UNDERSTANDING TEACHER AND CONTEXTUAL FACTORS THAT INFLUENCE THE ENACTMENT OF COGNITIVELY DEMANDING MATHEMATICS TASKS

By

Anne L. Garrison

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Learning, Teaching, and Diversity

August, 2013

Nashville, Tennessee

Approved:

Professor Paul A. Cobb

Professor Ilana S. Horn

Professor Marcy Singer-Gabella

Professor Thomas M. Smith

ACKNOWLEDGEMENTS

The research reported here was supported by Institute of Education Sciences (IES) pre-doctoral research training program, grant number R305B080025, and the National Science Foundation under grants No. ESI-0554535 and No. DRL-0830029. The opinions expressed are those of the author and do not necessarily reflect the views of the U.S. Department of Education or the National Science Foundation. I would like to thank members of the MIST research team for their help in collecting the data analyzed within this dissertation.

TABLE OF CONTENTS

Appendix

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

New reform goals and standards for students' mathematical learning have been put in place over the past two decades (e.g., see National Council of Teachers of Mathematics [NCTM], 1989, 2000; National Governors Association for Best Practices & Council of Chief State School Officers, 2010). These goals for students' mathematical learning imply new expectations for mathematics teachers' work in their classrooms. The *Curriculum and Evaluation Standards* and *Principles and Standards for School Mathematics* documents published by the National Council of Teachers of Mathematics (1989, 2000) and the more recent *Common Core State Standards* (National Governors Association for Best Practices & Council of Chief State School National Governors Association for Best Practices & Council of Chief State School Officers, 2010) reflect a consensus within the mathematics education research and policy communities for comprehensive reforms in how mathematics is taught. A fundamental aspect of high quality, inquiry-oriented mathematics instruction proposed in these documents is the use of challenging, or cognitively demanding, mathematical tasks. The level of challenge of the tasks that students solve and discuss impacts students' mathematical learning opportunities (Doyle, 1988; Hiebert & Grouws, 2007; Stein, Remillard, & Smith, 2007). In particular, there is evidence that challenging mathematical tasks support students in developing conceptual understanding (Stein & Lane, 1996).

The cognitive demand of a task refers to "the cognitive processes students are required to use in accomplishing it" (Doyle, 1988, p. 170). Stein, Grover, and Henningsen (1996) classified mathematics tasks into those with low and high cognitive demand. Mathematics tasks with low cognitive demand require students to memorize or reproduce facts, or perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks with high cognitive demand (or CDTs) require students to make connections to the underlying mathematical ideas. In addition, students are asked to engage in the disciplinary activities of explanation, justification, and generalization, or to use procedures to solve tasks that are open with regard to which procedures to use. I define the *enactment of CDTs* as involving two aspects: 1) *selecting* such tasks; and 2) *maintaining the cognitive demand* of those tasks during classroom implementation.

Teachers in the U.S. generally do not use cognitively demanding mathematics tasks in their classrooms, and when they do, they often implement them in ways that make them less challenging for students (Hiebert et al., 2003; Hiebert et al., 2005; Stein et al., 1996). Further, there is considerable evidence that it is challenging for teachers to develop the types of instructional practices described in the *Standards* documents, including enacting CDTs (e.g., Ball & Cohen, 1999; Elmore, Peterson, & McCarthey, 1996; Lambdin & Preston, 1995; C. L. Thompson & Zeuli, 1999), and even when they believe they are teaching in a manner consistent with the reforms, their classroom practices frequently diverge from the reforms (e.g., D. K. Cohen, 1990). Therefore, in order to determine how to support teachers to enact CDTs, the field needs to know more about which *teacher* and *contextual* factors might influence the enactment of CDTs. Further, we need to understand how teacher and contextual factors influence teachers' development of high quality instructional practice so that we can design effective and efficient ways to support teachers' development of such practice at scale.

The distinction between teacher and contextual factors requires clarification. Teacher factors include straightforward characteristics such as years of experience teaching as well as more latent and difficult-to-measure characteristics such as teacher knowledge and beliefs. A factor's inclusion in this category does not imply anything about whether the factor is static or changeable or whether it is exclusively a teacher factor or is influenced by the context. Contextual factors refer to aspects of the school settings in which teachers work that can potentially influence teachers' knowledge and practice. For example, when considering the practice of teaching, various student characteristics (e.g., students' current mathematical knowledge) are contextual factors.

Five recent studies have investigated teacher and contextual factors that influence the enactment of CDTs (Boston & Smith, 2009; Charalambous, 2010; Choppin, 2011; Son, 2008; Stein & Kaufman, 2010). Across these studies, there is evidence that a number of factors might influence the enactment of CDTs including: knowledge of students' thinking, conceptions of knowing and doing mathematics, teaching goals, alignment between teaching goals and textbooks, teacher perceptions about student achievement, test accountability, and teacher professional development. Further, there are mixed findings regarding the influence of mathematical knowledge for teaching and the type of curriculum (i.e., reform-oriented or traditional). These five studies contribute to our understanding of the enactment of CDTs, but there is still much to be learned about both critical factors that influence the enactment of CDTs and supporting teachers in enacting CDTs effectively. These three dissertation papers—one research synthesis and two empirical papers—attempt to build upon these five studies and other relevant literature to identify promising directions for future research and begin to address some of those questions.

In paper 1, I set an agenda for investigating the enactment of CDTs. After reviewing the five existing studies of different factors related to the enactment of CDTs, I cast a wide net to determine which other teacher and contextual factors might be worthy of investigation. In doing so, I identified 13 potentially relevant factors that were empirically and theoretically justified in their potential to be related to the process of enacting CDTs. This list sets an agenda for future studies of the enactment of CDTs. In addition to identifying these factors, another important aspect of the proposed research agenda concerns research methods. I argue that future large-scale studies of the enactment of CDTs should account for contingencies in expected relationships. For example, although we might expect that teachers' knowledge is related to their instructional practice, there is evidence that this relationship might be contingent upon their beliefs. Without accounting for that possible contingency, the relationship between knowledge and practice might not be statistically apparent across a large sample of teachers. In sum, the goal of paper 1 is to set the agenda for future research on the enactment of CDTs by suggesting both factors to investigate and an approach for carrying out those investigations in large-scale studies.

The empirical studies reported in papers 2 and 3 investigate some of those teacher and contextual factors, and the associated contingencies. Paper 2 investigates how mathematics teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics are related to the enactment of CDTs. In this analysis, I examine task selection and maintenance of the cognitive demand separately to investigate whether knowledge and beliefs are related to these two aspects of the enactment of CDTs in different ways. Also, I account for potential contingences in how mathematical knowledge for teaching and beliefs about teaching and learning mathematics are related to the enactment of CDTs to allow for possible interrelationships between teachers' knowledge and beliefs. An understanding of how teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics are related to their enactment of CDTs, will allow us to design better supports for teachers' development in enacting CDTs.

Current research on teacher learning and professional development suggests that ongoing interactions with relatively accomplished colleagues involving activities that are close to practice might support teachers' development. On the surface, work with a coach, collaborative teacher meetings, advice-seeking interactions, and professional development meet those criteria. Paper 3 investigates changes in teachers' enactment of CDTs over time and whether teachers' interactions with colleagues are related to change in their enactment of CDTs. In this paper, I investigate the influence of teachers' interactions in different settings (i.e., work with a coach, collaborative teacher meetings, advice-seeking interactions, and formal professional development) and the expertise available within those interactions. By examining interactions that have the potential to support teachers in enacting CDTs more effectively, I hope to contribute to the field's understanding of how to design better supports for teachers.

In sum, I seek to understand how to support and improve mathematics teachers' enactment of CDTs at scale. Together, these three interrelated studies set a direction for ongoing research in service of this goal, and begin to address some the key unresolved questions. In particular, the two empirical studies help to resolve some of the previously contradictory or weak findings within the recent studies of factors related to the enactment of CDTs and contribute ideas about how to support teachers in enacting CDTs.

CHAPTER II

PROMISING DIRECTIONS FOR FUTURE RESEARCH ON THE ENACTMENT OF COGNITIVELY DEMANDING MATHEMATICS TASKS

Introduction

The mathematics education research community has reached a general consensus on several key aspects of high quality mathematics instruction (National Council of Teachers of Mathematics [NCTM], 1989, 2000). One key aspect is the use of *cognitively demanding*, or challenging, mathematics tasks. The use of cognitively demanding tasks (CDTs) in the classroom has been linked to greater conceptual gains for students (Stein & Lane, 1996).

Developing high quality instructional practices is challenging for teachers (Stein et al., 1996). It involves changes in teacher knowledge and beliefs, as well as the development of new routines of practice (Ball & Cohen, 1999; Stein, Smith, & Silver, 1999; C. L. Thompson & Zeuli, 1999). Hence, teachers need considerable support to develop this type of practice. In order to determine what productive supports might entail, the field needs to know more about which factors might influence teachers' selection and implementation of CDTs. Knowledge of how various factors influence teachers' development of high quality instructional practice in general can inform the design of supports for teachers at scale. Well-designed large-scale studies can contribute to the development of this knowledge because they allow for generalization to larger populations. In particular, well-designed large-scale studies complement the analysis about processes produced by well-designed small-scale studies by examining phenomena in the aggregate and by providing information about general trends and critical patterns of variation. For example, a well-designed large-scale study might examine the effect of a policy on teachers' development, and identify meaningful variation in the effects (e.g., variation in policy implementation by school leadership or variation due to compatibility with other existing supports) with the aim of making policy recommendations to other school districts.

I define the *enactment of CDTs* as involving two aspects: 1) the *selection* of such tasks; and 2) the *maintenance of the cognitive demand* of those tasks during classroom implementation. Although it is very useful to understand what takes place in classrooms, it is also important to consider the influence of *teacher* and *contextual* factors on the enactment of CDTs. The distinction between these two types of factors requires clarification. Teacher factors might include both straightforward characteristics such as years of experience teaching and gender, and more latent and difficult-to-measure characteristics such as teacher knowledge and beliefs. A factor's inclusion in this category does not imply anything about whether the factor is static or changeable, or whether it is exclusively a teacher factor or is influenced by the context. Contextual factors refer to aspects of the school situations in which teachers work that can potentially influence teachers' knowledge and practice. For example, when considering the practice of teaching, various student characteristics (e.g., students' current mathematical knowledge) are contextual factors.

In this paper, I attempt to summarize and organize what is currently known in this area in order to understand the following: 1) What are potentially important teacher and contextual factors to study? 2) What is an effective approach for studying those factors on a large scale?

The first major section of this paper is a review of the existing literature on CDTs. I begin by reviewing studies that describe processes associated with the selection of CDTs and the maintenance of the cognitive demand of those tasks. Second, I define teacher and contextual factors and describe the implications for the study of the enactment of CDTs. Finally, I review the five studies that have explored how teacher and contextual factors are related to the enactment of CDTs. The results from these studies are mixed and inconclusive, and there is considerable variation in how they measure the enactment of CDTs. Hence, the findings of these five studies are insufficient to serve as a coherent basis for strategies or policies that would support teachers' enactment of CDTs. I therefore turn to the broader educational research literature to identify factors that might also influence the enactment of CDTs and are worthy of further investigation.

