisted of 3,553 students, 789 teachers, and 450 schools and 410 students, 111 teachers, and 100 schools, respectively.

Assessments and Measures

Criterion measure

Student performance on the mathematics portion of the direct cognitive assessment served as the dependent variable. More specifically, because separate analyses were conducted at each grade level (kindergarten, first grade, third grade, and fifth grade), mathematics performance from the current-grade level of analysis served as the dependent variable.

Control variables

Demographic measures. Demographic characteristics such as the presence of minority students and those receiving free and reduced lunch are associated with student achievement and are often used as proxies for socioeconomic status (Beaver, Wright, & Maume, 2008; Lee & Bryk, 1989; Rimm-Kaufman, Fan, Chiu, & You, 2007; Xue & Meisels, 2004). To control for their influence on student achievement, continuous variables representing classroom percentages of minority students and students receiving free and reduced lunch were included in the analysis. The percentage of minority students in the classroom was provided by NCES at each sampled grade level.

Unfortunately, the ECLS-K database only provides information on class-wide percentages of students receiving free and reduced lunch for the third-grade sample. Information regarding school-wide percentages of these students, however, is available

at all sampled grade levels. In their hierarchical linear modeling (HLM) analysis of ECLS-K data, Beaver, Wright, and Maume (2008) used these school-wide measures as the classroom-level variables. This decision resulted in multiple classrooms within each school having identical values on these variables. The authors, however, stated that because most of the schools contained very few sampled classrooms using these school-wide percentages at the classroom-level of analysis would not unduly bias the results. To assess for bias, Beaver et al. re-ran their models without the variable and produced identical results. This led them to conclude that “including the school-level measures thus helped to control for effects that may confound the relationship between the school classroom measure and low self-control [the dependent variable]” (p. 180).

Similar to Beaver et al. (2008), most of the schools included in the present study contained very few sampled classrooms (range = 1.1-2.7 sampled classrooms per school). Therefore, the above findings of Beaver et al. suggest that using the school-wide percentages for students receiving free or reduced lunch at the classroom-level is appropriate for the current analysis. Further support for the appropriateness of this procedure was obtained by alternating class-wide and school-wide versions of this variable within models at the third grade level. The results (available upon request), confirm that substituting school-wide percentages for students receiving free or reduced lunch for the classroom-level percentages produced no substantial changes to the model. Based on this finding and the findings of Beaver et al. (2008), the decision was made to use school-wide percentages for students receiving free or reduced lunch at the classroom-level in the current study.

Previous achievement. Previous scholastic achievement is highly predictive of current achievement. To control for previous mathematics achievement at the classroom level, class-wide aggregates of student mathematics scores from the immediately preceding data collection were created. Specifically, scores from the fall of kindergarten were used when conducting analyses on the spring kindergarten sample, scores from the spring of kindergarten were used for the first grade assessment, scores from first grade were used for the third grade assessment, and scores from third grade were used for the fifth grade assessment.

Aggregated rather than individual scores served as control variables because teacher use of *Standards*-based instructional practices varies between classrooms rather than individuals. Therefore, to accurately estimate the influence of these practices on achievement, we want the classrooms to be as similar as possible with respect to other characteristics associated with the outcome measure (i.e., previous achievement, percentage of minority students, and percentage of students receiving free or reduced lunch).

Instructional scales

Student exposure to *Standards*-based instructional practices within the classroom was the variable of interest in the present research.

Item pools. At each sampled grade-level, one or more composite scale(s) representing teacher usage of *Standards*-based instructional practices was created from items on the instructional practices section of the teacher questionnaire. This section of the questionnaire solicited information concerning the frequency with which teachers engaged in certain instructional practices. In the subsection pertaining to mathematics (sample copies of which are included in Appendices A, B, C, and D),

teachers answered the question, “How often do children in this class do each of the following mathematics activities?” for each of a number of listed instructional activities.

Kindergarten and first-grade teachers responded to these items using a six-point rating scale ranging from 1 (never) to 6 (daily). Third and fifth-grade teachers responded using a four-point rating scale ranging from 1 (almost every day) to 4 (never or hardly ever). Because the direction of the scales differ in their progression from less to more frequent between the kindergarten and first grade, and third and fifth grade questionnaires, responses to the latter two questionnaires were transformed so a response of 1 indicated infrequent use of a technique and a response of a 4 indicated frequent use. The number of instructional activities described in the mathematics subsection varied between 12 and 19 depending on data-collection period, but always contained several items pertaining to the use of *Standards*-based classroom practices. The form and wording of these specific items, however, varied across data waves.

Based upon the review of quantitative and qualitative literature regarding NCTM-recommended instructional practices along with the NCTM sources themselves (1989, 2000), a list of items representing *Standards*-based instructional practices was compiled at each grade level. At the kindergarten level, the following items were selected: (a) explain how a mathematics problem is solved, (b) work on math-problems that reflect real-life situations, (c) solve mathematics problems in small groups or with a partner, (d) work in mixed achievement groups on mathematics activities, and (e) peer tutoring. The same five items were selected at the first grade level along with a new item asking whether students work on problems for which there are several appropriate methods or solutions.

At the third-grade level, six items were considered to represent *Standards*-based practices. These items assessed the frequency with which students (a) solve mathematics problems in small groups or with a partner, (b) write a few sentences about how to solve a mathematics problem, (c) talk to the class about their mathematics work, (d) write reports or do mathematics projects, (e) discuss solutions to mathematics problems with other children, and (f) work on and discuss mathematics problems that reflect real-life situations. At the fifth-grade level, the following items were considered to represent *Standards*-based practices: (a) write a few sentences about how to solve a mathematics problem, (b) solve mathematics problems in small groups or with a partner, (c) discuss solutions to mathematics problems with other children, and (d) work on and discuss mathematics problems that reflect real-life situations. Tables 2-1, 2-2, 2-3, and 2-4 show the means, standard deviations, and correlations for the individual items at the kindergarten, first-grade, third-grade, and fifth-grade levels, respectively.

Scale construction. Exploratory factor analyses were conducted on the above items to assess the underlying structure of the selected items at each grade level. It was hypothesized that at least one latent variable representing teacher use of *Standards*-based instructional practices would result. To obtain results maximally representative of the instructional practices of teachers at the kindergarten, first-grade, third-grade, and fifth-grade levels, factor analyses were conducted on the total sample of teachers at each grade level, as opposed to the subset of teachers matched to students meeting the study's inclusion criteria. A total of 3,305 teachers from the spring of 1999, 5,046 from the spring of 2000, 5,017 from the spring of 2002, and 3,842 from the spring of 2004 received the teacher practices questionnaire. Teachers missing data on all of the

selected items were excluded from the analyses at each grade level, as were teachers exhibiting inconsistent responses across students (i.e., responding with a two on the questionnaire for one student and with a three for another student). After applying these criteria, factor analyses were conducted on 3,059; 3,720; 3,489; and 2,438 teachers at the kindergarten, first-grade, third-grade, and fifth-grade levels, respectively.

The latent factor structure of the item scores was examined via exploratory factor analysis using the Mplus version 5 statistical software program (Muthen & Muthen, 2004). The estimation method was diagonally weighted least squares with robust estimation of standard errors. This method uses data from students who have scores on all items and from students who have incomplete data and is appropriate for ordinal item scores. Mean adjusted goodness of fit chi-square statistics were calculated.

Among the multiple solutions produced at each grade level, the most appropriate was identified based on goodness-of-fit indices and the interpretability of the geomin (oblique)-rotated factor loadings. Based on Hu and Bentler's (1999) recommendations, the current study used a two-index presentation strategy to assess the degree of fit between the model and the sample. This strategy specifies the use of the standardized root mean-square residual (SRMR), an absolute fit index, along with an incremental fit index in model evaluation. With regard to the latter, the present research used Bentler's Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) as the indices of incremental fit. According to Hu and Bentler (1999), TLI or CFI scores of .95 or greater and SRMR scores of 0.09 or below indicate good fit between the model and data. Table 2-5 presents the goodness-of-fit indices for the proposed models at each grade level.

Regarding model interpretability, in solutions with more than one factor, only items with their highest loading on a given factor were considered in the interpretation of that factor (See Hamilton & Guarino, 2005). Additionally, because factor loadings exceeding .30 are considered to be meaningful (Floyd & Widaman, 1995, p.294), items with a loading below .30 were not considered in the interpretation of a factor.

At the kindergarten level, the factor analysis produced one-and two-factor solutions. Only the two-factor solution, however, met Hu and Bentler's (1999) goodness-of-fit criteria (See Table 2-5) and was also easily interpretable. Table 2-6 presents the factor loadings for the two-factor solution. The two items comprising the first factor, labeled Problem Solving (PS), represent teacher use of realistic-learning contexts and emphasizing of problem-solving processes over products in class discussion. The second factor, labeled Collaborative Learning (CL), consisted of three items measuring the frequency with which the teacher allows students to engage in collaborative-learning activities.

As presented in Table 2-5, both the one-and two-factor solutions met Hu and Bentler's (1999) criteria at the first-grade level. The two-factor solution was selected because the TLI, CFI, and SRMR indicated a superior fit of the model to the data and the two solutions were equally interpretable. Table 2-7 presents the factor loadings for the two-factor solution. The factor structure of the solution was similar to that found at the kindergarten level and the factors were again labeled PS and CL. At this grade level, however, an additional item pertaining to problem solving loaded on the PS factor. In both the kindergarten and first-grade samples, the PS and CL factors were moderately correlated with one another ($r = .65$ and $r = .61$, respectively).

At the third-grade level, the goodness-of-fit indices suggested a two-factor solution. Despite meeting Hu and Bentler's (1999) criteria, however, the factor loadings from the two-factor solution were uninterpretable and the presence of negative residual variances indicated an over-extraction of factors (Muthen, 2005). As the one-factor solution demonstrated adequate fit to the data (CFI of .98, TLI of .97, SRMR of .05) and yielded interpretable results, it was selected for use over the two-factor solution. The six items representing the single factor, labeled *Standards-Based Instructional Practices* (SBIP), represented teacher emphasis on student-centered instruction, realistic-learning contexts, and collaborative learning. Table 2-8 presents the factor loadings for the one-factor solution.

At the fifth-grade level, MPLUS was unable to estimate solutions with more than one factor. The one-factor model, however, demonstrated adequate fit to the data with a CFI of .99, a TLI of .96, and a SRMR of .04. The four items comprising the single factor, labeled *Standards-based instructional practices* (SBIP), represent teacher emphasis on student-centered instruction, realistic-learning contexts, and collaborative learning. Table 2-9 presents the factor loadings for this solution.