While few studies focused explicitly on teachers' enactment of cognitively demanding mathematical tasks, a significant number have investigated the relationships among teacher factors, contextual factors, and teachers' instructional practice more generally. In addition, many of those studies focused on aspects of teachers' instructional practice that are closely related to the enactment of CDTs. To identify all potentially relevant studies, I systematically searched databases and mined the reference lists of seminal literature reviews. I then limited the final set of studies to those that: (a) produced empirical results, (b) focused on aspects of mathematics teachers' instructional practice that are broadly related to the enactment of CDTs, (c) utilized classroom observation or student surveys to characterize instruction, and (d) went beyond descriptive characterizations of instructional practice to examine factors influencing practice. Applying these criteria resulted in a set of 63 studies that yielded 33 different teacher or contextual factors identified as potentially related to the enactment of CDTs. In the second section of the paper, I report the results of this review.

These factors, of course, vary in how they might be related to the enactment of CDTs. In the third section of this paper, I attempt to synthesize and reduce the set of potential factors to a more manageable and relevant subset: factors were excluded or adjusted if they could not be linked to specific processes associated with enactment of CDTs or if they were closely related to another key factor. I maintained the two broad categories of teacher factors and contextual factors. Six major teacher factors were identified: (a) mathematical knowledge for teaching, (b) knowledge of students' mathematical thinking, (c) beliefs about mathematics, (d) beliefs about teaching and learning mathematics, (e) beliefs about students' capabilities, and (f) classroom management skills. Seven major contextual factors were identified: (a) class time, (b) class size, (c) student characteristics (e.g., background knowledge, student expectations), (d) departmental culture, (e) school leaders' expectations for instructional improvement, (f) nature of the curriculum (e.g., whether it is inquiry-oriented, there are supports for teachers in teacher's guide), and (g) learning opportunities through interactions (e.g., work with a math coach or formal professional development). For each of factor, I describe the links to processes of enacting CDTs. I contend that we need to investigate how these 13 factors influence teachers' practice if we are to understand how we can successfully support teachers' enactment of CDTs in their classrooms.

While conducting the broad review, I tried to identify effective approaches for studying how factors are related to the enactment of CDTs on a large scale. In doing so it became clear that most of the large-scale studies assume that the factors are related to teacher practice in straightforward ways, while the small-scale studies are much more likely to describe contingencies in the way that particular factors influence practice. For example, large-scale studies have investigated relationships between teachers' knowledge and their practice while small-scale studies frequently describe how the expected relationship between teachers' knowledge and their practice might not hold under particular circumstances. I argue that it is both necessary and possible to consider contingencies in large-scale studies. We need to better understand the nuanced relationships in how specific factors influence the enactment of CDTs on a large scale by considering contingencies. In the fourth section, I illustrate a methodological approach for addressing contingencies by focusing on two categories of teacher factors: teacher beliefs and teacher knowledge. I describe some of the contingencies that should be considered in future large-scale studies of the relationship between teachers' knowledge, beliefs, and the enactment of CDTs.

Cognitive Demand of Mathematical Tasks

Over the last 25 years, mathematics educators and researchers have proposed new approaches for teaching mathematics that change the nature of activity in the classroom. The *Curriculum and Evaluation Standards* and *Principles and Standards for School Mathematics* documents published by the National Council of Teachers of Mathematics (1989, 2000) and the more recent *Common Core State Standards* (National Governors Association for Best Practices & Council of Chief State School National Governors Association for Best Practices & Council of Chief State School Officers, 2010) reflect a consensus within the mathematics education research and policy communities for comprehensive reforms in how mathematics is taught. Two fundamental aspects of high quality mathematics instruction outlined in these documents are the use of genuine, challenging tasks and students' participation in classroom discourse that focuses on key mathematical ideas that emerge from individual and collective efforts to solve such problems. While these two aspects center on students' activity in the classroom, they have clear implications for the role of the teacher (Hiebert et al., 1997). For example, the teacher is expected to choose and set up the challenging tasks for students and to orchestrate productive discourse within the classroom (Stein, Engle, Smith, & Hughes, 2008). Although classroom discourse both while solving the task and sharing solution methods provides critical learning opportunities for students, the level of challenge of the tasks selected is the foundation for those learning opportunities. For example, a task that requires students to reproduce memorized facts is unlikely to provide conceptual learning opportunities for students, no matter how well-orchestrated the classroom discourse. Hence, the cognitive demand, or level of challenge, of tasks is a critical aspect of high-quality mathematics instruction that requires further investigation. In the following paragraphs, I define cognitive demand, describe processes of enacting CDTs in the classroom, and review the five studies that have investigated how teacher and contextual factors are related to the enactment of CDTs.

The cognitive demand of a task refers to “the cognitive processes students are required to use in accomplishing it” (Doyle, 1988, p. 170). When examining the cognitive demand of mathematical tasks, Doyle (1988) chose *familiar* and *novel* as descriptors of two categories of mathematical tasks. Familiar tasks ask students to engage in routinized activities, whereas novel tasks are flexible with regard to how to carry out the task. Stein, Grover, and Henningsen (1996) built on Doyle’s work by more systematically delineating the cognitive demand of different types of mathematical tasks. They classified tasks into those with low and high cognitive demand (with parallels to familiar and novel tasks, respectively). Tasks with low cognitive demand require students to memorize or reproduce facts, or perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks with high cognitive demand (or CDTs) require students to make connections to the underlying mathematical ideas. In addition, students are asked to engage in one or more of the disciplinary activities of explanation, justification, and generalization, or to use procedures to solve tasks that are open with regard to which procedures to use. Although implied in the definition, it is important to emphasize that the distinctions between familiar and novel tasks, and between high and low cognitive demand tasks, are relative to students' current understanding and, thus, are situation-dependent.

There is evidence that CDTs can provide critical learning opportunities for all students. Stein and Lane (1996) found that the use of tasks with high cognitive demand was related to greater student gains on an assessment requiring high levels of mathematical thinking and reasoning. In particular, the greatest gains occurred when teachers assigned tasks that were initially of high cognitive demand, and teachers and students maintained the cognitive demand throughout the lesson. Further, there is evidence that high cognitive demand thinking affords valuable learning opportunities for all students, not just previously high-achieving students (Zohar & Dori, 2003). The enactment of high cognitive demand tasks in the classroom therefore appears to be important in supporting all students' learning.

Unfortunately, it is clear that CDTs are not often enacted in US classrooms. In attempting to understand more about changes in cognitive demand during a lesson, Stein, Grover, and Henningsen (1996) documented the initial cognitive demand of mathematical tasks as written or verbally posed to students, and examined whether teachers and students maintained, increased, or decreased the demand in different phases of a math lesson. They found that in classrooms where tasks with the potential for high levels of cognitive demand were used, teachers and/or students often decreased the cognitive demand during implementation of the tasks. The results from the 1999 Trends in International Math and Science Study (TIMSS) video study are consistent with those of Stein and colleagues in that they suggest that the mathematical activity in US middle school classrooms tends to be procedural in nature, and when teachers do select high-level tasks they often implement them in low-level ways (Hiebert et al., 2003; Hiebert et al., 2005).

The Math Task Framework proposed by Stein, Grover, and Henningsen (1996) is useful when analyzing how teachers enact tasks. The Framework divides a lesson up into phases and transitions between phases of the lesson (see Figure 1). The squares denote different phases of the lesson and flow from left to right. In the subsequent paragraphs, I describe how my definition of the enactment of CDTs maps onto the Math Task Framework. In this analysis, I focus on the cognitive demand of the written task as selected from the curricular materials (represented in the leftmost square) and then the changes in the cognitive demand from how it is written to the implementation by the teacher and students (denoted in the second and third squares). Both of these aspects of cognitive demand influence whether students engage in cognitively demanding mathematical activity in the classroom.

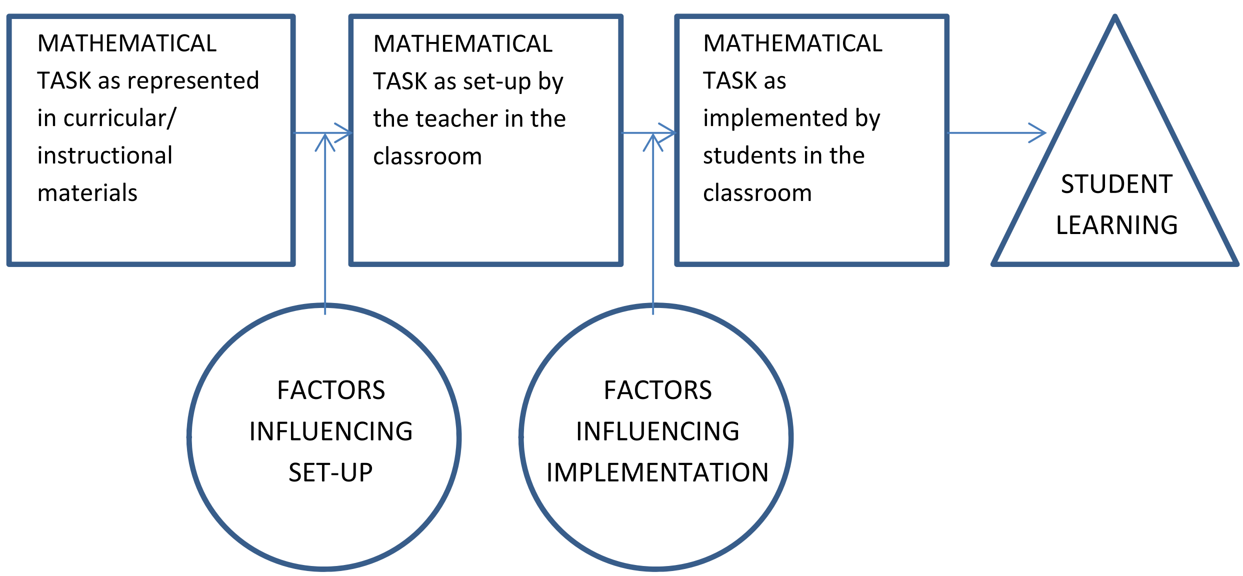


Figure 1. Modified "Math Tasks Framework" (Stein et al., 1996, p.459)

First, teachers select a task from the curriculum materials (in Figure 1, the leftmost square). In selecting a task to pose to students, teachers might choose to use a task directly from the curriculum as suggested by the district pacing guide, use other tasks from the curriculum, adjust tasks from the curriculum, use tasks from supplementary materials, or create their own tasks. In doing any one of these options, the teacher may have to balance a range of issues including: teaching goals, class time, knowledge of the students, what he or she feels capable of, and others' expectations. For example, a teacher who only has 45 minutes for each class period might decide that she does not have enough class time to engage students in a high cognitive demand task. It is unlikely that teachers consider each of these concerns individually, but instead they weigh them against each other when deciding which to prioritize and how they fit with their goals for the class period (Remillard, 1999). The cognitive demand of the selected task sets the stage for the cognitive demand over the course of the lesson.