The *Standards-based instruction* scale(s) at each grade level were created by taking the mean of teacher responses to the items loading on each factor. In situations with more than one factor, only items with their highest loading on a given factor were averaged. This method of combining information from a set of items into a single scale is commonly used and the resulting scale is easier to interpret than when constructed from factor scores (Hamilton & Guarino, 2005; Hausken & Rathbun, 2004). Previous researchers creating scales from the instructional practices questionnaire of the ECLS-

K (Guarino, Hamilton, Lockwood, & Rathbun, 2006; Hamilton & Guarino, 2005; Hausken & Rathbun, 2004; Milesi & Gamoran, 2006) used similar methods of scale creation.

Missing data. Missing data from the selected teacher questionnaire items represented a major obstacle to scale construction. In two separate studies involving the creation of instructional practices scales from ECLS-K data, Hamilton and Guarino dealt with missing teacher response data by imputing the average of the items that had been answered for the given scale (Guarino et al., 2006; Hamilton & Guarino, 2005). That is, “for teachers who answered some, but not all, items on a given scale, the average score of those items that they answered was used” (Guarino et al., 2006, p.12). Use of this approach, however, can lead to the generation of scale scores which may be inconsistent with a teacher’s true practices. For example, a teacher answering only one of the five items loading on a scale would have that score generalized to represent the frequency of his or her engagement in several instructional practices. Hausken and Rathbun (2004) adopted a more conservative approach to dealing with missing teacher responses in composite scale creation by excluding teachers missing more than one item on a given scale from the study. Due to the presence of one two-item scale and several three-item scales in the present study, however, the most conservative approach to aggregation (listwise deletion) was adopted.

Hierarchical Linear Modeling

The primary aim of the current study was to describe the influence of *Standards-*based instructional practices on mathematics achievement. Given the nested nature (i.e., students within classrooms) of the dataset, the hierarchical linear modeling (HLM) methodology (Raudenbush & Bryk, 2002) was selected for the analyses. Using HLM,

the mathematics scores of individual students within a given classroom are considered to vary around the classroom mean. Similarly, the classroom means are considered to vary around the grand mean (i.e., the mean of all class-mean scores). In other words, HLM partitions the total variation in mathematics achievement into two levels: within-classroom variance (Level 1) and between-classroom variance (Level 2). The focus of the current study is at this second level.

At each grade level, a preliminary analysis was conducted to estimate an unconditional model (Baseline Model). This model provided information about the grand mean of mathematics achievement and proportions of within-and between-classroom variance. To determine the confidence interval for classroom mean scores, the equation $\gamma_{00} \pm 1.96(\tau_{00}^{1/2})$ is used. To determine the proportion of classroom-level to individual-level variance in student achievement scores, an intraclass correlation (ICC) is computed. The ICC represents the ratio of classroom-level variance to the total variance across both students and classrooms and is computed through the following equation: $ICC = \tau_{00} / (\tau_{00} + \sigma^2)$. In this equation, σ^2 represents the variance of r_{ij} (a random effect representing the deviation of student $_{ij}$'s mathematics score from the mean of classroom $_j$) and τ_{00} represents the variance of μ_{0j} (a random effect representing the deviation of classroom $_j$'s mathematics score from the grand mean).

Expanding the baseline model to include the instructional scale(s) at the classroom level permitted the determination of whether *Standards*-based instruction was predictive of mean class achievement (Model 1). The resulting variable coefficients were standardized to allow for comparison among variables as to the increase or decrease a one SD change in a particular variable would predict in student achievement

scores. Variable coefficients were standardized by multiplying the coefficient by its own SD and dividing it by the SD of the individual student mathematics scores.

Next, classroom-level averages of sampled students' previous exposure to Standards-based instructional practices were introduced to the model (Model 2). This model estimated the influence of a class's previous exposure to *Standards*-based instructional practices on mathematics achievement while controlling for the influence of current teacher practices. Due to a lack of exposure to previous instruction, this model was not applied to the kindergarten-level sample. Applied at the other grade levels, this model required the inclusion of two, four, and five variables for the first-grade, third-grade, and fifth-grade samples, respectively. For example, when examining the influence of prior exposure to NCTM-recommended practices at the fifth-grade level, the classroom-level averages of sampled students' a) kindergarten teachers' scores on the PS and CL scales, b) first-grade teachers' scores on the PS and CL scales, and c) third-grade teachers scores on the SBIP scale in were included in Model 2.

To examine the influence of past and present exposure to these instructional practices on achievement while controlling for previous school achievement, mean class mathematics scores from the immediately preceding data collection occasion was added to the second level of the analysis (Model 3). In Model 4, the influence of the demographic variables (percentages of minorities and students receiving free or reduced lunch) was also controlled.

In conclusion, Model 1 examined the influence of NCTM-recommended instructional practices on achievement; Model 2 examined the influence of previous exposure to these techniques on current achievement; and Models 3 and 4 determined

whether the influence of current and past exposure to these teaching practices remained after controlling for the demographic characteristics of the classroom. All models were applied to each grade-level sample with the exception of Model 2. Additionally, all variables lacking a meaningful origin were grand-mean centered to aid interpretability. Furthermore, any non-significant variables were not included in subsequent models. Finally, equations for all models and explanations of model terms at each grade level are provided in Appendices E, F, G, and H. All models were estimated using the HLM 6.0 (Raudenbush, Bryk, & Congdon, 2006).

Sample Weights

Due to the complex stratified sampling design of the ECLS-K, some students, teachers, and schools had a higher probability of selection than others. This complex sample design violates the assumption of independence of observations that underlies parametric statistics (See Hahs-Vaughn, 2005). To correct for this problem, the ECLS-K project staff provides sample weights to be applied to the datasets. Use of these weights also corrects for nonresponse rates and the oversampling of Asian students and private schools. Failure to apply the weights can result in incorrect parameter estimates or standard errors. When applied, the weights produce nationally representative estimates for children who attended kindergarten in the fall of 1998 (NCES, 2001).

In the present study, the child-level, cross-sectional sample weights, provided by ELCS-K staff, were applied at the student-level of the appropriate sample. Specifically, the weights labeled C2CW0, C4CW0, C5CW0, and C6CW0 were used at the kindergarten, first grade, third grade, and fifth grade samples, respectively. The above weights were designed for use in cross-sectional analyses of information (i.e., math

scores, answers to teacher questionnaire items) provided by both students and teachers. In addition, the teacher-level, cross-sectional sample weight (B2TW0)—provided by the ECLS-K staff—was applied at the teacher level of the multi-level model used for the kindergarten sample. This procedure allowed the production of coefficient estimates representative of the national population of kindergarten teachers. Sample weights are not provided for a teacher-level analysis after the base-year of the study because the ECLS-K did not seek to provide a nationally representative sample of teachers in the first, third, and fifth grades.

Table 2-1. Item means, standard deviations and correlations: Kindergarten sample ($n = 3,059$)

Item	M (SD)	Item 1	Item 2	Item 3	Item 4	Item 5
1. Explain how a mathematics problem is solved	3.92 (1.56)	1.00				
2. Solve mathematics problems in small groups or with a partner	3.52 (1.51)	0.45	1.00			
3. Work on math-problems that reflect real-life situations	3.92 (1.45)	0.56	0.53	1.00		
4. Work in mixed achievement groups on mathematics activities	4.16 (1.76)	0.33	0.45	0.38	1.00	
5. Peer tutoring	3.12 (1.79)	0.35	0.48	0.38	0.45	1.00

Table 2-2. Item means, standard deviations and correlations: First-grade sample ($n = 3,720$)

Item	M (SD)	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
1. Explain how a mathematics problem is solved	4.91 (1.12)	1.00					
2. Solve mathematics problems in small groups or with a partner	4.06 (1.27)	0.35	1.00				
3. Work on math-problems that reflect real-life situations	4.43 (1.16)	0.49	0.49	1.00			
4. Work in mixed achievement groups on mathematics activities	4.16 (1.59)	0.28	0.49	0.42	1.00		
5. Peer tutoring	3.65 (1.58)	0.29	0.47	0.34	0.43	1.00	
6. Work on problems for which there are several appropriate methods or solutions	3.78 (1.44)	0.45	0.44	0.49	0.41	0.40	1.00

Table 2-3. Item means, standard deviations and correlations: Third-grade sample ($n = 3,489$)

Item	M (SD)	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
1. Solve mathematics problems in small groups or with a partner	1.92 (0.76)	1.00					
2. Write a few sentences about how to solve a mathematics problem	2.43 (0.92)	0.33	1.00				
3. Talk to the class about their mathematics work	2.07 (1.03)	0.32	0.42	1.00			
4. Write reports or do mathematics projects	3.58 (0.65)	0.32	0.39	0.35	1.00		
5. Discuss solutions to mathematics problems with other children	1.93 (0.88)	0.47	0.39	0.60	0.29	1.00	
6. Work and discuss mathematics problems that reflect real-life situations	1.91 (0.84)	0.34	0.35	0.47	0.29	0.65	1.00

Table 2-4. Item means, standard deviations and correlations: Fifth-grade sample ($n = 2,438$)

Item	M (SD)	Item 1	Item 2	Item 3	Item 4
1. Solve mathematics problems in small groups or with a partner	1.83 (0.83)	1.00			
2. Write a few sentences about how to solve a mathematics problem	5.46 (0.93)	0.35	1.00		
3. Discuss solutions to mathematics problems with other children	1.82 (0.88)	0.63	0.48	1.00	
4. Work and discuss mathematics problems that reflect real-life situations	1.88 (0.79)	0.37	0.47	0.63	1.00

Table 2-5. Goodness of fit indices for exploratory factor analyses

Grade	Factors Extracted	Bentler's Comparative Fit Index	Tucker-Lewis Index	Standardized root mean-square residual.
Kindergarten	1	0.97	0.40	0.05
	2	1.00	1.00	0.00
First	1	0.98	0.96	0.04
	2	0.99	0.99	0.02
Third	1	0.98	0.97	0.05
	2	0.99	0.99	0.01
Fifth	1	0.99	0.96	0.04

Table 2-6. Factor loadings for the two-factor solution at the kindergarten level

Item	Problem Solving	Collaborative Learning
Explain how a mathematics problem is solved	0.61*	0.09
Solve mathematics problems in small groups or with a partner	0.35	0.46*
Work on math-problems that reflect real-life situations	0.84*	-0.01
Work in mixed achievement groups on mathematics activities	0.09	0.58*
Peer tutoring	-0.01	0.72*

Note. An asterisk appears besides each item's highest loading.