Once a CDT has been selected, maintenance of the cognitive demand concerns both the task set-up and implementation (in Figure 1, the second and third squares). In the first of those phases, the task is set up for the students in the classroom (in Figure 1, the second square from the left). In other words, the teacher explains what students are expected to do to complete the task. During this phase a teacher can alter the cognitive demand of the task by clarifying or changing the expectations set out in the written task. For example, a teacher might tell students to complete only part of the written task or might go through a series of examples that change the nature of the task in which students will engage.

The next phase (in Figure 1, the third square from the left) is the implementation of the task by the students in the classroom. This phase includes all of the remaining class time spent on the task. For example, it might include both student work time and a concluding whole-class discussion. The cognitive demand of the task can also change within this phase of instruction, depending on teacher and student actions as they carry out the task. There are several reasons why cognitive demand might decrease: 1) the teachers' expectations for students' work might be unclear, 2) the classroom environment might not be conducive to engaging in challenging mathematical activity (e.g., poor classroom management or unproductive classroom norms), 3) the task might not be appropriate for students given their current mathematical knowledge, or 4) the appropriate level of scaffolding or teacher support for students to productively engage in high-level ways might not occur (Henningsen & Stein, 1997; Stein et al., 1996). Given the range of the reasons why cognitive demand might decrease, it is clear that maintaining the cognitive demand of CDTs can be quite difficult.

Teacher and Contextual Factors

To this point the description I have given of the enacting CDTs highlights classroom-specific reasons why the level of challenge in the classroom might be reduced. This, however, gives little attention to teacher factors and the contexts in which teachers work. Possible teacher factors include both straightforward characteristics like years of experience teaching and gender, and more latent and difficult to measure characteristics like teacher knowledge and beliefs. A factor's inclusion in this category does not imply anything about whether the factor is static or changeable, or whether it is exclusively a teacher factor or is influenced by the context. For example, mathematical knowledge for teaching is a teacher factor but is specific to the context of teaching mathematics and can develop through practice (Hill, Sleep, Lewis, & Ball, 2007; Sherin, 2002).

From the perspective of supporting teachers' instructional practice, the notion of contextual factors refers to aspects of the school and district settings in which teachers work that are relevant to teachers' practice. For example, when considering the practice of teaching, various student characteristics (e.g., the current mathematical knowledge of students) are contextual factors. In addition to teacher and classroom factors, school and district contextual factors are likely to influence teacher's practice (Cobb & Smith, 2008). For example, school factors might include the opportunities for teacher collaboration, the school culture, and the expectations of the principal. District factors might include the curriculum and professional development opportunities for teachers. When considering a particular teacher's practice, I define *curriculum* to be the adopted curriculum and/or supplementary materials that a teacher uses in the classroom. Contextual factors can influence a teacher's current practice and/or influence whether teachers change their practice. For example, the adopted curriculum might influence teacher's current practice, but it might also include supports for teachers in developing their practice. I include both types of influence when considering factors influencing practice. In sum, when accounting for teachers' instructional practices, it is also important to look beyond teacher factors to the contextual factors within the school and district to understand their practices (Cobb, McClain, Lamberg, & Dean, 2003; Coburn, 2005).

The school and district contexts in which teachers work are themselves influenced by the broader state and national policy environment. However, the impact of the state and national policy environment on teachers' practice is typically mediated by decisions made by district and school leaders (Coburn, 2001, 2005; Coburn & Russell, 2008; James P. Spillane et al., 2002). For example, the federal No Child Left Behind Act (NCLB) requires states to assess whether schools are making adequate yearly progress on standardized tests, and this can result in different responses by district or school leaders', which are then likely to influence teachers' practice. Some district leaders might respond to the NCLB requirements by pressing teachers to develop high quality instructional practices whereas others might press teachers to focus on test preparation. Therefore, the broader state and national policy environment is not likely to have a direct impact on teachers' practice, but instead affects teachers' practice indirectly through aspects of the school and district contexts. In the following literature reviews, I limit factors to those that are likely to directly influence teachers' practice.

Review of Literature of Factors influencing the Enactment of CDTs

In recent years, five studies have investigated particular teacher and contextual factors as they relate to teachers' enactment of CDTs. Son (2008) examined elementary teachers' mathematics textbook use with an emphasis on teachers' patterns of cognitive demand using survey data from 169 teachers, with classroom observational data and artifacts from eight of those teachers. In the observational analysis Son focused on three different patterns of cognitive demand between the problems used from the textbook and the types of questions asked of the students over the course of the lesson: high-level problems to high-level questions, high-level problems to low-level questions, and low-level problems to low-level questions. This approach is similar to examining the changes in cognitive demand from the task as selected to the task as implemented. Several key factors related to the cognitive demand patterns emerged including: teachers' conceptions of knowing and doing mathematics, teachers' goals for student learning, alignment between the teachers' goals and textbooks, the nature of the textbook, teachers' use of the textbook, teachers' perceptions about student achievement, time for mathematics instruction, and test accountability. While these factors were identified through observation and case study analysis, the author did not provide explanations of the process by which they were related to the cognitive demand patterns. Therefore, it is hard to discern which of these factors are most critical or most likely to influence teachers' enactment of CDTs in the future. An additional factor that emerged from the survey was the teachers' perceived mathematical knowledge for teaching. Teachers were asked to rate their content, pedagogical, and curriculum knowledge (e.g., Mathematics content knowledge on whole numbers) from very poor to excellent. However, teacher self-reports of this type are difficult to interpret with regard to what is actually being assessed. In fact, Son's study corroborates this difficulty: Son found that two of the six case study teachers who reported that they used high-level problems and then asked high-level questions, did use high-level problems but tended to ask low-level questions in their classrooms. In other words, the teachers' self-reports of their classroom practice were not consistent with their actual practice.

In another study that considered how teacher knowledge relates to maintenance of the cognitive demand of mathematical tasks, Charalambous (2010) examined the relationship between teachers' mathematical knowledge for teaching and their "task unfolding" (i.e., changes in the cognitive demand over the course of a lesson) for two teachers who differed dramatically in their mathematical knowledge for teaching. He found that mathematical knowledge for teaching was related to maintenance of the cognitive demand of CDTs and as a result of his case analyses, he proposed several hypotheses about how high levels of mathematical knowledge for teaching are integral to the work of teaching. For example, his first hypothesis was that "strong mathematical knowledge for teaching supports teachers in using representations to attach meaning to mathematical procedures rather than to simply show answers" (Charalambous, 2010, p. 273). His hypotheses suggest mechanisms by which mathematical knowledge for teaching is integral to teacher practice. These hypothesized mechanisms make a fairly convincing argument for how mathematical knowledge for teaching is integral to the enactment of CDTs. As a part of his analysis of the cases, Charalambous also suggested other factors that may influence the relationship between mathematical knowledge for teaching and task unfolding, but his focus was *how* mathematical knowledge for teaching is integral to maintenance of the cognitive demand, rather than examining multiple factors simultaneously.

Stein and Kaufman (2010) examined how several different factors are related to inquiry-oriented curriculum implementation. In particular, they focused on curricular materials and teacher capacity, conceptualized as teacher education, experience, professional development, and mathematical knowledge for teaching, as they related to inquiry-oriented curriculum implementation (with one of three foci being maintaining high levels of cognitive demand) for 48 elementary teachers in two large, urban school districts. To investigate the enactment of CDTs, they created a total cognitive demand score that was the sum of the cognitive demand of the task as selected and as implemented: if high-level tasks were selected and the cognitive demand was maintained, then the total cognitive demand score would be at its maximum value. They found that the way teachers used curriculum materials, hours of professional development, and perceptions of curriculum usefulness each were positively related to the total cognitive demand in the classroom for the 24 observed elementary teachers in one large, urban district. However, those relationships were not significant for the 24 teachers in the other large, urban school district. In addition, although they expected that teacher knowledge, education, and experience would be related to curriculum implementation, they did not find any significant relationships between those teacher factors and instructional quality. In fact, they found that mathematical knowledge for teaching was not significantly related to the cognitive demand of the tasks in the classroom, directly contradicting Charalambous' findings. This suggests that the relationship between mathematical knowledge for teaching and the cognitive demand of mathematical tasks might not be straightforward, and should be explored further.

Choppin (2011) investigated how the things that teachers noticed when they reviewed video-recordings of their own teaching were related to their enactment of CDTs. In particular, he studied five middle school teachers who had at least three years of experience in using the same inquiry-oriented curriculum and found that teachers who attended to student thinking when they viewed recordings, used their knowledge of students' thinking to enact CDTs, whereas teachers who only evaluated student thinking as right or wrong when reviewing recordings of their teaching knew less about their students' thinking and often decreased the cognitive demand of tasks. Further, he found that by attending to students' thinking, teachers developed in their general understanding of students' learning (i.e., learning trajectories) over time, which also influenced their enactment of CDTs. Results from this study suggest that teachers' knowledge of students' thinking, another dimension of teachers' knowledge, is integral to teachers' enactment of CDTs.

Lastly, Boston and Smith (2009) studied the effects of professional development (PD) focused specifically on the cognitive demand of mathematical tasks on teachers' enactment of CDTs. Their sample included 18 secondary mathematics teachers who participated in the PD and 10 contrast teachers who did not. The participating and contrast teacher samples were split with regard to whether their schools used inquiry-oriented or traditional curricula. Boston and Smith found that their PD program had a significant and positive influence on both teachers' choices of tasks posed to students and the implementation of those tasks in the classroom. In particular, after participation in professional development, teachers chose more CDTs and were more likely to implement them in high-level ways. It is important to note that for some teachers the cognitive demand of high-level tasks still decreased over the course of the lesson. Therefore, there is still more to understand about supporting teachers to consistently maintain the cognitive demand of high-level tasks. Another dimension that the researchers explored was the influence of the type of curriculum on those PD effects. They were surprised to find that there were not significant curriculum-related differences in terms of the cognitive demand of tasks or effects of PD. This contradicts the significant curriculum-related findings of Son (2008) and Stein and Kaufman (2010), and suggests that, like the relationship between MKT and the enactment of CDTs, the relationship between curriculum materials and the enactment of CDTs requires further exploration.

Looking across the findings of these five studies, there is some evidence that enacting CDTs may be related to both teacher and contextual factors. Only the Boston and Smith (2009) study and the Choppin (2011) study examined the cognitive demand of tasks posed to students and the maintenance of the cognitive demand of high-level tasks separately. All of the other studies focused on task enactment patterns (i.e., “task unfolding”) or they used a measure of cognitive demand that aggregated across the selection and implementation. Yet, there is some evidence that numerous factors might influence the enactment of CDTs, including: knowledge of students’ thinking, conceptions of knowing and doing mathematics, teaching goals, alignment between teaching goals and textbooks, teacher perceptions about student achievement, test accountability, and professional development. However evidence concerning the influence of teachers’ mathematical knowledge for teaching and the type of curriculum materials used is inconsistent. These inconsistencies may be related to either a lack of attention to the contingencies associated with different factors or the way mathematical knowledge for teaching and type of curriculum are defined and measured.