Table 2-7. Factor loadings for the two-factor solution at the first-grade level

Item	Problem Solving	Collaborative Learning
Explain how a mathematics problem is solved	0.73*	0.00
Solve mathematics problems in small groups or with a partner	0.08	0.69*
Work on math-problems that reflect real-life situations	0.48*	0.32
Work in mixed achievement groups on mathematics activities	0.01	0.68*
Peer tutoring	-0.01	0.64*
Work on problems for which there are several appropriate methods or solutions	0.39*	0.36

Note. An asterisk appears besides each item's highest loading.

Table 2-8. Factor loadings for the one-factor solution at the third-grade level

Item	Standards-Based Instructional Practices
Solve mathematics problems in small groups or with a partner	0.53
Write a few sentences about how to solve a mathematics problem	0.55
Talk to the class about their mathematics work	0.69
Write reports or do mathematics projects	0.48
Discuss solutions to mathematics problems with other children	0.86
Work and discuss mathematics problems that reflect real-life situations	0.71

Table 2-9. Factor loadings for one-factor solution at the fifth-grade level

Item	Standards-Based Instructional Practices
Solve mathematics problems in small groups or with a partner	0.64
Write a few sentences about how to solve a mathematics problem	0.57
Discuss solutions to mathematics problems with other children	0.93
Work and discuss mathematics problems that reflect real-life situations	0.69

CHAPTER 3 RESULTS

Results of the HLM analyses are provided below. Due to the cross-sectional nature of these analyses, the results for each grade level are presented separately.

Kindergarten

Results of the kindergarten-level analyses are presented in Table 3-1 (fixed effects) and Table 3-2 (random effects). Estimation of the baseline model revealed significant differences among teachers in the mean mathematics achievement of their students, $\chi^2(901, n = 902) = 2015.92, p < .001$. Calculation of the ICC indicated that 20% of the total variance in mathematics achievement occurs between classrooms and 80% occurs between students within classrooms. In other words, four fifths of the achievement variance is attributable to individual variables and one fifth is attributable to classroom variables

According to this model, a score of 27.09 represents the grand mean of mathematics achievement at the kindergarten level. This coefficient has a standard error of 0.20. The confidence interval for classroom means was calculated by the equation $\gamma_{00} \pm 1.96(\tau_{00}^{1/2})$. Based on this calculation, it is expected that 95% of the class means fall between 19.57 and 34.62 points on the achievement scale.

The fixed effects of Model 1 indicate that student mathematics achievement is significantly associated with teacher use of problem-solving instruction ($\gamma_{01} = 0.41, p < .05$), but not with use of collaborative learning structures ($\gamma_{02} = -0.05, p = .80$). To standardize γ_{01} , the coefficient was multiplied by its SD (1.24) and divided by the SD of individual student mathematics scores (8.59). The standardized coefficient of 0.06

indicates that a one SD increase in the PS variable predicts a 0.06 SD increase in student mathematics achievement.

After accounting for teacher use of *Standards*-based instructional practices in the model, significant differences continued to exist between teachers in the mean mathematics achievement of their students, $\chi^2(899, n = 902) = 1999.57, p < .001$. Importantly, the variance (τ_{00}) represents a conditional variance in Model 1 and all subsequent models. That is, in Model 1, τ_{00} represents the variance in classroom-level mathematics scores after controlling for teacher use of *Standards*-based instructional practices (PS and CL). As shown in Table 3-4, the residual variance is slightly smaller in Model 1 ($\tau_{00} = 14.53$) than in the baseline model ($\tau_{00} = 14.74$). These coefficients can be compared to determine the reduction in the classroom-level score variance resulting from the inclusion of the instructional scales by using the following equation:

$$[\tau_{00}(\text{baseline}) - \tau_{00}(\text{subsequent model})] / \tau_{00}(\text{baseline}). \quad (2-1)$$

Equation 1 demonstrates that variation in teacher use of student-centered instructional and realistic instructional contexts accounts for 1% of the classroom-level variance.

In Model 2, use of student-centered and realistic instruction ($\gamma_{01} = 0.22, p < .01$) remained significantly associated with the outcome variable; however, its strength as a predictor was reduced. Specifically, the standardized coefficient indicated that a one SD increase in the PS variable predicts a 0.03 SD increase in student mathematics achievement after controlling for prior achievement compared to the 0.06 SD increase prior to controlling for prior achievement.

Both the fixed and random effects of Model 2 demonstrated that the mean mathematics score of a class in the fall of kindergarten functions as a strong predictor of

that class's score in the spring of kindergarten ($\gamma_{02} = 1.09, p < .01$). Specifically, a one SD increase in prior achievement predicted a 0.53 SD increase in the dependent variable (DV). Furthermore, no significant differences remained among teachers in the mean mathematics achievement of their students following the inclusion of prior achievement in the model, $\chi^2(897, n = 902) = 642.82, p > .50$. As shown in Table 3-4, the residual variance for Model 2 ($\tau_{00} = 0.12$) was greatly reduced from the baseline model ($\tau_{00} = 14.91$). Using Equation 1, the inclusion of prior achievement and PS in the model accounted for 99% of the classroom-level variance in mathematics scores.

Based on the results of re-running Model 2 (not shown) excluding the PS instructional scale, previous achievement alone accounted for 99% of the classroom-level variance. This finding indicates negligible contributions from any other classroom-level variables to the prediction of mathematics achievement in kindergarten after controlling for previous achievement.

The fixed effects of Model 3 confirmed this prediction. Previous achievement ($\gamma_{02} = 0.99, p < .01$) was again the strongest predictor of mathematics scores in the spring of kindergarten with a 1.00 SD change in the IV predicting a 0.49 SD change in the DV. The contributions of the percentage of minority students in the classroom ($\gamma_{04} = 0.02, p < .01$) and the PS instructional scale ($\gamma_{01} = 0.33, p < .01$) to mathematics achievement were significant, but weak, with a one SD change in either variable failing to predict more than a 0.08 SD change in the DV. Finally, the percentage of students receiving free and reduced lunch was not significantly associated with student mathematics scores ($\gamma_{04} = 0.02, p = .52$).

First Grade

Table 3-3 and Table 3-4 present the results of the first-grade analyses. Estimation of the baseline model revealed significant differences among teachers in the mean mathematics achievement of their students, $\chi^2(1264, n = 1265) = 3020.26, p < .01$. Computing the ICC indicated that 23% of the total variance in achievement exists between classrooms. The grand mean of mathematics scores for first-grade students was 43.90 (SE = 0.16) with 95% of the classroom means predicted to fall between 36.03 and 51.77 points on the achievement scale.

The fixed effects of Model 1 indicate that student mathematics achievement was significantly associated with teacher scores on the PS ($\gamma_{01} = 1.24, SE = 0.19$) and CL scales ($\gamma_{02} = -0.70, SE = 0.16$). To standardize the coefficient values, γ_{01} and γ_{02} were multiplied by their own SDs (0.99 and 1.15, respectively) and divided by the SD of individual student mathematics scores (8.51). The standardized coefficient of 0.14 for γ_{01} indicates that a one SD increase in the PS variable predicts a 0.14 SD increase in student mathematics achievement, whereas a one SD increase in the CL variable predicts a 0.09 decrease.

After accounting for teacher use of *Standards*-based instructional practices in the model, significant differences in mathematics achievement still existed among classrooms, $\chi^2(1262, n = 1265) = 2924.30, p < .001$. As shown in Table 3-4, the residual variance is slightly smaller in Model 1 ($\tau_{00} = 15.22$) than in the baseline model ($\tau_{00} = 16.14$). Using Equation 1, we find that teacher usage of *Standards*-based practices accounts for 6% of the classroom-level variance in achievement scores.

The addition of variables representing previous (kindergarten) exposure to *Standards*-based instructional practices to the model failed to increase this percentage

although the p-values of both variables indicated significance (See Table 3-4). This indicated that exposure to realistic and student-centered instruction or collaborative learning environments during kindergarten were weak predictors of first grade mathematics achievement.

The fixed and random effects of Model 3 demonstrated that the mean mathematics score of a class in the spring of kindergarten functions as a strong predictor of the class's score in the spring of the first-grade year ($\gamma_{05} = 0.83$, $SE = 0.02$). Specifically, a one SD increase in prior achievement predicted a 0.51 SD increase in the DV. Use of student-centered and realistic instruction ($\gamma_{01} = 0.50$, $SE = 0.11$) and collaborative learning ($\gamma_{02} = -0.29$, $SE = 0.10$) remained significantly associated with the outcome variable; however, the strength of this association was reduced after controlling for previous achievement. Specifically, a one SD increase in the PS variable now predicted a 0.06 SD increase in student mathematics achievement compared to the previous 0.14 SD increase. Similarly, a one SD increase in the CL variable now predicted a 0.04 SD decrease in student mathematics achievement compared to the previous 0.09 SD decrease. After controlling for prior achievement, the variables representing previous (kindergarten) exposure to *Standards*-based instructional practices were no longer significant and were excluded from subsequent models.

After accounting for exposure to *Standards*-based instructional practices and prior achievement, no significant differences remained among teachers in the mean mathematics achievement of their students, $\chi^2 (1259, n = 1265) = 1150.41, p > .500$. As shown in Table 3-4, the residual variance is much smaller in Model 3 ($\tau_{00} = 0.08$) than in

the baseline model ($\tau_{00} = 16.14$). Using Equation 1, the variables included in Model 3 accounted for 99% of the classroom-level variance in mathematics scores.

Based on the results of re-running Model 3 (not shown) excluding all instructional scales, previous achievement alone accounted for 99% of the classroom-level variance. This finding indicates negligible contributions from any other classroom-level variables to the prediction of mathematics achievement in kindergarten after controlling for previous achievement.

The fixed effects of Model 4 confirmed this prediction. With previous achievement in the model, a one SD change in any of the instructional practice or demographic variables failed to predict a change of even one-tenth of a SD in the DV.

Third Grade

Tables 3-5 and 3-6 present the results of the third-grade analyses. Estimation of the baseline model revealed significant differences among teachers in the mean mathematics achievement of their students, $\chi^2(788, n = 789) = 1633.84, p < .01$. Computing the ICC indicated that variance among class-mean scores accounted for 19% of the total variance in mathematics achievement. The grand mean of mathematics scores for third-grade students was 87.96 (SE = 0.39) with 95% of the classroom means predicted to fall between 71.53 and 102.23 points on the achievement scale.

The fixed effects of Model 1 indicate that third grade mathematics achievement was significantly associated with teacher use of *Standards*-based instructional practices ($\gamma_{01} = -1.79, SE = 0.71$), such as student-centered instruction, situating learning in realistic contexts, and encouraging collaboration between peers. To standardize the coefficient value, γ_{01} was multiplied by its own SD (0.56) and divided by the SD of individual student mathematics scores (16.41). The standardized coefficient of 0.06 for

γ_{01} indicates that a one SD increase in this variable predicts a 0.06 SD decrease in mathematics achievement.

After accounting for teacher use of *Standards*-based instructional practices in the model (Model 1), significant differences in mathematics achievement still existed between classrooms, $\chi^2 (787, n = 789) = 1617.33, p < .001$. As shown in Table 3-6, the residual variance was slightly smaller in Model 1 ($\tau_{00} = 49.56$) than in the baseline model ($\tau_{00} = 50.45$). Using Equation 1, teacher usage of *Standards*-based practices accounted for 2% of the classroom-level variance in achievement scores.

The results of Model 2 indicate that previous exposure to *Standards*-based instructional practices affected mathematics achievement in the third grade. Specifically, students' kindergarten exposure to collaborative learning environments ($\gamma_{03} = -1.17, p < .01$) and first-grade exposure to realistic and student-centered instruction ($\gamma_{04} = 1.59, p < .01$) were significantly associated with the dependent variable as was the third grade instructional scale ($\gamma_{01} = 1.71, p < .01$). After standardizing these coefficients, however, all variables proved to be weak predictors of mathematics achievement as a one SD change in any IV failed to predict a change of even a one-tenth SD in the DV. Computing Equation 1 from the variance components of Model 2 ($\tau_{00} = 47.41$) and the baseline model ($\tau_{00} = 50.45$) indicated that 6% of the classroom-level variance in achievement was explained by students' previous and current exposure to *Standards*-based instruction.

The fixed and random effects of Model 3 demonstrated that the mean mathematics score of a class in Grade 1 functions as a strong predictor of the class's score in Grade 3 ($\gamma_{04} = 0.89, p < .01$). Specifically, a one SD increase in prior

achievement predicted a 0.48 SD increase in the DV. The fixed effects of Model 3 reveal that the addition of this predictor weakened the association of students' kindergarten exposure to collaborative learning environments ($\gamma_{02} = -0.79, p < .01$) and first-grade exposure to student-centered instruction ($\gamma_{03} = -.07, p = .58$) with the dependent variable. After accounting for prior achievement, the latter variable was no longer a significant predictor of achievement. Current exposure to *Standards*-based instruction remained significantly, associated with grade-level mathematics achievement ($\gamma_{01} = 1.83, p < .01$).

After accounting for exposure to *Standards*-based instructional practices and prior achievement, no significant differences remained among teachers in the mean mathematics achievement of their students, $\chi^2(784, n = 789) = 705.76, p > .5$. As shown in Table 3-6, the residual variance was much smaller in Model 3 ($\tau_{00} = 0.43$) than in the baseline model ($\tau_{00} = 50.45$). Using Equation 1, the variables in Model 3 accounted for 99% of the classroom-level variance in mathematics scores.

Based on the results of re-running Model 3 (not shown) excluding all instructional scales, previous achievement alone accounted for 98% of the classroom-level variance in achievement. This finding indicates negligible contributions from any other classroom-level variables to the prediction of mathematics achievement in kindergarten after controlling for previous achievement.

The fixed effects of Model 4 confirmed these results. With previous achievement in the model, the percentage of students receiving free or reduced lunch (standardized coefficient of -0.16) was the only IV where a one SD change predicted a change of more than one-tenth of a SD in the DV.

Fifth Grade

Table 3-7 (fixed effects) and 3-8 (random effects) present the results of the fifth-grade analyses. Estimation of the baseline model revealed significant differences among teachers in the mean mathematics achievement of their students, $\chi^2(110, n = 111) = 245.50, p < .01$. Computing the ICC indicated that variance among class-mean scores accounted for 25% of the total variance in mathematics achievement among students. The grand mean of mathematics scores for fifth-grade students was 122.19 (SE = 1.10) with 95% of the classroom means predicted to fall between 107.25 and 137.13 points on the achievement scale.

The fixed effects of Model 1 indicated that fifth-grade mathematics achievement was not significantly associated with teacher use of *Standards*-based instructional practices ($\gamma_{01} = 3.32, p = .13$). Similarly, the fixed effects of Model 2 revealed that the mathematics achievement of fifth grade students was also unaffected by previous exposure to *Standards*-based instruction. Because no variable of interest functioned as a predictor, none of the subsequent models featuring the inclusion of control variables were estimated.

Table 3-1. Kindergarten hierarchical linear modeling fixed effects

Model	Fixed Effect	Coefficient	SE	<i>t</i> -ratio	<i>p</i>
Baseline	Intercept	27.09	0.20	137.36	0.00
Model 1	Intercept	27.11	0.20	138.27	0.00
	Problem Solving	0.41	0.19	2.18	0.03
	Collaborative Learning	-0.05	0.19	-0.25	>0.50
Model 2	Intercept	27.26	0.12	219.68	0.00
	Problem Solving	0.22	0.10	2.21	0.03
	Fall classroom math mean	1.09	0.03	34.11	0.00
Model 3	Intercept	27.78	0.20	135.56	0.00
	Problem Solving	0.28	0.10	2.81	0.01
	Fall classroom math mean	1.02	0.04	27.28	0.00
	% of free/reduced lunch students	0.15	0.30	0.50	>0.50
	% of minority students	- 1.49	0.29	-5.18	0.00

Table 3-2. Kindergarten hierarchical linear modeling random effects

Model	Random Effect	Variance	df	χ^2	p
Baseline	Class (μ_{0j})	14.74	901.00	2015.92	0.00
	Student (ϵ_{ij})	57.81			
Model 1	Class (μ_{0j})	14.53	899.00	1990.57	0.00
	Student (ϵ_{ij})	57.82			
Model 2	Class (μ_{0j})	0.13	899.00	649.82	>0.50
	Student (ϵ_{ij})	54.86			
Model 3	Class (μ_{0j})	0.09	897.00	617.51	>0.50
	Student (ϵ_{ij})	54.51			

Table 3-3. First grade hierarchical linear modeling fixed effects

Model	Fixed Effect	Coefficient	SE	t-ratio	p
Baseline	Intercept	43.90	0.16	271.25	0.00
Model 1	Intercept	43.90	0.16	278.89	0.00
	Problem Solving (first grade)	1.24	0.19	6.55	0.00
Model 2	Collaborative Learning (first grade)	-0.70	0.16	-4.22	0.00
	Intercept	43.90	0.16	279.67	0.00
	Problem Solving (first grade)	1.19	0.19	6.17	0.00
	Collaborative Learning (first grade)	-0.66	0.17	-3.93	0.00
	Problem Solving (kindergarten)	0.36	0.17	2.09	0.04
Model 3	Collaborative Learning (kindergarten)	-0.40	0.18	-2.26	0.02
	Intercept	43.84	0.10	452.61	0.00
	Problem Solving (first grade)	0.50	0.11	4.52	0.00
	Collaborative Learning (first grade)	-0.29	0.10	-2.94	0.00
	Problem Solving (kindergarten)	-0.15	0.10	-1.43	0.15
	Collaborative Learning (kindergarten)	-0.03	0.11	-0.27	>0.50
Model 4	Spring kindergarten classroom math mean	0.83	0.02	39.75	0.00
	Intercept	44.32	0.15	303.21	0.00
	Problem Solving (first grade)	0.51	0.11	4.70	0.00
	Collaborative Learning (first grade)	-0.28	0.10	-2.90	0.00
	Spring kindergarten classroom math mean	0.78	0.02	33.80	0.00
	% of free/reduced lunch students	-0.39	0.23	-1.67	0.10
	% of minority students	-0.85	0.24	-3.51	0.00

Table 3-4. First grade hierarchical linear modeling random effects

Model	Random Effect	Variance	df	χ^2	p
Baseline	Class (μ_{0j})	16.14	1264	3020.26	0.00
	Student (r_{ij})	54.90			
Model 1	Class (μ_{0j})	15.22	1262	2924.30	0.00
	Student (r_{ij})	54.86			
Model 2	Class (μ_{0j})	15.22	1262	2924.30	0.00
	Student (r_{ij})	54.86			
Model 3	Class (μ_{0j})	0.08	1259	1150.41	>0.50
	Student (r_{ij})	53.26			
Model 4	Class (μ_{0j})	0.07	1259	1132.78	>0.50
	Student (r_{ij})	53.08			

Table 3-5. Third grade hierarchical linear modeling fixed effects

Model	Fixed Effect	Coefficient	SE	t-ratio	<i>p</i>
Baseline	Intercept	87.96	0.39	226.33	0.00
Model 1	Intercept	87.97	0.39	227.58	0.00
	Standards-based Instructional Practices	-1.79	0.71	-2.53	0.01
Model 2	Intercept	87.97	0.38	229.73	0.00
	Standards-based Instructional Practices	-1.71	0.69	-2.47	0.01
	Problem Solving (kindergarten)	0.43	0.42	1.03	0.30
	Collaborative Learning (kindergarten)	-1.17	0.43	-2.70	0.01
	Problem Solving (first grade)	1.59	0.61	2.62	0.01
	Collaborative Learning (first grade)	-0.60	0.53	-1.13	0.26
Model 3	Intercept	88.06	0.24	353.75	0.00
	Standards-based Instructional Practices	-1.83	0.47	-3.88	0.00
	Collaborative Learning (kindergarten)	-0.76	0.26	-2.96	0.00
	Problem Solving (first grade)	0.19	0.34	0.56	>0.50
	First grade classroom math mean	0.89	0.03	27.99	0.00
Model 4	Intercept	90.63	0.45	199.74	0.00
	Standards-based Instructional Practices	-1.58	0.46	-3.45	0.00
	Collaborative Learning (kindergarten)	-0.53	0.25	-2.06	0.04
	First grade classroom math mean	0.80	0.03	24.24	0.00
	% of free/reduced lunch students	-0.08	0.01	-5.42	0.00
	% of minority students	0.01	0.01	0.45	>0.50

Table 3-6. Third grade hierarchical linear modeling random effects

Model	Random Effect	Variance	df	χ^2	p
Baseline	Class (μ_{0j})	50.45	788	1633.84	0.00
	Student (ϵ_{ij})	216.20			
Model 1	Class (μ_{0j})	49.56	787	1617.33	0.00
	Student (ϵ_{ij})	216.23			
Model 2	Class (μ_{0j})	47.41	783	1574.15	0.00
	Student (ϵ_{ij})	216.41			
Model 3	Class (μ_{0j})	0.43	784	705.76	>0.50
	Student (ϵ_{ij})	209.42			
Model 4	Class (μ_{0j})	0.48	783	665.33	>0.50
	Student (ϵ_{ij})	205.83			