	Category	Factor	References	Included	Excluded	Closely Related
Teacher Factors	Teacher Knowledge	Mathematical Knowledge for Teaching	(Charalambous, 2010; Escudero & Sánchez, 2007; Hill, Ball, Blunk, Goffney, & Rowan, 2007; Hill, Blunk, et al., 2008)	X		
		Knowledge of Mathematics	(Ball, 1991; Baumert et al., 2010; D. K. Cohen, 1990; Lambdin & Preston, 1995; Manouchehri & Goodman, 2000; Sherin, 2002; Yun-peng, Chi-chung, & Ngai-ying, 2006)	X		
		Pedagogical Content Knowledge	(Baumert et al., 2010; Manouchehri & Goodman, 1998, 2000; Philipp, Flores, Sowder, & Schappelle, 1994; Sherin, 2002; Yun-peng et al., 2006)	X		
		Knowledge of Student Thinking	(Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Choppin, 2011; Peterson, Carpenter, & Fennema, 1989)	X		
		General Pedagogical Knowledge or Skills	(Manouchehri & Goodman, 2000; Opdenakker & Van Damme, 2006)	X		
	Teacher Beliefs	Beliefs about Mathematics	(Aguirre & Speer, 1999; Beswick, 2005; Cross, 2009; Lloyd & Wilson, 1998; Putnam, Heaton, Prawat, & Remillard, 1992; Raymond, 1997; Remillard, 1999; Remillard & Bryans, 2004; Stipek, Givvin, Salmon, & MacGyvers, 2001; Stodolsky & Grossman, 2000; A. G. Thompson, 1984)	X		
		Beliefs about Curriculum	(Lambdin & Preston, 1995; Lloyd, 1999; Manouchehri & Goodman, 2000; Remillard & Bryans, 2004; Sowder, Philipp, Armstrong, & Schappelle, 1998; Superfine, 2009)			X
		Beliefs about Teaching and Learning Mathematics	(Aguirre & Speer, 1999; D. K. Cohen, 1990; Cross, 2009; Fennema et al., 1996; Lambdin & Preston, 1995; Lloyd, 1999; Manouchehri & Goodman, 1998; Philipp et al., 1994; Putnam et al., 1992; Remillard, 1999; Remillard & Bryans, 2004; Schoenfeld, 2011; Skott, 2001; Stipek et al., 2001; Superfine, 2009; A. G. Thompson, 1984)	X		
		Beliefs about Teaching	(Barrett Paterson, 2009; Beswick, 2005; D. K. Cohen, 1990; Manouchehri & Goodman, 2000; A. G. Thompson, 1984; Wood, Cobb, & Yackel, 1991)	X		
		Beliefs about Student Learning	(Cooney, 1985; Jamar & Pitts, 2005; Manouchehri, 2004; Prawat & Jennings, 1997; Son, 2008; Sowder et al., 1998; Stipek et al., 2001; Stodolsky & Grossman, 2000; Sullivan & Leder, 1992; Sztajn, 2003; Turner, Warzon, & Christensen, 2011; Yun-peng et al., 2006)	X		
		Beliefs about Affective Issues	(Aguirre & Speer, 1999; Hill, Blunk, et al., 2008)			X
	Teacher Affect	Job Satisfaction	(Opdenakker & Van Damme, 2006)		X	
		Reflectiveness	(Philipp et al., 1994; Smith, 2000; A. G. Thompson, 1984)		X	
		Disposition toward Math	(Stipek et al., 2001)		X	

	Teacher Goals	Teachers' Goals	(Aguirre & Speer, 1999; Manouchehri & Goodman, 2000; Schoenfeld, 2011; Skott, 2001; Son, 2008; Sowder et al., 1998; Stodolsky & Grossman, 2000)			X
	Teacher Experience	Experience As Students	(Anderson, White, & Sullivan, 2005; Cross, 2009; Raymond, 1997)			X
		Experience Teaching Math	(Charalambous, 2010; Manouchehri & Goodman, 1998; Superfine, 2009)		X	
		Experience with Curriculum	(Remillard & Bryans, 2004)		X	
Contextual Factors	Aspects of Teachers' Work	Time Demands	(Barrett Paterson, 2009; Cooney, 1985; Rousseau & Powell, 2005; Son, 2008).	X		
		Class Size	(Rousseau & Powell, 2005)	X		
	Student-Related Factors	Background Knowledge of Students	(McGinnis, Parker, & Graeber, 2004; Opdenakker & Van Damme, 2006; Son, 2008)	X		
		Student Expectations for Instruction	(Cooney, 1985; Herbel-Eisenmann, Lubienski, & Id-Deen, 2006; McGinnis et al., 2004; Sullivan & Leder, 1992)	X		
		Mobility/Absenteeism	(Rousseau & Powell, 2005)	X		
		Students' Backgrounds	(Anderson et al., 2005; Goos, Dole, & Makar, 2007; Raymond, 1997; Yun-peng et al., 2006)	X		
	School Context	School leadership	(Manouchehri & Goodman, 1998; Philipp et al., 1994)	X		
		Culture of math department	(Anderson et al., 2005; Manouchehri & Goodman, 1998; McGinnis et al., 2004; Raymond, 1997; Stodolsky & Grossman, 2000)	X		
		Curriculum	(Barraugh, 2011; Diaz, 2004; Herbel-Eisenmann et al., 2006; Lambdin & Preston, 1995; Manouchehri & Goodman, 2000; McGinnis et al., 2004; Remillard, 1999; Remillard & Bryans, 2004; Rousseau & Powell, 2005; Schoenfeld, 2011; Son, 2008; Stein & Kaufman, 2010; Wang & Paine, 2003)	X		
		Parents' Expectations	(Herbel-Eisenmann et al., 2006; McGinnis et al., 2004)		X	
	Accountability	Accountability Pressures	(Barraugh, 2011; McGinnis et al., 2004; Rousseau & Powell, 2005; Son, 2008; Yun-peng et al., 2006)		X	
	Interactions	Access to Expertise	(Diaz, 2004; Manouchehri & Goodman, 1998; Neuberger, 2010)	X		
		Teacher Collaboration	(Barraugh, 2011; Bruce & Ross, 2008; Franke, Carpenter, Levi, & Fennema, 2001; Smith, 2000; Wang & Paine, 2003; Yun-peng et al., 2006)	X		
		Formal Professional Development	(Barton, 2005; Boston & Smith, 2009; Carpenter et al., 1989; Fennema et al., 1996; Franke, Carpenter, Fennema, Ansell, & Behrend, 1998; Neuberger, 2010; Owston, Sinclair, & Wideman, 2008; Remillard, 1999; Swafford, Jones, Thornton, Stump, & Miller, 1999; Turner et al., 2011; Walker, 2007; Yun-peng et al., 2006)	X		

Figure 2. Factors Related to Teachers' Instructional Practice, by Category

Review of Literature of Factors Influencing Teachers' Instructional Practice

Although few studies have investigated the enactment of CDTs, a significant number of studies have investigated the relationships between teacher factors, contextual factors, and teachers' instructional practice more generally. Importantly, many of these additional studies have focused on aspects of teachers' instructional practice that are closely related to the cognitive demand of mathematical tasks in the classroom. These aspects range from general characteristics of instructional quality (e.g., inquiry-oriented curriculum implementation) to specifics of the classroom activity (e.g., whether the activity in the classroom is procedurally or conceptually oriented). Given the widespread inclusion of aspects of instructional practice that are closely related to the cognitive demand of mathematical tasks in the classroom, I use the general term *teachers' instructional practice* to refer to them. In this section I report on a review of studies that sought to identify relationships between teacher factors, contextual factors, and teachers' instructional practice.

The first stage of the review process was to search the literature for relevant studies. To generate the list of studies to include in this review, I utilized several complementary search strategies. First, I drew on 1992 and 2007 handbook chapters in mathematics education pertaining to teacher knowledge, beliefs, and curriculum implementation (Fennema & Franke, 1992; Hill, Sleep, et al., 2007; Philipp, 2007; Stein et al., 2007; A. G. Thompson, 1992). In addition, I read all pertinent studies referenced in the handbook chapters and obtained references from those studies. Lastly, I searched ERIC and PsychINFO databases for dissertations and published studies pertaining to K-12 mathematics instruction that attended to influential factors[1]. This database search resulted in a list of 1511 unique studies, 140 of which warranted closer review based on reading the abstract. I read each article to determine whether the study was empirical in nature (rather than a literature review or theoretical article), which aspects of teachers' instructional practice were investigated and how they were measured, and whether the study examined factors that influenced the quality of those aspects of instructional practice. I then limited the final set of studies that I reviewed to those that produced empirical results, focused on aspects of mathematics teachers' instructional practice that are broadly related to the enactment of cognitively demanding mathematical tasks, utilized classroom observation or student surveys to characterize instruction, and went beyond characterizations of instructional practice to examine factors influencing practice. Recall that by "factors influencing practice" I mean both factors that influence a teacher's practice and factors that influence whether teachers change their practice. A set of 63 studies met these criteria and were included in this broad review.

Taken together, these studies suggest that a large number of potentially relevant teacher and contextual factors might influence teachers' enactment of CDTs (see Figure 2). Because the review draws on 30 years of research in the mathematics education field which has continued to develop over time, some factors studied in the past now seem either less central or less useful in accounting for the quality of teachers' instructional practices (e.g., teacher reflectiveness). Other

[1] The following search strings were used: Eric:DE=("teaching methods" or teachers or "mathematics instruction" or "mathematics teachers") and AB=math* and AB=(factor* or influence* or effects*) and NOT DE=("college mathematics" OR "community colleges" OR "preservice teachers")
Psychinfo:DE=(teaching or "task complexity" or "teacher characteristics") and AB=(math* and (teach* or instruct*)) and AB=(factor* or influence* or effect*) AND NOT DE=("college teachers" or colleges)

factors have continued to be refined over the years and have been linked to teachers' instructional practice (e.g., beliefs about teaching and learning mathematics).

Synthesis: Refining Factors Relevant to the Enactment of CDTs

In an effort to narrow the list of factors from the broad literature review to a set of potentially relevant factors, I developed criteria for excluding, adjusting, or retaining potential factors. The most important criterion for whether factors were potentially relevant with regard to the enactment of CDTs was whether the factors could plausibly be linked to processes associated with enacting CDTs. A second criterion was whether factors were closely related to another key factor. When closely related and the other factor had more evidence linking it to the enactment of CDTs, the second factor was removed from the list of potentially relevant factors. In this section I first describe how I excluded or adjusted factors if they could not be linked to specific processes associated with enactment of CDTs or if they were closely related to another key factor. Then, for each of the factors that could be plausibly linked to processes of enacting CDTs, I describe those links.