Table 3-7. Fifth grade hierarchical linear modeling fixed effects

Model	Fixed Effect	Coefficient	SE	t-ratio	p
Baseline	Intercept	122.19	0.39	110.94	0.00
Model 1	Intercept	122.11	1.08	112.76	0.00
	Standards-based Instructional Practices	-3.32	2.18	-1.52	0.13

Table 3-8. Fifth grade hierarchical linear modeling random effects

Model	Random Effect	Variance	df	χ^2	p
Baseline	Class (μ_{0j})	58.08	110	245.50	0.00
	Student (ϵ_{ij})	173.07			
Model 1	Class (μ_{0j})	56.18	109	239.36	0.00
	Student (ϵ_{ij})	172.99			

CHAPTER 4 DISCUSSION

The present study explored whether student exposure to Standards-based instructional practices is related to mathematics achievement. The results indicate that a weak relationship exists between student achievement and teacher usage of Standards-based instructional practices at kindergarten, first grade, and third grade. While statistically significant, the practical significance of these findings is questionable: after controlling for previous achievement, a one standard deviation change in an instruction scale failed to result in more than a one-tenth standard deviation change in mathematics achievement at any grade. Furthermore, the association between the instructional scale (SBIP) and achievement at the third-grade level is negative. As discussed below, however, when interpreted in the context of the present study's limitations, these findings offer limited support for the use of reform mathematics. While inconsistent with the literature on Standards-based instructional practices and mathematics achievement, the results of the present study are consistent with the findings of previous large-scale, regression-based studies of instructional practices (Guarino et al., 2006; Milesi & Gamoran, 2006). These results are consistent with the larger conclusions regarding teacher effects on student achievement drawn by Scheerens and Bosker (1997) in their review of HLM-based studies of the educational system. These authors reported that student-level factors account for 60-70% of the variance in academic achievement, while classroom-level (i.e., class size, class demographic characteristics, and teacher-related variables) and school-level factors each accounted for only 15-20% of this variance. On the basis of such findings, the weak relationships between instructional practices and student achievement identified in

the present study and in other studies using a similar methodology and data source (i.e., Guarino et al., 2006; Milesi & Gamoran, 2006) are unsurprising.

The repeated findings of a weak association between mathematics achievement and teacher practices from regression-based studies do not disconfirm the strong positive relationships reported in the merged qualitative/experimental literature reviewed above (e.g., Rittle-Johnson, 2006; Steencken & Maher, 2003). Despite significant methodological differences between them, the qualitative and experimental research methodologies both examined Standards-based instructional practices by deliberately manipulating student exposure to this form of teaching. The present study, however, was limited to an examination of naturally-occurring variation in self-reported use of these practices among teachers. Furthermore, the infrequent usage of these techniques in U.S. classrooms (Hiebert et al., 2005; Porter, 1989) suggests even teachers in the present samples with high scores on the pertinent instructional scales (see below) engaged in substantially lower levels of Standards-based instructional practices than their peers in the qualitative and experimental studies. Thus, as students in the present study were likely not exposed to the classroom environments described by Lampert (1990), Cobb et al. (1991), and their colleagues, the results of the present study do not detract from the positive findings reported by these authors.

Instead, the findings of the present study generally support the arguments espoused by proponents of reform: namely, inquiry-based instruction is the foundation of Standards-based instructional practices and exposure to it enhances student learning. The specific supports for this argument come from the contrasting relationships of the PS and CL scales to mathematics achievement at the kindergarten

and first-grade levels. The PS scales represent teacher efforts to facilitate process-focused discourse on mathematical problem solving. As discussed in Chapter 1, one of the central principles of reform mathematics is that mathematics learning is enhanced by exposure to this form of mathematical discourse. Although the above teaching experiments (e.g., Cobb et al., 1991) used multiple instructional methods (e.g., collaborative learning, use of manipulatives), these techniques were specifically used to facilitate the process-focused discourse on mathematical problem-solving characteristic of inquiry-based instruction. According to the principles of reform mathematics (NCTM, 2000), these “faciliatory” teaching practices are not considered to be capable of positively impacting mathematics achievement without fostering such discourse (Yackel et al., 1991).

The PS scales contain questions pertaining to teacher facilitation of process-focused discourse within the classroom; therefore, the positive—though weak—relationship between these scales and mathematics achievement is expected based on the principles of reform mathematics. The CL scales do not contain such questions; they measure only the teacher’s frequency of assigning group work. Therefore, the nonsignificant (kindergarten) and negative (first grade) relationships between these scales and mathematics achievement are unsurprising for two reasons. First, without explicit incentives, group-work assignments fail to produce increases in either process-focused discourse amongst students (Yackel et al., 1991) or achievement gains relative to traditional instructional practices (Ginsburg-Block & Fantuzzo, 1998; Hurley et al., 2005; Janicki & Peterson, 1981; Madden & Slavin, 1983). Second, engagement in group-work without such a dialogue can negatively impact student learning as it may

promote the copying of answers by lower-performing members of the group (Yackel et al., 1991).

In summary, the relationships of the PS and CL scales to mathematics achievement at the kindergarten and first grade levels can be argued as support for the use of reform-based mathematics in early elementary education. Although the relationships are small, when placed in the context of the limited range of naturally-occurring usage of Standards-based instructional practices among U.S. teachers, these results suggest that higher levels of exposure to these practices benefits student learning.

Third-and Fifth-Grade Findings

Given the study's findings at the kindergarten and first-grade levels, the negative association between mathematics and the Standards-based Instructional Practices (SBIP) instructional scale at third grade can be interpreted in two ways. First, the negative association between the instructional scale and mathematics achievement may indicate that exposure to Standards-based instruction practices inhibits the learning of mathematics in third grade. More likely, however, the negative association is the result of a methodological flaw in the study: measuring teacher use of Standards-based instructional practices using questions not explicitly designed for this purpose. This flaw is more pronounced at the third and fifth grades because the items selected for the factor analysis at these grade levels less clearly pertained to teacher engagement in inquiry-based instruction than at the kindergarten and first-grade levels.

That is, the same items pertaining to Standards-based instructional practices (i.e., process-focused dialogue, collaborative learning, and situating mathematics in realistic contexts) were used in both the kindergarten and first grade teacher

questionnaires with a single exception. Many of these items, however, were discarded or modified in the third-and fifth-grade teacher questionnaires. The association of the new items to the principles of inquiry-based instruction was less clear than at the kindergarten and first grade-levels. For example, items questioning the extent to which teachers encourage their students to “talk to the class about their mathematics work” (NCES, 2002b, p. 53) and “discuss solutions to mathematics problems with other children” (NCES, 2002b, p. 53) are open to multiple interpretations, unlike items assessing the frequency with which students are asked to “explain how a mathematics problem is solved” (NCES, 2000, p. 51) or “work on problems for which there are several appropriate solutions or methods” (NCES, 2000, p. 51).

The lack of items pertaining to process-oriented teaching techniques in the SBIP instructional scale potentially explain both the failure to observe a two-factor solution and the resulting scale’s nonpositive relationship with mathematics achievement. As discussed in the previous section, the items used in the teacher questionnaire to assess use of collaborative learning do not represent the type of collaborative learning recommended in the Standards as they do not assess whether the teaching technique was used to promote mathematical dialogue among students (NCTM, 2000). Thus, without clear indicators of the extent to which process-focused discourse was promoted, the SBIP scale cannot be interpreted as a measure of Standards-based instructional practices. Therefore, its negative affect on achievement may not represent the true relationship between Standards-based instructional practices and achievement within the third grade. Attempting to find a suitable interpretation for this factor is beyond the scope of the present study.

As described above, at least one instructional scale demonstrated a significant but weak relationship with student achievement at all grades except the fifth. This exception likely results from the smaller sample size at the fifth grade level (i.e., 410 fifth-grade students compared to 3,553, 5,999, and 4,841 students in the third grade, first grade, and kindergarten, respectively). Due to the smaller sample size, the analysis lacked sufficient power to detect weaker relationships among variables. As such, the results of the fifth-grade analysis should be interpreted with caution.

Delayed Effects of Standards-based Instructional Practices

After controlling for previous achievement and classroom demographics, the only instructional scale to impact future achievement was the kindergarten CL scale. This scale was weakly but significantly associated with student achievement in third grade. The failure of any other instructional scale to affect future achievement and the difficulty with interpreting the finding (e.g., why is achievement affected at third grade rather than at first or fifth grade); however, suggest this finding may represent a Type I error. Therefore, this finding should be interpreted with caution.

Limitations

The use of instructional scales composed of items not explicitly designed to assess Standards-based instructional practices constitutes the main limitation of the study. This limitation calls into question the extent to which the instructional scales measure the application of Standards-based instructional practices in the classroom. For example, the present study considers the PS scales to represent teacher application of the principles of reform mathematics at the kindergarten and first grade levels; however, the information provided by the two scales suffers in comparison to the rich qualitative descriptions of Standards-based instruction provided by the teaching

experiments described in the literature reviewed for the present study (e.g., Steencken & Maher, 2003). Additionally, as the PS scales were composed of items not specifically designed to assess teacher use of Standards-based instructional practices, they likely overestimate teacher application of these instructional techniques within the classroom. That is, based on the non-specific nature of the items, the low reported incidence of American teachers engaging in these types of practices (Inagaki et al., 1999), and the extensive external supports provided in the teaching experiments (e.g., Cobb et al., 1991; Lampert, 1990), it seems unlikely that the classrooms of even teachers with the highest scores on the PS scale resemble those described by Lampert (1990), Civil (2002), and their colleagues. Thus, the present study likely does not measure the influence of “true” Standards-based instructional practices on student achievement.

In addition to the items used to assess Standards-based practices, the use of a one-time survey to measure their use in the classroom also constitutes a threat to the validity of the present study. Research has demonstrated that teacher logs and/or classroom observations are more accurate (although more costly) measures of instructional practices than surveys (Burstein et al., 1995; Rowan, Correnti, & Miller, 2002). Regarding the veracity of teachers' self-reported use of specific instructional practices, however, research suggests that information obtained from such surveys is generally valid, although less fine-grained than that obtained through other measures (Burstein et al., 1995; Rowan et al., 2002).

Despite the previously described limitations, strengths of using the ECLS-K dataset to assess the impact of Standards-based instructional practices include its large sample size, longitudinal design, and use multiple informants. Furthermore, the nested

structure of the data—students nested within teachers/classrooms—allowed for analyses that provide new information in this area.

Future Studies

Future studies on this subject should seek to address the limitations of the present study by using more accurate measures of Standards-based instructional practices. Future research should also examine one of the key assumptions of this study: that the weak relationship between student achievement and Standards-based instructional practices is the result of the limited range of the latter variable. This assumption could be tested through experimentally manipulating the level of student exposure to these practices. In such a study, both the quality and the quantity of Standards-based instruction would be manipulated. The effect of the quantity of this form of instruction on student achievement could be assessed by manipulating the number of days per week that students received access to Standards-based instructional practices (e.g., zero, two, or five days per week). The quality of Standards-based instructional practices could be manipulated by varying the level of external supports for Standards-based instructional practices between conditions. For example, one group would be provided with the type of external supports described in Cobb et al. (1991) and a second group of teachers would be provided with several in-services on this instructional methodology, but not with frequent, individualized consultation during the school year. Finally, the control group of teachers would be provided with no instruction in or encouragement of the use of Standards-based instructional practices.

Another area of future research lies in developing a quantitative measure of process-focused mathematical dialogue between peers to determine the relationship of this variable to both Standards-based instruction and student achievement. Specifically,

future research should seek to determine whether such discourse mediates the relationship between instructional practices and student learning.

Conclusions

Although not conclusive support for reform mathematics, the results of this study, particularly at the kindergarten and first-grade levels, suggest that Standards-based instructional practices positively influence mathematics achievement. As all previous studies of Standards-based instruction involved environmental manipulations to artificially increase student exposure to these practices, the positive findings within the limited naturally-occurring range of this variable represent an important contribution to arguments for reform.

Additionally, the present study provides insight into the role of inquiry-based instruction in reform mathematics. Specifically, the respective negative and positive relationships of the CL and PS scales to student achievement suggested that collaborative learning is ineffective without a concomitant emphasis on the use of process-focused mathematical dialogue amongst peers. With the extensive support found in both the quantitative and qualitative literature on Standards-based instructional practices (e.g., Rittle-Johnson, 2006; Steencken & Maher, 2003), this conclusion suggests that elementary school mathematics is better learned through discussion than drill. Thus, the findings of the present study support the use of reform mathematics within the school system, while also highlighting the need for further research in this area.

APPENDIX A
 MATHEMATICS INSTRUCTIONAL ACTIVITY ITEMS FROM TEACHER
 QUESTIONNAIRES

31. How often do children in this class do each of the following **MATH** activities? **CIRCLE ONE NUMBER ON EACH LINE.**

	Never	Once a month or less	Two or three times a month	Once or twice a week	Three or four times a week	Daily
a. Count out loud	1	2	3	4	5	6
b. Work with geometric manipulatives.....	1	2	3	4	5	6
c. Work with counting manipulatives to learn basic operations.....	1	2	3	4	5	6
d. Play math-related games	1	2	3	4	5	6
e. Use a calculator for math.....	1	2	3	4	5	6
f. Use music to understand math concepts	1	2	3	4	5	6
g. Use creative movement or creative drama to understand math concepts	1	2	3	4	5	6
h. Work with rulers, measuring cups, spoons, or other measuring instruments.....	1	2	3	4	5	6
i. Explain how a math problem is solved.....	1	2	3	4	5	6
j. Engage in calendar-related activities.....	1	2	3	4	5	6
k. Do math worksheets	1	2	3	4	5	6
l. Do math problems from their textbooks	1	2	3	4	5	6
m. Complete math problems on the chalkboard	1	2	3	4	5	6
n. Solve math problems in small groups or with a partner.....	1	2	3	4	5	6
o. Work on math problems that reflect real-life situations	1	2	3	4	5	6
p. Work in mixed achievement groups on math activities	1	2	3	4	5	6
q. Peer tutoring	1	2	3	4	5	6

Figure A-1. Mathematics instructional activities items at kindergarten. (Source: <http://nces.ed.gov/ecls/pdf/kindergarten/springteachersABC.pdf>. Last accessed May, 2010).

51. How often do children in this class do each of the following **MATH** activities? CIRCLE ONE NUMBER ON EACH LINE.

	Never	Once a month or less	Two or three times a month	Once or twice a week	Three or four times a week	Daily
a. Count out loud.....	1	2	3	4	5	6
b. Work with geometric manipulatives	1	2	3	4	5	6
c. Work with counting manipulatives to learn basic operations	1	2	3	4	5	6
d. Play math-related games.....	1	2	3	4	5	6
e. Use a calculator for math.....	1	2	3	4	5	6
f. Use music to understand math concepts.....	1	2	3	4	5	6
g. Use creative movement or creative drama to understand math concepts....	1	2	3	4	5	6
h. Work with rulers, measuring cups, spoons, or other measuring instruments	1	2	3	4	5	6
i. Explain how a math problem is solved.....	1	2	3	4	5	6
j. Engage in calendar-related activities	1	2	3	4	5	6
k. Do math worksheets.....	1	2	3	4	5	6
l. Do math problems from their textbooks.....	1	2	3	4	5	6
m. Complete math problems on the chalkboard.....	1	2	3	4	5	6
n. Solve math problems in small groups or with a partner	1	2	3	4	5	6
o. Work on math problems that reflect real-life situations	1	2	3	4	5	6
p. Work in mixed achievement groups on math activities	1	2	3	4	5	6
q. Peer tutoring.....	1	2	3	4	5	6
r. Work on problems for which there are several appropriate methods or solutions	1	2	3	4	5	6
s. Do worksheet or workbook page emphasizing routine practice or drill.....	1	2	3	4	5	6

Figure A-2. Mathematics instructional activities items at first grade. (Source: <http://nces.ed.gov/ecls/pdf/firstgrade/teachersABC.pdf>. Last accessed May, 2010).

53. How often do children in your class engage in the following? CIRCLE ONE NUMBER ON EACH LINE.

	Almost every day	Once or twice a week	Once or twice a month	Never or hardly ever
a. Solve mathematics problems from their textbooks	1	2	3	4
b. Solve mathematics problems on worksheets	1	2	3	4
c. Solve mathematics problems in small groups or with a partner	1	2	3	4
d. Work with measuring instruments, e.g., rulers	1	2	3	4
e. Work with manipulatives, e.g., geometric shapes	1	2	3	4
f. Use a calculator	1	2	3	4
g. Take mathematics tests.....	1	2	3	4
h. Write a few sentences about how to solve a mathematics problem.....	1	2	3	4
i. Talk to the class about their mathematics work.....	1	2	3	4
j. Write reports or do mathematics projects.....	1	2	3	4
k. Discuss solutions to mathematics problems with other children.....	1	2	3	4
l. Work and discuss mathematics problems that reflect real-life situations.....	1	2	3	4
m. Use a computer for math.....	1	2	3	4

Figure A-3. Mathematics instructional activities items at third grade. (Source: <http://nces.ed.gov/ecls/pdf/thirdgrade/teachersABC.pdf>. Last accessed May, 2010).

15. How often does the child identified on the cover of this questionnaire engage in the following as part of mathematics instruction? CIRCLE ONE NUMBER ON EACH LINE.

	Almost every day	Once or twice a week	Once or twice a month	Never or hardly ever
a. Solve mathematics problems from textbooks or worksheets.....	1	2	3	4
b. Solve mathematics problems from the blackboard or overhead.....	1	2	3	4
c. Solve mathematics problems in small groups or with a partner	1	2	3	4
d. Work with measuring instruments, e.g., rulers	1	2	3	4
e. Work with manipulatives, e.g., geometric shapes	1	2	3	4
f. Use a calculator.....	1	2	3	4
g. Take mathematics tests/quizzes.....	1	2	3	4
h. Write a few sentences about how to solve a mathematics problem.....	1	2	3	4
i. Discuss solutions to mathematics problems with other children.....	1	2	3	4
j. Work on and discuss mathematics problems that reflect real-life situations.....	1	2	3	4
k. Use a computer for math.....	1	2	3	4
l. Use visual representations (e.g., diagrams, tables, models)	1	2	3	4

Figure A-4. Mathematics instructional activities items at fifth grade. (Source: <http://nces.ed.gov/ecls/pdf/fifthgrade/teacherMath.pdf>. Last accessed May, 2010).

APPENDIX B
HIERARCHICAL LINEAR MODELING EQUATIONS (KINDERGARTEN)

Baseline Model

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean.

Model 1

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving})_j + \gamma_{02}(\text{Collaborative Learning})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving variable on the mean achievement of classroom j after controlling for Collaborative Learning.

γ_{02} : The effect of the Collaborative Learning variable on the mean achievement of classroom j after controlling for Problem Solving.

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for PS and Collaborative Learning.

Model 2

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving})_j + \gamma_{02}(\text{Fall kindergarten mathematics IRT score})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving variable on the mean achievement of classroom j after controlling for achievement scores from the fall of kindergarten.

γ_{02} : The effect of achievement scores from the fall of kindergarten on the mean achievement of classroom j after controlling for Problem Solving.

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Problem Solving and achievement scores from the fall of kindergarten.

Model 3

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving})_j + \gamma_{02}(\text{Fall kindergarten mathematics IRT score})_j + \gamma_{03}(\text{Percentage of students receiving free and reduced lunch})_j + \gamma_{04}(\text{Percentage of minority students})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving variable on the mean achievement of classroom j after controlling for CL, achievement scores from the fall of kindergarten, percentage of

minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

γ_{02} : The effect of achievement scores from the fall of kindergarten on the mean achievement of classroom j after controlling for Problem Solving, percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

γ_{03} : The effect of the percentage of student receiving free or reduced lunch in the classroom on the mean achievement of classroom j after controlling for Problem Solving, achievement scores from the fall of kindergarten, and percentage of minority students in the classroom.

γ_{04} : The effect of the percentage of minority students in the classroom on the mean achievement of classroom j after controlling for Problem Solving, achievement scores from the fall of kindergarten, and

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Problem Solving, achievement scores from the fall of kindergarten, percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

APPENDIX C
HIERARCHICAL LINEAR MODELING EQUATIONS (FIRST GRADE)

Baseline Model

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean.

Model 1

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving—first grade})_j + \gamma_{02}(\text{Collaborative Learning—first grade})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade).

γ_{02} : The effect of the Collaborative Learning (first grade) variable on the mean achievement of classroom j after controlling for Problem Solving (first grade).

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Problem Solving (first grade) and Collaborative Learning (first grade).

Model 2

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving—first grade})_j + \gamma_{02}(\text{Collaborative Learning—first grade})_j + \gamma_{03}(\text{Problem Solving--kindergarten})_j + \gamma_{04}(\text{Collaborative Learning—kindergarten})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (kindergarten), and Collaborative Learning (kindergarten).