Some factors that were identified as related to teachers' instructional practice in the broad review were excluded because they could not be linked to specific processes associated with enactment of CDTs. For example, several of the reviewed studies focused on teacher reflectiveness as a characteristic that is related to teachers' instructional practice in the classroom (Philipp et al., 1994; Smith, 2000; A. G. Thompson, 1984). While teachers who are reflective may learn from their practice, there are no direct links between reflectiveness and selecting high cognitive demand tasks or maintaining the cognitive demand of those tasks. For example, teachers who are reflective may analyze their teaching to determine how to make the procedural practice in their classroom more effective while others, who teach in more inquiry-oriented ways, might analyze students' questions or responses to learn about their students' thinking. But reflectiveness in itself is too broadly defined to influence the enactment of CDTs. Therefore, teacher reflectiveness was one of the previously investigated factors that I excluded from the set of potentially relevant factors. For similar reasons, I also excluded several other factors including job-satisfaction, teachers' dispositions toward math, and experience teaching math and with the curriculum (see Figure 2 and the column labeled "Excluded"). I excluded parent expectations and accountability pressures because they may be indirectly related to teachers' enactment of CDTs and buffered through school leader expectations, which is included in the set of potential factors.

In addition to excluding some factors, I also adjusted other factors identified in prior studies so that they could be linked to specific processes associated with enactment of CDTs. For example, school leader expectations was a factor that was indicated in only two studies, yet has the potential to influence teachers' enactment of CDTs. It required additional adjustment (mostly in the form of specification) in order to link it to specific processes associated with enactment of CDTs. The specific adjustments related to school leader expectations are described in detail below. A second potential factor that I also found necessary to adjust was professional development which is included in the category of learning opportunities through interactions with colleagues.

Lastly, I excluded some factors from the set of potentially relevant factors because they were closely related to another key factor. For example, teachers' goals, beliefs about mathematics, and beliefs about teaching and learning mathematics are all factors that were identified in prior studies. Yet, there is considerable evidence that teachers' goals are closely related to their beliefs (Aguirre & Speer, 1999; Schoenfeld, 2011; Sowder et al., 1998). Because of this close relationship, it is reasonable to focus on beliefs about mathematics and beliefs about teaching and learning mathematics and exclude teachers' goals. This is not to say that teachers' goals are inconsequential for teachers' enactment of CDTs, but rather that teachers' goals do not appear to add much insight to teachers' enactment of CDTs if beliefs about mathematics and beliefs about teaching and learning mathematics have already been considered. The factors that I excluded because they were closely related to teachers' knowledge or beliefs are teachers' beliefs about the curriculum, teachers' beliefs about affective issues, teachers' goals, and teachers' experiences as students (see Figure 2 and the column labeled "Closely Related").

As I made decisions about the potential significance of factors, two different categories of factors emerged: teacher factors and contextual factors. As specified above, teacher factors are specific to the individual teachers whereas contextual factors pertain to the classroom, school, or district context. For all of the potentially significant factors, I define the factor, summarize supporting evidence from the five studies of factors influencing the enactment of CDTs and the broad literature review, and give conceptual examples of how the factors might be linked to processes associated with the enactment of CDTs.

Teacher Factors

In this synthesis, I focus on three categories of teacher factors that encompass the bulk of the studies contained in the broad literature review (see Figure 2): teacher knowledge, teacher beliefs, and teacher skills. I concentrate on six constructs across the three categories and describe each below.

Teacher knowledge. Recent work has made progress in conceptualizing what it means for teachers to understand the content they teach and in clarifying other aspects of teacher knowledge that fall outside the traditional conceptualization of content knowledge yet are integral to the work of teaching. Based on this work and understandings of the processes associated with the cognitive demand of mathematical tasks in the classroom, there is some indication that two key categories of mathematics teacher knowledge might be integral to selecting high cognitive demand tasks and maintaining the cognitive demand of those tasks: mathematical knowledge for teaching, and knowledge of students' thinking.

Mathematical knowledge for teaching. Within the mathematics education community, mathematical knowledge for teaching (MKT) is generally conceptualized as a combination of subject matter knowledge and pedagogical content knowledge (PCK) (Hill, Ball, & Schilling, 2008). In the view of Hill and colleagues, subject matter knowledge is conceptual knowledge of mathematics that is necessary for solving mathematics problems but that is not specific to the work of teaching. On the other hand, PCK is specific to the work of teaching. Shulman (1986, 1987) introduced the notion of PCK and suggested that it "represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (1987, p. 228). Hence, MKT goes beyond pure mathematical content knowledge to also include PCK for mathematics teachers.

As indicated above, the empirical findings pertaining to the relationship between mathematical knowledge for teaching and the enactment of CDTs are mixed. However, the process-oriented argument made in Charlambous' (2010) study and the considerable evidence from the broad literature review provides some evidence that MKT might be related to teachers' instructional practice. Within this broader literature, some studies have explored mathematics PCK (Manouchehri & Goodman, 1998; Philipp et al., 1994), some have explored mathematics content knowledge , some have explored both aspects separately (Baumert et al., 2010; Escudero & Sánchez, 2007), and others have explored the two aspects together (Charalambous, 2010; Escudero & Sánchez, 2007; Hill, Ball, et al., 2007; Hill, Blunk, et al., 2008; Manouchehri & Goodman, 2000; Sherin, 2002; Yun-peng et al., 2006). These studies suggest that mathematical knowledge, pedagogical content knowledge, and combinations of the two (i.e., mathematical knowledge for teaching) are related to teachers' instructional practice.

The conceptual connections between mathematical knowledge for teaching and processes of selecting CDTs and maintaining the cognitive demand of those tasks provide further evidence that teachers with deeper MKT may be more likely to provide more cognitively demanding learning opportunities to their students. There are indications that MKT is integral to teachers' work as they try to understand the mathematical ideas within tasks and decide whether those ideas and thus the tasks are worthwhile (Clarke, 2008). In addition, teachers with deeper MKT are more likely to feel more confident in their ability to handle CDTs in the classroom (Sowder et al., 1998). This, in turn, might make it more likely that teachers will select CDTs for their students.

Teachers' MKT is also likely to play a critical role in maintaining the cognitive demand of high-level tasks. MKT is integral to teachers' decisions about how to support students and to their ability to provide that support (Henningsen & Stein, 1997; Herbst, 2003; Stein et al., 1996). For example, MKT seems to be integral to teachers' selection and use of representations to help students make sense of mathematical ideas (Charalambous, 2010), which might have an impact on maintenance of the cognitive demand of high-level tasks (Herbst, 2003). This same argument can be made for the potential importance of MKT on other instructional decisions teachers make in the classroom (e.g., which questions they ask students and when to ask those questions). Therefore, there is some indication that MKT might be related to the enactment of CDTs in the classroom.

Knowledge of students' thinking. Most definitions of MKT include "knowledge of students and content." This aspect of MKT includes deep knowledge of how students typically progress in learning particular mathematical concepts (Hill, Schilling, & Ball, 2004). As it is defined, that type of knowledge is not specific to the students in a teacher's classroom, but instead, concerns typical students. For instance, it might include a teacher's understanding of how students (in general) develop their understanding of fractions, but it does not include knowledge about how the students in the teacher's classroom fit within that learning trajectory.

Another aspect of teachers' knowledge of students is important: knowledge of their individual students' thinking. This is the type of understanding that Choppin (2011) found is related to teachers' enactment of CDTs. He found that teachers who attended to their students' thinking were better at selecting CDTs and implementing them in cognitively demanding ways. Further, there is a body of work pertaining to teachers' knowledge of their students that has stemmed from work on Cognitively Guided Instruction (CGI) (e.g., Carpenter et al., 1989; Franke & Kazemi, 2001; Peterson et al., 1989). One study of the CGI program demonstrated that teachers' knowledge of students' thinking is related to teachers' instructional practice (Peterson et al., 1989). In particular, the findings suggest that teachers with greater knowledge of students' thinking are more likely to provide cognitively challenging learning opportunities for students. Also, CGI professional development focused on helping teachers understand the development of children's mathematical thinking changed teachers' instructional practices (Carpenter et al., 1989; Fennema et al., 1996). Those studies found that changes in teachers' knowledge of students' thinking coincided with changes in teachers' instructional practice. This suggests that there is a relationship between teachers' knowledge of students' thinking and teachers' instructional practice.

The connections between teachers' knowledge of students' thinking and processes of selecting CDTs and maintaining the cognitive demand of those tasks provide further evidence that teachers with knowledge of their students' mathematical thinking might be more likely to provide more cognitively demanding learning opportunities to their students. Knowledge of students' thinking is likely to feature prominently in teachers' decision-making as it involves coordinating information about students with their teaching goals (Henningsen & Stein, 1997; Lampert, 2001). For example, if teachers know more about their individual students' thinking, then they might be better able to judge whether tasks are appropriate for their students (Peterson et al., 1989). As with MKT, teachers' knowledge of students' thinking is also likely to play a critical role in maintaining the cognitive demand of high-level tasks through the provision of the appropriate levels of support in the classroom. If teachers have deep knowledge of their individual students' thinking, then they might be more likely to provide appropriate scaffolding that supports students without decreasing the cognitive demand of high-level tasks (Choppin, 2011; Henningsen & Stein, 1997; Peterson et al., 1989). Hence, given the evidence from Choppin's (2011) study and the processes associated with the selection of tasks and maintenance of the cognitive demand of high-level tasks, knowledge of students' thinking might be related to the enactment of CDTs in the classroom.

Teacher beliefs. I follow Thompson (1992) in drawing the following distinction between knowledge and beliefs: beliefs can be held to varying degrees and they are not necessarily agreed upon whereas knowledge is agreed upon by a professional community and assumed to be true. In this review, I consider references to teachers' "views", "conceptions", "ideas", or "beliefs" pertaining to mathematics or teaching and learning mathematics as *teachers' mathematics-related beliefs*. A considerable number of studies have investigated how mathematics teachers' beliefs are related to their instructional practice. Before reviewing the studies, I define three crucial dimensions of teachers' mathematics-related beliefs: beliefs about mathematics, beliefs about teaching and learning mathematics, and beliefs about students' mathematical capabilities.

Beliefs about mathematics pertain to conceptions of the discipline of mathematics (A. G. Thompson, 1992). One key distinction that underlies many of the characterizations is the distinction between a relational understanding of mathematics and an instrumental understanding of mathematics (Skemp, 1978). An instrumental understanding of mathematics involves thinking about mathematics as a set of unrelated tools used to solve problems in a procedural manner. In contrast, a relational understanding of mathematics involves thinking about mathematics as a set of conceptual relationships that enable people to solve problems in a variety of ways. Viewing mathematics as relational is more compatible with the enactment of CDTs in that CDTs allow for multiple solution methods and require students to draw on their conceptual understanding to devise an appropriate solution to the task.

Beliefs about teaching and learning mathematics extend beyond conceptions of the discipline and include ideas about how students learn best and what should happen instructionally. These beliefs concern what teachers' consider to be high quality teaching rather than what they do themselves in their classrooms, although the two might be related. For the sake of clarity, I define beliefs about teaching and learning mathematics that are compatible with inquiry-oriented instruction as *inquiry-oriented* beliefs. Different aspects of inquiry-oriented beliefs appear to be interrelated. For example, teachers' beliefs about the types of tasks in which students should engage are likely to be linked to their beliefs about how students learn by working on tasks (A. G. Thompson, 1992). Analogously, a teacher who has developed inquiry-oriented beliefs might believe that students learn math best by participating in a classroom where those aspects of instruction are present (A. G. Thompson, 1992). Some key aspects of inquiry-oriented beliefs about teaching mathematics include the importance of high cognitive demand tasks, the importance of discourse, and the proactive role of the teacher in scaffolding students' work on those tasks and facilitating discussion (Munter, Under review).