γ_{02} : The effect of the Collaborative Learning (first grade) variable on the mean achievement of classroom j after controlling for Problem Solving (first grade), Problem Solving (kindergarten), and Collaborative Learning (kindergarten).

γ_{03} : The effect of the Problem Solving (kindergarten) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (first grade), and Collaborative Learning (kindergarten).

γ_{04} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (first grade), and Problem Solving (kindergarten).

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Collaborative Learning (first grade), Problem Solving (first grade), Collaborative Learning (kindergarten) and Problem Solving (kindergarten).

Model 3

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving—first grade})_j + \gamma_{02}(\text{Collaborative Learning—first grade})_j + \gamma_{03}(\text{Problem Solving--kindergarten})_j + \gamma_{04}(\text{Collaborative Learning--kindergarten})_j + \gamma_{05}(\text{Spring kindergarten mathematics IRT score})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (kindergarten), Collaborative Learning (kindergarten), and achievement scores from the spring of kindergarten.

γ_{02} : The effect of the Collaborative Learning (first grade) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (kindergarten), Collaborative Learning (kindergarten), and achievement scores from the spring of kindergarten.

γ_{03} : The effect of the Problem Solving (kindergarten) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (first grade), Collaborative Learning (kindergarten), and achievement scores from the spring of kindergarten.

γ_{04} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade),

Problem Solving (first grade), Problem Solving (kindergarten), and achievement scores from the spring of kindergarten.

γ_{05} : The effect of achievement scores from the spring of kindergarten on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), Problem Solving (first grade), Collaborative Learning (kindergarten), and Problem Solving (kindergarten).

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Collaborative Learning (first grade), Problem Solving (first grade), Collaborative Learning (kindergarten), and Problem Solving (kindergarten), and achievement scores from the spring of kindergarten.

Model 4

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Problem Solving—first grade})_j + \gamma_{02}(\text{Collaborative Learning—first grade})_j + \gamma_{03}(\text{Spring kindergarten mathematics IRT score})_j + \gamma_{04}(\text{Percentage of students receiving free and reduced lunch})_j + \gamma_{05}(\text{Percentage of minority students})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Collaborative Learning (first grade), achievement scores from the spring of kindergarten, percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

γ_{02} : The effect of the Collaborative Learning (first grade) variable on the mean achievement of classroom j after controlling for Problem Solving (first grade), achievement scores from the spring of kindergarten, percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

γ_{03} : The effect of achievement scores from the spring of kindergarten on the mean achievement of classroom j after controlling for Problem Solving (first grade), Collaborative Learning (first grade), percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

γ_{04} : The effect of the percentage of student receiving free or reduced lunch in the classroom on the mean achievement of classroom j after controlling for Problem Solving (first grade), Collaborative Learning (first grade), achievement scores from the spring of kindergarten, and the percentage of student minority students in the classroom

γ_{05} : The effect of the percentage of student receiving free or reduced lunch in the classroom on the mean achievement of classroom j after controlling for Problem Solving (first grade), achievement scores from the spring of kindergarten, and percentage of minority students in the classroom.

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Problem Solving (first grade), Collaborative Learning (first grade), achievement scores from the spring of kindergarten , percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

APPENDIX D
HIERARCHICAL LINEAR MODELING EQUATIONS (THIRD GRADE)

Baseline Model

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean.

Model 1

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Standards-based Instructional Practices})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Standards-based Instructional Practices variable on the mean achievement of classroom j .

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Standards-based Instructional Practices.

Model 2

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Standards-based Instructional Practices})_j + \gamma_{02}(\text{Problem Solving—first grade})_j + \gamma_{03}(\text{Collaborative Learning—first grade})_j + \gamma_{04}(\text{Problem Solving--kindergarten})_j + \gamma_{05}(\text{Collaborative Learning--kindergarten})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Standards-based Instructional Practices variable on the mean achievement of classroom j after controlling for Problem Solving (kindergarten), Collaborative Learning (kindergarten), Problem Solving (first grade), and Collaborative Learning (first grade).

γ_{02} : The effect of the Problem Solving (kindergarten) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten), Problem Solving (first grade), and Collaborative Learning (first grade).

γ_{03} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Problem Solving (kindergarten), Problem Solving (first grade), and Collaborative Learning (first grade).

γ_{04} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Problem Solving (kindergarten), Collaborative Learning (kindergarten), , and Collaborative Learning (first grade).

γ_{05} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Problem Solving (kindergarten), Collaborative Learning (first grade), and Problem Solving (first grade).

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Standards-based Instructional Practices, Problem

Solving (kindergarten), Collaborative Learning (kindergarten), Problem Solving (first grade), and Collaborative Learning (first grade).

Model 3

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Standards-based Instructional Practices})_j + \gamma_{02}(\text{Collaborative Learning—kindergarten})_j + \gamma_{03}(\text{Problem Solving—first grade})_j + \gamma_{04}(\text{Spring first grade mathematics IRT score})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Standards-based Instructional Practices variable on the mean achievement of classroom j after controlling for Collaborative Learning (kindergarten), Problem Solving (first grade) and first-grade achievement scores.

γ_{02} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Problem Solving (first grade), and first-grade achievement scores.

γ_{03} : The effect of the Problem Solving (first grade) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten), and first-grade achievement scores.

γ_{04} : The effect of first-grade achievement scores on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten) and Problem Solving (first grade).

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten) Problem Solving (first grade), and first-grade achievement scores.

Model 4

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Standards-based Instructional Practices})_j + \gamma_{02}(\text{Collaborative Learning—kindergarten})_j + \gamma_{03}(\text{Spring first grade mathematics IRT score})_j + \gamma_{04}(\text{Percentage of students receiving free or reduced lunch})_j + \gamma_{05}(\text{Percentage of minority students})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the Standards-based Instructional Practices variable on the mean achievement of classroom j after controlling for Collaborative Learning (kindergarten), first-grade achievement scores, percentage of student receiving free or reduced lunch in the classroom, and percentage of minority students in the classroom.

γ_{02} : The effect of the Collaborative Learning (kindergarten) variable on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, first-grade achievement scores, percentage of student receiving free or reduced lunch in the classroom, and percentage of minority students in the classroom

γ_{03} : The effect of first-grade achievement scores on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten), percentage of student receiving free or reduced lunch in the classroom, and percentage of minority students in the classroom.

γ_{04} : The effect of the percentage of student receiving free or reduced lunch in the classroom on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten), first-grade achievement scores, and percentage of minority students in the classroom.

γ_{05} : The effect of the percentage of minority students in the classroom on the mean achievement of classroom j after controlling for Standards-based Instructional Practices, Collaborative Learning (kindergarten), first-grade achievement scores, and the percentage of student receiving free or reduced lunch in the classroom

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for SBIP, Collaborative Learning (kindergarten), first-

grade achievement scores percentage of minority students in the classroom and percentage of student receiving free or reduced lunch in the classroom.

APPENDIX E
HIERARCHICAL LINEAR MODELING EQUATIONS (FIFTH GRADE)

Baseline Model

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean.

Model 1

Level 1:

$$\text{Math}_{ij} = \beta_{0j} + r_{ij}$$

Where:

Math_{ij} : Mathematics achievement of student i in classroom j .

β_{0j} : Mean mathematics achievement of classroom j .

r_{ij} : A random effect representing the deviation of student i in classroom j 's mathematics score from the mean of classroom j .

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Standards-based Instructional Learning})_j + \mu_{0j}$$

Where:

γ_{00} : Grand mean of mathematics achievement scores

γ_{01} : The effect of the SBIP variable on the mean achievement of classroom j .

μ_{0j} : A random effect representing the deviation of classroom j 's mathematics score from the grand mean after controlling for SBIP.

LIST OF REFERENCES

- Anand, P. G., & Ross, S. M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79*, 72-78. doi: 10.1037/0022-0663.79.1.72
- Beaver, K. M., Wright, J. P., & Maume, M. O. (2008). The effect of school classroom characteristics on low self-control: A multilevel analysis. *Journal of Criminal Justice, 36*, 174-181. doi: 10.1016/j.jcrimjus.2008.02.007
- Benbow, C. P., & Faulkner, L. R. (2008). Rejoinder to the critiques of the National Mathematics Advisory Panel final report. *Educational Researcher, 37*, 645-648. doi: 10.3102/0013189x08329195
- Boaler, J. (2008). When politics took the place of inquiry: A response to the National Mathematics Advisory Panel's review of instructional practices. *Educational Researcher, 37*, 588-594. doi: 10.3102/0013189x08327998
- Bosse, M. J. (1995). The NCTM standards in light of the new math movement: A warning! *Journal of Mathematical Behavior, 14*, 171-201. doi: 10.1016/0723-3123(95)90004-7
- Brown, A. L. (1994). The advancement of learning. *Educational Researcher, 23*, 4-12. doi: 10.3102/0013189X023008004
- Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., & Guitton, G. (1995). *Validating national curriculum indicators*. Santa Monica, CA: Rand Corporation.
- Civil, M. (2002). Chapter 4: Everyday mathematics, mathematicians' mathematics, and school mathematics: Can we bring them together? *Journal for Research in Mathematics Education. Monograph, 11*, 40-62. doi: 10.2307/749964
- Cobb, P. (1995). Continuing the conversation: A response to Smith. *Educational Researcher, 24*, 25-27. doi: 10.3102/0013189X024007025
- Cobb, P., Wood, T., & Yackel, E. (1993). Discourse, mathematical thinking, and classroom practice. In N. Minick, E. Forman & A. Stone (Eds.), *Education and mind: Institutional, social, and developmental processes* (pp. 91-119). New York, NY: Oxford University Press.
- Cobb, P., Wood, T., Yackel, E., Nicholls, J., Wheatley, G., Trigatti, B., & Perlwitz, M. (1991). Assessment of a problem-centered second-grade mathematics project. *Journal for Research in Mathematics Education, 22*, 3-29. doi: 10.2307/749551

- Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. *Economics of Education Review*, 26, 312-324. doi: 10.1016/j.econedurev.2005.05.008
- Ellis, M. W., & Berry, R. Q. I. (2005). The paradigm shift in mathematics education: Explanations and implications of reforming conceptions of teaching and learning. *The Mathematics Educator*, 15, 7-17.
- Fey, J. T. (1979). Mathematics teaching today: Perspectives from three national surveys. *Mathematics Teacher*, 72, 490-504.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7, 286-299. doi: 10.1037/1040-3590.7.3.286
- Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology*, 90, 560-569. doi: 10.1037/0022-0663.90.3.560
- Ginsburg, H. P., & Baroody, A. J. (1990). *The Test of Early Mathematics*. Austin, TX: PRO-ED.
- Guarino, C. M., Hamilton, L. S., Lockwood, J. R., & Rathbun, A. (2006). *Teacher qualifications, instructional practices, and reading and mathematics gains of kindergartners (NCES 2006-031)*. Washington, DC: National Center for Education Statistics.
- Hahs-Vaughn, D. L. (2005). A primer for using and understanding weights with national datasets. *Journal of Experimental Education*, 73, 221-248. doi: 10.3200/jexe.733.3.221-248
- Hamilton, L. S., & Guarino, C. M. (2005). *Measuring the practices, philosophies, and characteristics of kindergarten teachers (WR-199-EDU)*. Santa Monica, CA: RAND Corporation.
- Hausken, E. G., & Rathbun, A. (2004). *Mathematics instruction in kindergarten: Classroom practices and outcomes*. Paper presented at the American Educational Research Association, San Diego, CA.
- Hearst, E. (1999). After the puzzle boxes: Thorndike in the 20th century. *Journal of the Experimental Analysis of Behavior*, 72, 441-446. doi: 10.1901/jeab.1999.72-441
- Hergenhahn, B. R. (1992). *An introduction to the history of psychology* (2nd ed.). Belmont, CA: Wadsworth.

- Hiebert, J. (1999). Relationships between research and the NCTM standards. *Journal for Research in Mathematics Education*, 30, 3-19. doi: 10.2307/749627
- Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis*, 27, 111-132. doi: 10.3102/01623737027002111
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus. *Structural Equation Modeling*, 6, 1-55. doi: 10.1080/10705519909540118
- Hurley, E. A., Boykin, A. W., & Allen, B. A. (2005). Communal versus individual learning of a math-estimation task: African american children and the culture of learning contexts. *Journal of Psychology: Interdisciplinary and Applied*, 139, 513-527. doi: 10.3200/jrlp.139.6.513-528
- Huttenlocher, J., & Levine, S. C. (1990). *The Primary Test of Cognitive Skills*. New York, NY: CTB/McGraw Hill.
- Inagaki, K., Morita, E., & Hatano, G. (1999). Teaching-learning of evaluative criteria for mathematical arguments through classroom discourse: A cross-national study. *Mathematical Thinking and Learning*, 1, 93-111. doi: 10.1207/s15327833mtl0102_1
- Institute of Education Sciences. (2007). *Understanding sampling weights*. Paper presented at the Using the Early Childhood Longitudinal Study , Kindergarten Class of 1998-1999 (ECLS-K) Database for Research and Policy Discussion, Washington, DC.
- Janicki, T. C., & Peterson, P. L. (1981). Aptitude-treatment interaction effects of variations in direct instruction. *American Educational Research Journal*, 18, 63-82. doi: 10.3102/00028312018001063
- Jennings, J. L., & DiPrete, T. A. (2009). *Teacher effects on academic and social outcomes in elementary school*. Unpublished Manuscript. Department of Sociology, Columbia University. New York, NY.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27, 29-63. doi: 10.3102/00028312027001029
- Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172-192. doi: 10.2307/2112866
- Lipman, M. (1987). Ethical reasoning and the craft of moral practice. *Journal of Moral Education*, 16, 139-147. doi: 10.1080/0305724890160206

- Madden, N. A., & Slavin, R. E. (1983). Effects of cooperative learning on the social acceptance of mainstreamed academically handicapped students. *The Journal of Special Education, 17*, 171-182. doi: 10.1177/002246698301700208
- Markwardt, F. C., Jr. (1989). *Peabody Individual Achievement Test-Revised*. Circle Pines, MN: American Guidance Services.
- Milesi, C., & Gamoran, A. (2006). Effects of class size and instruction on kindergarten achievement. *Educational Evaluation and Policy Analysis, 28*, 287-313. doi: 10.3102/01623737028004287
- Muthen, B. O. (2005). Loading greater than one in an efa. Retrieved March 1, 2009, from <http://www.statmodel.com/cgi-bin/discus/discus.cgi?pg=prev&topic=8&page=181>
- Muthen, L. K., & Muthen, B. O. (2004). *MPLUS user's guide* (3rd ed.). Los Angeles, CA: Muthen & Muthen.
- Muthukrishna, N., & Borkowski, J. G. (1995). How learning contexts facilitate strategy transfer. *Applied Cognitive Psychology, 9*, 425-446. doi: 10.1002/acp.2350090506
- National Assessment Governing Board. (1996). *Mathematics frameworks of the 1996 National Assessment of Educational Progress*. Washington, DC: Government Printing Office.
- National Center for Education Statistics. (2000). Spring 2000 teacher questionnaire: Part b Retrieved from <http://nces.ed.gov/ecls/pdf/firstgrade/teachersABC.pdf>
- National Center for Education Statistics. (2001). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99: Base year public-use data files user's manual (NCES 2001–029)*. Washington, DC: Author.
- National Center for Education Statistics. (2002a). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99: First-grade public-use data files user's manual (NCES 002–134)*. Washington, DC: Author.
- National Center for Education Statistics. (2002b). Spring 2002 teacher questionnaire: Part a Retrieved from <http://nces.ed.gov/ecls/pdf/firstgrade/teacherABC.pdf>
- National Center for Education Statistics. (2004). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99: Third-grade public-use data files user's manual (NCES 2003–003)*. Washington, DC: Author.
- National Center for Education Statistics. (2006). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99: Fifth-grade public-use data files user's manual (NCES 2006-032)*. Washington, DC: Author.

- O'Connor, M. C. (1998). Chapter 2: Can we trace the "Efficacy of social constructivism"? *Review of Research in Education*, 23, 25-71. doi: 10.3102/0091732x023001025
- Piaget, J. (1954). *The construction of reality in the child*. New York, NY: Basic.
- Porter, A. (1989). A curriculum out of balance: The case of elementary school mathematics. *Educational Researcher*, 18, 9-15. doi: 10.3102/0013189x018005009
- Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *The Elementary School Journal*, 93, 213-228. doi: 10.1086/461723
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2006). HLM: Hierarchical linear and nonlinear modeling (Version 6.03). Chicago, IL: Scientific Software International.
- Resnick, L. B., & Hall, M. W. (1998). Learning organizations for sustainable education reform. *Daedalus*, 127, 89-118.
- Rimm-Kaufman, S. E., Fan, X., Chiu, Y.-J., & You, W. (2007). The contribution of the responsive classroom approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, 45, 401-421. doi: 10.1016/j.jsp.2006.10.003
- Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77, 1-15. doi: 10.1111/j.1467-8624.2006.00852.x
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey results tell us about teacher effects on student achievement: Insights from the prospects study of elementary schools. *Teachers College Record*, 104, 1525-1567.
- Salomon, G., & Perkins, D. N. (1998). Chapter 1: Individual and social aspects of learning. *Review of Research in Education*, 23, 1-24. doi: 10.3102/0091732x023001001
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness research*. Oxford, UK: Pergamon.
- Schoenfeld, A. (2007). Problem solving in the United States, 1970-2008: Research and theory, practice and politics. *ZDM*, 39, 537-551. doi: 10.1007/s11858-007-0038-z
- Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, 35, 13-21. doi: 10.3102/0013189x035002013

- Schumer, G. (1999). Mathematics education in Japan. *Journal of Curriculum Studies*, 31, 399-427. doi: 10.1080/002202799183061
- Senk, S. L., & Thompson, D. R. (2003). School mathematics curricula: Recommendations and issues. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do students learn* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum Associates.
- Slavin, R. E. (2008). Perspectives on evidence-based research in education--what works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5-14. doi: 10.3102/0013189x08314117
- Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, 78, 427-515. doi: 10.3102/0034654308317473
- Steencken, E. P., & Maher, C. A. (2003). Tracing fourth graders' learning of fractions: Early episodes from a year-long teaching experiment. *The Journal of Mathematical Behavior*, 22, 113-132. doi: 10.1016/S0732-3123(03)000017-8
- Stigler, J. W., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction. *Phi Delta Kappan*, 79, 14-21.
- The National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- The National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston VA: Author.
- Thorndike, E. L. (1913). *The psychology of learning*. New York, NY: Teacher's College.
- Thorndike, E. L. (1922). *The psychology of arithmetic*. New York, NY: Macmillan
- Verschaffel, L., & Corte, E. D. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education*, 28, 577-601. doi: 10.2307/749692
- Vukmir, L. (2001). 2+2=5: Fuzzy math invades Wisconsin schools. *Wisconsin Interest*, 10, 9-16.
- Wilkins, J. L. M. (2008). The relationship among elementary teachers' content knowledge, attitudes, beliefs, and practices. *Journal of Mathematics Teacher Education*, 11, 139-164. doi: 10.1007/s10857-007-9068-2
- Woodcock, R. W., & Bonner, M. (1989). *The Woodcock-Johnson Tests of Achievement-Revised*. Itasca, IL: Riverside Publishing.

- Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with LD. *Journal of Special Education, 36*, 89-101. doi: 10.1177/00224669020360020401
- Xue, Y., & Meisels, S. J. (2004). Early literacy instruction and learning in kindergarten: Evidence from the early childhood longitudinal study: Kindergarten class of 1998-1999. *American Educational Research Journal, 41*, 191-229. doi: 10.3102/00028312041001191
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education, 27*, 458-477. doi: 10.2307/749877
- Yackel, E., Cobb, P., & Wood, T. (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. *Journal for Research in Mathematics Education, 22*, 390-408. doi: 10.2307/749187
- Yackel, E., Cobb, P., & Wood, T. (1998). The interactive constitution of mathematical meaning in one second grade classroom: An illustrative example. *The Journal of Mathematical Behavior, 17*, 469-488. doi: 10.1016/S0732-3123(99)00003-6

BIOGRAPHICAL SKETCH

Jack Robbins Dempsey was born in Bryn Mawr, Pennsylvania in 1979. He grew up in Villanova, Pennsylvania, where he graduated from The Shipley School in 1998. He earned his B.A. in Psychology from Bowdoin College in 2002, after which he was awarded a two-year post-baccalaureate research fellowship at the National Institute on Drug Abuse (NIDA). While at NIDA, Jack met his future wife, Allison. The two were married in 2005.

In 2004, Jack began his doctoral degree in the school psychology program at the University of Florida. While at the University of Florida, Jack conducted several research projects and engaged in practicum work in school and clinic settings. Jack completed his pre-doctoral internship at the Munroe-Meyer Institute in Nebraska, before graduating with his Ph.D. in 2010.