In contrast to beliefs about mathematics and about teaching and learning mathematics, teachers' beliefs about students' mathematical capabilities have received less attention in the literature but could be integral to teachers' enactment of CDTs. These beliefs about students' mathematical capabilities include teachers' beliefs about student motivation, inherent ability, and teacher and student agency with regard to altering motivation and ability. The notion of students' mathematical capabilities builds on Horn's (2007) work. Horn examined teachers' category systems with regard to students' mathematical capabilities as groups of teachers discussed students' learning of mathematics. She found that some teachers described students' capabilities in terms of inherent characteristics (e.g., lazy kids, fast kids, slow kids, etc.) whereas other teachers situated students' capabilities with respect to the learning opportunities provided to them in the classroom. Further, when teachers characterized students' capabilities in terms of inherent characteristics, they tended to describe the problem of student achievement as residing with the students or their families and communities rather than as something that they could influence through instruction. If we are interested in enabling all students to engage in cognitively demanding mathematical activity, then a particular set of beliefs about students' mathematical capabilities is more productive: that is, beliefs that do not treat students' ability as inherent, but instead explain students' capabilities in terms of learning opportunities provided to them.

There are signs that the relationship between the three types of teachers' mathematics-related beliefs and the enactment of CDTs might be complicated. In particular, it seems that teachers' beliefs about mathematics and beliefs about teaching and learning mathematics are related to teachers' instructional practice, but that their influence depends on their alignment. In other words, if one type of a teacher's beliefs is compatible with the enactment of CDTs, but another type is not compatible with the enactment of CDTs then that teacher might be less likely to enact CDTs than a teacher with aligned and compatible beliefs. Because of the potential complexity of these relationships, I first review the evidence that the individual types of beliefs are related to teachers' instructional practice, discuss the evidence pertaining to the complexity of the relationships in greater detail, and then review the evidence that these same relationships might hold for the enactment of CDTs.

Only one study of the enactment of CDTs focused specifically on mathematics teacher's beliefs. Son's (2008) study identified several dimensions of teachers' beliefs about teaching and learning mathematics as factors that contributed to the enactment of CDTs. However, her analysis did not explicate the mechanisms behind those relationships. Aspects of beliefs about teaching and learning mathematics on which she focused included conceptions of knowing and doing mathematics and perceptions of student achievement (described as "students' diversity in terms of mathematics ability" (Son, 2008, p.219). Her definition of teachers' perceptions of student achievement seems to be related to teachers' beliefs about students' mathematical capabilities. Although Son's results give some suggestion that teachers' beliefs about teaching and learning mathematics, including beliefs about students' mathematical capabilities, might be related to their enactment of CDTs, additional process-oriented investigations and large-scale studies are needed to better understand this relationship. None of the studies of factors related to teachers' enactment of CDTs have focused on teachers' beliefs about mathematics.

A number of studies included in the larger review found relationships between one of the three types of mathematics teachers' beliefs and their practice. In particular, there is evidence that beliefs about mathematics (Beswick, 2005; Philipp et al., 1994; Putnam et al., 1992; Remillard, 1999; Sowder et al., 1998; Stipek et al., 2001; Stodolsky & Grossman, 2000; A. G. Thompson, 1984), beliefs about teaching (Barrett Paterson, 2009; Beswick, 2005; D. K. Cohen, 1990; Manouchehri & Goodman, 2000; A. G. Thompson, 1984), beliefs about student learning (Son, 2008; Stodolsky & Grossman, 2000; Yun-peng et al., 2006), beliefs about students' needs (Prawat & Jennings, 1997; Sullivan & Leder, 1992; Sztajn, 2003), expectations for students (Jamar & Pitts, 2005), beliefs about student motivation (Cooney, 1985; Manouchehri, 2004; Turner et al., 2011), beliefs about teaching and learning mathematics (Lloyd, 1999; Manouchehri & Goodman, 1998; Philipp et al., 1994; Putnam et al., 1992; Remillard, 1999; Remillard & Bryans, 2004; Stipek et al., 2001; Superfine, 2009) and mathematics teachers' beliefs, in general (Aguirre & Speer, 1999; Schoenfeld, 2011; Skott, 2001), are related to teachers' instructional practice. Overall, the evidence suggests that beliefs about teaching and learning mathematics, beliefs about students' mathematical capabilities, and beliefs about mathematics are related to teachers' instructional practices. However, several studies suggest that it is important to consider multiple types of mathematics-related beliefs when attempting to understand how they relate to teachers' instructional practice because the relationship is not always straightforward. In the following paragraphs, I review the evidence that the enactment of CDTs might require that not only are teachers' mathematics-related beliefs compatible with the enactment of CDTs, but also that different types of beliefs are all aligned.

There are some indications that the relationship between beliefs about teaching and learning mathematics and teachers' practice is not straightforward. In particular, several case studies suggest that teachers' beliefs about teaching and learning mathematics are not consistent with teachers' actual classroom practices. These studies of individual teachers attempted to document both participants' beliefs about teaching and learning mathematics and their teaching practices, and then examined the degree of congruence (D. K. Cohen, 1990; Cooney, 1985; Raymond, 1997; Skott, 2001). For example, Cohen (1990) studied one teacher's practice in the context of state-wide reform of mathematics instruction. He found that she believed she had adopted the reform notions of teaching and learning but the changes in her practices did not match, she had filtered the reform practices through her traditional approach to teaching.

One possible explanation for this lack of congruence between teachers' beliefs about teaching and learning mathematics and their instructional practice focuses on a lack of alignment between teachers' beliefs about teaching and learning mathematics and their beliefs about mathematics. Several studies suggest that beliefs about mathematics are, in fact, more closely related to teachers' instructional practice than teachers' beliefs about teaching and learning mathematics (Cross, 2009; Lloyd & Wilson, 1998; Raymond, 1997). Results from Raymond's (1997) case study of one beginning elementary teacher suggest that the teacher's beliefs about mathematics, rather than beliefs about teaching mathematics, were more closely aligned with her instructional practices. Both Cross (2009) and Lloyd and Wilson (1998) suggest that beliefs about mathematics are part of the foundation for mathematics teachers' beliefs and practice, and that changes in teachers' beliefs about mathematics can lead to changes in teachers' beliefs about teaching and learning mathematics. In the case of Cohen's (1990) focal teacher, although she seemed to have adopted more reform-oriented beliefs about teaching mathematics, she may have struggled to enact the reform teaching practices because her beliefs about mathematics had not changed.

There is some evidence that the existence of a relationship between teachers' beliefs about mathematics or teaching and learning mathematics and their instructional practices might also require that their beliefs about students' mathematical capabilities are aligned. Cooney (1985) studied one novice teacher and found that despite the fact that his beliefs about mathematics and teaching mathematics were consistent with a problem-solving approach, his beliefs about students and motivation appeared to more directly influence his choice of tasks for students. The teacher believed that he needed to use "recreational" mathematics problems to catch his students' interest and used superficially interesting problems that were unconnected to key mathematical ideas to capture students' attention. Hence, his beliefs about student motivation were not aligned with his other mathematics-related beliefs and seemed to be more closely linked to his instructional practice. Therefore, there is evidence that alignment between different types of teachers' mathematics-related beliefs is likely to be important in identifying relationships between particular dimensions of teachers' beliefs and their instructional practice.

To this point, I have summarized the one study examining how teachers' mathematics-related beliefs are related to their enactment of CDTs (Son, 2008) and summarized the evidence from the larger review that suggests how the different types of teachers' mathematics-related beliefs are related to teachers' instructional practice. In this section I review the conceptual evidence that teachers' beliefs about mathematics, teaching and learning mathematics, and students' mathematical capabilities might be integral to teachers' enactment of CDTs. Further, it is likely that it will also be equally important that different types of teachers' mathematics-related beliefs be aligned in order for any one of them to be related to teachers' enactment of CDTs. First, with regard to the selection of tasks to pose to students, teachers' goals for instruction and their beliefs about what it means to do mathematics and beliefs about how students learn best are likely to influence their decision (Schoenfeld, 2011). For example, teachers who hold an instrumental view of mathematics or believe that students learn best by practicing procedures that they have been shown how to use might be more likely to choose tasks that give students opportunities to practice procedures. This relationship becomes more complicated if there are teachers who hold instrumental views of mathematics but believe that students learn best through productive struggle with CDTs. Those teachers are unlikely to consistently enact CDTs despite their inquiry-oriented beliefs about student learning. Therefore, individual types of teachers' mathematics-related beliefs might be related to the enactment of CDTs provided that they are aligned with other dimensions of those teachers' mathematics-related beliefs.

Beyond the selection of high-level tasks, teachers' mathematics-related beliefs might also be related to maintenance of the cognitive demand of those tasks. In particular, they could influence maintenance of the cognitive demand through the level or type of support that teachers provide to students. For example, as with the choice of tasks to pose to students, teachers' beliefs about how students learn best might influence the types of supports they provide for students in the classroom (Smith, 2000; Sztajn, 2003). If teachers believe that students learn best by practicing procedures that they have been shown how to use, yet the teachers do select a CDT, they might attempt to support students by first demonstrating how to solve similar tasks, thereby decreasing the cognitive demand of those tasks. On the other hand, if they believe that students should struggle to solve CDTs teachers might be less likely to intervene in ways that decrease the cognitive demand of high-level tasks. Further, if they generally believe that students learn best through productive struggle with CDTs but believe that the majority of the students in their class are not capable of solving CDTs, then they might not be as likely to enact CDTs with that class of students (Henningsen & Stein, 1997). Again, the compatibility of particular types of beliefs with the enactment of CDTs along with the alignment between different types might be critical to how teachers support students. Therefore, there is evidence that it might be important to consider teachers' beliefs about mathematics, teaching and learning mathematics, and students' mathematical capabilities, as well as the alignment between those different dimensions of teachers' mathematics-related beliefs in considering how they relate to the enactment of CDTs.

Teacher skills. One teacher capability that should not be overlooked in its potential to facilitate the enactment of CDTs is classroom management skills. Studies of teacher quality from over 20 years ago identified links between teachers' classroom management skills and student achievement (Brophy, 1986; Evertson, Emmer, & Brophy, 1980; Good & Brophy, 1987). That line of work has been criticized for its inattention to the content (e.g., mathematics) of instruction (Confrey, 1986), and recent work has taken on a stronger content-based focus. While this criticism has some validity, classroom management skills are still acknowledged as critical in more recent studies of mathematics teachers' instructional practice. Many studies, particularly those of new teachers, indicate the importance of teachers developing adequate classroom management if they are to provide their students cognitively challenging learning opportunities (e.g., D. K. Cohen, 1990; Hill, Blunk, et al., 2008; Opdenakker & Van Damme, 2006; Raymond, 1997; Skott, 2001; Sullivan & Leder, 1992). Stein and colleagues found that a lack of classroom management was one reason for decline in cognitive demand of high-level tasks (Henningsen &

Stein, 1997; Stein et al., 1996). In particular, if teachers are unable to manage how their students engage in classroom activities, they might find it difficult to enact CDTs because high cognitive demand tasks tend to gives students more freedom and depend on students working together in groups.

In sum, teacher factors that have the potential to be related to teachers' enactment of CDTs and should be investigated further include: (a) mathematical knowledge for teaching (including content knowledge and pedagogical content knowledge), (b) knowledge of students' thinking, (c) beliefs about mathematics, (d) beliefs about teaching and learning mathematics, (e) beliefs about students' mathematical capabilities, and (f) classroom management skills.

Contextual Factors

The categories of contextual factors within the set of factors identified in the literature as promising include: class time, class size, characteristics of students, departmental culture, school leader expectations, the nature of the curriculum, and learning opportunities through interactions. Contextual factors have received less attention and are less developed in the literature than the teacher factors (see Figure 2), yet most of the studies of teachers' knowledge and beliefs mentioned the importance of considering the context in which teachers work. I report evidence from the broad literature review concerning the relationships between contextual factors and teachers' instructional practice and give examples of how the contextual factors might be related to processes associated with the enactment of CDTs to justify their inclusion in the list of promising factors to investigate.

Class time. Several studies give accounts of teachers who reported that the amount of class time available to work on challenging tasks influences their instructional practice (Barrett

Paterson, 2009; Cooney, 1985; Raymond, 1997; Son, 2008). In particular, teachers in Son's study of enacting CDTs who did not use CDTs in their classroom explained that they felt they had limited time to cover the content and that CDTs would take too much time. Specific to the enactment of CDTs in the classroom, Henningsen and Stein (1997) reported that time allocated to work on tasks (either too much or too little) was a common reason for decline in cognitive demand of high-level tasks. Although teachers generally see a lack of time as the primary constraint, too much time was also a reason for decline in cognitive demand in Henningsen and Stein's study. Hence, it seems that sufficient time needs to be allocated for work on CDTs and that time needs to be carefully managed to maintain the cognitive demand of high-level tasks.

Class size. Another contextual factor that appears to be an important consideration in teachers' enactment of CDTs is the size of the class (e.g., the number of students in the class). Rousseau and Powell (2005) studied four secondary teachers' mathematics reform efforts and found that the teachers mentioned the size of their classes as a reason why enacting CDTs was difficult. Teachers who had large classes explained that they did not engage students in collaborative activity because that type of activity would be difficult to manage with a large class. Productive learning opportunities for students with challenging tasks often involve students working in groups (E. G. Cohen, 1994). The large number of students in the classroom makes this type of group interactive activity more difficult, possibly due to space constraints or the challenge of keeping all groups engaged. Although the size of the class does not necessarily prohibit the enactment of CDTs in the classroom, it is likely to make it more challenging to do so.

Student characteristics. Students are crucial partners in the enactment of CDTs. As described above, there are important aspects of teachers' knowledge and beliefs that might be related to the enactment of CDTs, and the specific relationships between teachers' knowledge, beliefs, and practice will depend on the individual students in the class. Also, there is considerable evidence that teachers perceive students as having a great impact on their instructional practices. Several studies suggest that there is a relationship between students' background knowledge or the ability level of the class and teachers' instructional practice (Anderson et al., 2005; Goos et al., 2007; McGinnis et al., 2004; Opdenakker & Van Damme, 2006; Son, 2008; Yun-peng et al., 2006). One study suggests that the amount of student mobility and absenteeism is related to teachers' instructional practice (Rousseau & Powell, 2005). Lastly, a number of studies describe students' expectations as influential (Cooney, 1985; Herbel-Eisenmann et al., 2006; McGinnis et al., 2004; Sullivan & Leder, 1992). The inclusion of student characteristics in the list of promising factors to investigate should not be construed as implying that enacting CDTs is only possible with certain students. CDTs are appropriate for students with different levels of background knowledge (Zohar & Dori, 2003); yet, the challenges associated with and resources necessary for successful enactment might vary depending upon the individual students in the class (Stein et al., 1996). Therefore, in attempting to understand teachers' enactment of CDTs across a variety of contexts, it will be important to consider the characteristics of the students in the class.

Departmental culture. I define departmental culture to be the norms and attitudes of the mathematics colleagues with whom teachers work. Several studies have found that the culture of the mathematics department has an impact on teachers' instructional practice (Manouchehri & Goodman, 1998; McGinnis et al., 2004; Stodolsky & Grossman, 2000). For example, Stodolsky and Grossman (2000) combined case study and survey data to study teachers who worked in schools with changing student populations and found that if colleagues are generally resistant to changing their instructional practices, then it may be more difficult for individual teachers to implement new practices. Similarly, if the other teachers in the school do not teach in a way that is consistent with a teacher's practice, then that teacher may feel pressure to adjust her practice to match those of her colleagues (McGinnis et al., 2004). These pressures could have an impact on both teachers' selection of CDTs and maintenance of the cognitive demand of those tasks. With regard to selection, if the mathematics teachers in the school generally select low cognitive demand tasks then there might be normative pressure for all teachers to do the same (Bidwell & Yasumoto, 1999). Similarly, if the norm in the school is to use CDTs but teach in ways that decrease the cognitive demand of tasks as they are implemented, then a teacher might conform to those pressures and decrease the cognitive demand. Hence, if a teacher's colleagues are generally not attempting to enact CDTs, then that teacher could find it challenging to select high cognitive demand tasks and maintain the cognitive demand of those tasks in her classroom.

School leader expectations. Another contextual factor that is likely to have an impact on teachers' practice is school leader expectations for instructional practice. First, there is evidence that school leaders' instructional leadership is critical to teachers' instructional improvement (Bryk, Sebring, Allensworth, Luppesco, & Easton, 2010). For example, school leaders might respond to the NCLB requirements by expecting teachers to develop high quality instruction or by expecting teachers to engage in test preparation. These different expectations for teachers' instructional practice might have implications for their practice itself. Based on the evidence below, there are some signs that if school leaders' expectations include the enactment of CDTs, then teachers are more likely to select high cognitive demand tasks and maintain the cognitive demand of those tasks.

Several studies within the broad literature review identified the support from or expectations of the principal—the primary school leader—as a factor that influences teachers' instructional practice (Anderson et al., 2005; Manouchehri & Goodman, 1998; Philipp et al., 1994). Surprisingly, this factor was only investigated in three studies that qualified for the review. It is important to note that other studies that did not meet the review criteria have investigated the influence of principal support and expectations on teachers' practice or student achievement (Quinn, 2002; Robinson, Lloyd, & Rowe, 2008; Stein & Nelson, 2003; Trenkamp, 2007). These studies typically did not meet the review criteria because they assessed teachers' instructional practice via teacher survey rather than classroom observation. There is some evidence that if expectations from the principal are aligned with reform efforts, then those reform efforts are more effective (Coburn, 2005). Hence, in order for principals' expectations to more directly influence teachers' enactment of CDTs in the classroom, those expectations might need to specifically include the enactment of CDTs. Also, as mentioned above, schools leaders can buffer the influence of other external pressures. For example, the pressures of accountability (Barraugh, 2011; McGinnis et al., 2004; Rousseau & Powell, 2005; Son, 2008; Yun-peng et al., 2006) and parental expectations (Herbel-Eisenmann et al., 2006; McGinnis et al., 2004) were both identified as factors influencing teachers' instructional practice in the literature review. By clarifying expectations for teachers to enact CDTs in the face of testing pressures and communicate about those expectations with parents, school leaders can help to buffer the external pressures and support teachers' enactment of CDTs. In sum, there is some suggestion that if school leaders' expectations include the enactment of CDTs then teachers might be more likely to select high cognitive demand tasks and maintain the cognitive demand of those tasks.

Curriculum. The term curriculum has many different meanings. Recall that I use the term *curriculum* to mean the adopted curriculum and/or supplementary materials that a teacher uses in the classroom. In the following paragraphs I describe the evidence indicating that the more the curriculum is specifically designed to support teachers' enactment of CDTs, the more likely teachers might be to select high-level tasks and maintain the cognitive demand of those tasks.

A handful of studies within the broad literature review have concluded that straightforward curriculum-related characteristics such as the type of adopted curriculum (e.g., inquiry-oriented or traditional) (Barraugh, 2011; Herbel-Eisenmann et al., 2006; McGinnis et al., 2004; Rousseau & Powell, 2005; Son, 2008) and teachers reliance on the adopted curriculum (or use of supplementary resources) (Lambdin & Preston, 1995; Manouchehri & Goodman, 2000; Son, 2008) are related to teachers' instructional practice. These studies suggest that the more teachers have access to materials with high cognitive demand tasks as part of either the adopted curriculum or supplementary materials, the more likely they are to provide cognitively challenging learning opportunities for students. Therefore, some studies suggest that an important first step in supporting teachers' enactment of CDTs might be to provide them with curriculum materials containing CDTs.

As noted above, findings from the studies of factors related to the enactment of CDTs are less definitive: Son's (2008) study suggests that the type of textbook (inquiry-oriented or traditional) was related to enactment of CDTs, while Boston and Smith (2009) expected but did not find the same result. In the case of the Boston and Smith study, they characterized the type of curriculum adopted by the school district as either inquiry-oriented or traditional. There is evidence that teachers do not necessarily use the school or district-adopted curriculum in their classrooms (Freeman & Porter, 1989). Hence, this characterization might not capture the nature of the materials that the teachers in their study actually used in their classrooms. On the other hand, Son characterized the textbook that was the source of tasks used in the classroom, regardless of whether tasks came from the adopted curriculum or supplementary materials. And Son found that type of textbook characterized in this way was related to teachers' enactment of CDTs. This suggests that the nature of the materials from which teachers select tasks might have an impact on the extent to which they select high cognitive demand tasks. Further, their choice between the adopted curriculum and supplementary materials might be related to their beliefs about mathematics and teaching and learning mathematics (Remillard, 1999). Although it is important to consider the nature of the adopted curriculum as a critical resource for teachers' enactment of CDTs, it is also essential to consider any supplementary materials that they actually use as well.

Several studies in the broad review go beyond the availability of challenging tasks in the curriculum to suggest that the nature of support for teachers provided by the curriculum is related to teachers' instructional practice (Diaz, 2004; Remillard, 1999; Stein & Kaufman, 2010; Wang & Paine, 2003). Supports within the curriculum might include ties to the state standards or examples of student responses. Wang and Paine (2003) studied the development of a beginning Chinese middle-school teacher and found that the deliberately supportive structure of Chinese curriculum materials (e.g., well-articulated teaching objectives) supported her implementation of that curriculum.

Consistent with these findings, Stein and Kaufman found a relationship between the nature of support and learning opportunities within the curriculum and the enactment of CDTs. They studied teachers' use of two different inquiry-oriented elementary mathematics curricula and found that one of the curricula provided more support for teachers that resulted in higher quality enactment of CDTs. They defined support as “additional information written specifically for teachers in order to help them better understand and teach the lessons” (2010, p.666). For example, the more supportive curriculum that they studied gave teachers additional information (e.g., clear objectives, details about student thinking) to help them understand the big mathematical ideas (i.e., concepts or ideas that are the focus of the lesson). An understanding of the big mathematical idea can help teachers understand the overall goal for the lesson and how the different aspects of the task serve that goal, which might support them in maintaining the cognitive demand of tasks. Therefore, there is some indication that the greater the supports provided by the curriculum to enact CDTs, the more likely teachers might be to select high-level tasks and maintain the cognitive demand of those tasks.

Learning opportunities through interactions. Opportunities to learn on the job include formal PD as well as informal learning opportunities through conversations with colleagues and while working in the classroom with students. I follow Kaufman and Stein (2010) by defining define *learning opportunities through interactions with others* to include both formal PD and informal interactions with colleagues, but exclude work in the classroom with students. I exclude work in the classroom with students because those learning opportunities are likely to require design of a different nature. In the following paragraphs I summarize the evidence that specific types of interactions with colleagues might support teachers to improve their enactment of CDTs in the classroom.

First, there are strong *a priori* theoretical reasons to believe that certain types of interactions with colleagues might to support teachers’ development of enacting CDTs. In particular, studies of professional learning indicate that co-participation in activities that approximate the targeted practices with more accomplished others is critical for the learning of complex practices (Bruner, 1996; Lave & Wenger, 1991). Further, these activities typically occur over an extended period of time (Lave & Wenger, 1991). Therefore, there are theoretical grounds for the notion that interactions with colleagues that are ongoing in nature, involve co-participation in activities that are related to the enactment of CDTs, and involve colleagues who are relatively accomplished in enacting CDTs, might support teachers' in enacting CDTs.

In addition to theoretical grounds, there is empirical evidence that interactions with colleagues in several different settings can support teachers in developing their instructional practice, and in one case their enactment of CDTs. Reviewing the empirical evidence in light of the theories of professional learning helps to clarify what might be potentially productive interactions with colleagues.

There is evidence that PD can support teachers' development, and that ongoing PD focused on the enactment of CDTs can influence teachers' enactment of CDTs. First, the findings of several studies from the broad review suggest that participation in formal PD is related to change in teachers' instructional practice (Barton, 2005; Remillard, 1999; Yun-peng et al., 2006). Others go further to emphasize the content of the PD and suggest that PD focused on supporting teachers' development of particular types of knowledge, beliefs, and practice is likely to result in related changes in teachers' instructional practice (e.g., Carpenter et al., 1989; Fennema et al., 1996; Turner et al., 2011). As described above, Boston and Smith (2009) found that ongoing PD focused on the enactment of CDTs had an impact on participants' enactment of CDTs in the classroom. The PD program they studied was theoretically sound in that it was ongoing and involved activities that approximated the enactment of CDTs with accomplished peers and mathematics education experts. However, while teachers' enactment of CDTs generally improved, they did not all enact CDTs successfully after participating in PD. Therefore, that there is still more to be understood about designing effective PD for supporting teachers' enactment of CDTs.

Within the broad review, there is some evidence that work with expert colleagues, collaborative meetings and one-on-one interactions with colleagues can all support changes in teachers' instructional practice. However, there has been little simultaneous attention to expertise and interactions. For example, some studies suggest the importance of access to colleagues in the school with mathematics instructional expertise (Manouchehri & Goodman, 1998), often in the form of mathematics coaching. Other studies have found relationships between change in teachers' instructional practice and both teacher collaboration through meeting together (especially about the curriculum) (D. K. Cohen, 1990; Diaz, 2004; Smith, 2000; Wang & Paine, 2003) and peer coaching (Bruce & Ross, 2008). Despite the evidence from the professional learning literature suggesting that co-participation with relatively accomplished colleagues is an important criterion for the productivity of interactions, Diaz's (2004) study was the only study in the broad review that investigated interactions with experts and collaboration with colleagues. Diaz's dissertation study included two components: 1) a study of four second grade teachers' use of curriculum materials, and 2) a cross-site comparison of five studies—her findings and studies by Collopy (2003) Lloyd and Behm (2002), Remillard (2000), and Schneider and Krajcik (2002). Diaz's findings from the cross-site comparison suggest both the importance of interactions with experts and teacher meetings. However, she analyzed them as two separate supports and did not investigate whether expertise is important in teacher collaboration.

In addition, none of these studies of settings outside of formal PD have focused on the activities that occur within the interactions, which is an important aspect of the theoretical grounding for the potential productivity of such interactions. Further, there is some empirical evidence that the occurrence of interactions with colleagues is not necessarily sufficient to support teacher learning unless the interactions focus on problems of practice (Horn & Little, 2010). Given the productivity of the interactions described in the studies above, it is likely that the interactions focused on problems of practice, despite the lack of emphasis on the content of interactions within the studies. Therefore, there is theoretical and empirical evidence suggesting that interactions with colleagues can provide valuable learning opportunities for teachers, assuming that they involve ongoing interactions centered on the practice of enacting CDTs with relatively accomplished colleagues.

Accounting for Contingencies Associated with Factors Influencing the Enactment of CDTs

Beyond generating a list of promising factors to investigate, I also set out to understand an effective approach for studying these factors on a large scale. In conducting the review, I found that most of the large-scale studies treat the factors I have discussed as if they are related to teacher practice in straightforward ways, while the small-scale studies frequently describe contingencies associated with the ways that particular factors influence practice. For example, large-scale studies have investigated relationships between teachers' knowledge and their practice while a small-scale study might describe how the expected relationship between teachers' knowledge and their practice does not appear to hold for a particular teacher under particular circumstances. Despite this trend, it is necessary and possible to consider contingencies when conducting large-scale studies.

We need to investigate more precisely how factors are related to the enactment of CDTs on a large scale by considering factors and associated contingencies. I draw on information about contingencies associated with particular relationships from small-scale studies within the broad review to suggest contingencies that should be investigated in future large-scale studies of the enactment of CDTs. In this section, I illustrate this approach for two different categories of factors from the set of factors to investigate: teachers' knowledge and teachers' beliefs. One reason I focus on contingencies associated with teachers' knowledge and beliefs in this paper is because, of all of the factors, they have received the most attention in both small- and large-scale studies. Because contextual factors have received much less attention in the literature, it will be much harder to develop conjectures about contingencies.

Contingencies Associated with Mathematics Teacher Knowledge

Despite all of the evidence that MKT is integral to enacting CDTs, there is evidence from small-scale studies that the relationship between MKT and the enactment of such tasks has some associated contingencies. I focus on three key contingencies in the relationship between MKT and the enactment of CDTs: beliefs about teaching and learning math, beliefs about students' capabilities, and curriculum supports.

First, there is evidence that that the relationship between MKT and instructional practice is contingent on teachers' beliefs about teaching and learning mathematics (Ball, 1991; Schoenfeld, 2011; Turner et al., 2011). The studies signaling this contingency suggest that an inquiry-orientation in teachers' beliefs about teaching and learning mathematics is necessary for teachers who have developed sufficient mathematical knowledge for teaching to support conceptually-rich mathematical activity in the classroom. In particular, several studies suggest that beliefs about teaching and learning mathematics that are incongruent with inquiry-oriented practice can limit teachers' ability to enact mathematical tasks in conceptual ways. Schoenfeld

(2011) demonstrated that one teacher who seemed to have developed relatively sophisticated mathematical knowledge but his belief that "what he said should be an elaboration or clarification of what a student had said," (p.82) limited the conceptual resources within the classroom. Ball (1991) described a similar phenomenon and argued that teachers' beliefs about teaching and learning mathematics might interfere with whether or not they draw on their mathematical knowledge.

With regard to more specific beliefs about learning mathematics, there is also evidence that the relationship between teacher knowledge and the enactment of CDTs might be contingent on teachers' beliefs about students' capabilities. Turner, Warzon, and Christensen (2011) studied changes in three middle school mathematics teachers' beliefs about motivation and instructional practices, and found that one teacher with a high level of mathematical knowledge for teaching did not enact conceptually-rich mathematical activity in the classroom and held negative views of students' abilities. In other words, she did not believe that her students were motivated or capable of being motivated to engage in conceptually-challenging mathematical tasks. Hence, unproductive beliefs about students' mathematical capabilities limited her instructional practice even though her mathematical knowledge for teaching was deep. This finding indicates that the relationship between mathematics teacher knowledge and the enactment of CDTs might be contingent on productive beliefs about students' capabilities.

The relationship between mathematics teacher knowledge and the enactment of CDTs is likely to be contingent on the supports available within the curriculum used by teachers. One dissertation study of factors influencing seven upper elementary grade teachers' mathematics instruction found a positive effect of supportive curricular materials for teachers with low levels of MKT (Barraugh, 2011). The author suggests that in cases where teachers have less developed

MKT, by using well-designed, reform-oriented curricular materials, teachers may be supported to enact higher quality instructional practices than would otherwise be expected given their MKT. Therefore, there is some indication that the influence of MKT on the enactment of high cognitive demand tasks is contingent on supports available through the curriculum.

No studies have investigated contingencies associated with the relationship between knowledge of students' thinking and teachers' instructional practice. Yet, it is plausible that the same contingencies hypothesized for MKT and the enactment of CDTs also occur for knowledge of students' thinking.

Contingencies Associated with Mathematics Teacher Beliefs

Although there is evidence that aligned beliefs about mathematics and beliefs about teaching and learning mathematics might be related to the enactment of CDTs, there may be several key contingencies associated with these relationships. Those contingencies include: (a) teachers' knowledge, (b) curriculum supports, and (c) student characteristics.

First, the relationship between mathematics teachers' beliefs and the enactment of CDTs is likely to be contingent on aspects of teachers' mathematical knowledge for teaching. Ball (1991) describes mathematical knowledge as a "critical part of the resources available which comprise the realm of pedagogical possibility in teaching mathematics" (p.36). Putnam and colleagues (1992) studied the beliefs and mathematics instructional practices of four fifth grade teachers and reported a similar finding: they describe the cases of two teachers who, despite their inquiry-oriented beliefs about teaching mathematics, were limited by their MKT. In these situations, it appears that a sufficient level of MKT might be necessary for the relationship between mathematics teachers' beliefs about mathematics or beliefs teaching and learning mathematics and the enactment of CDTs to hold.

Second, the relationship between mathematics teachers' beliefs about teaching and learning mathematics and the enactment of CDTs might be contingent on curriculum supports. In particular, based on Boston and Smith's (2009) and Stein and Kaufman's (2010) findings, it seems that the nature of curriculum materials (e.g., whether they are inquiry-oriented or traditional) might be less consequential, and that it is really only the ways in which teachers use those curriculum materials (guided by the supports within the curriculum materials) that has an impact on the enactment of CDTs. Specifically, curriculum materials that support teachers to focus on the big ideas in tasks or lessons might increase the likelihood that teachers will enact CDTs (Stein & Kaufman, 2010). Therefore, it is likely that the relationship between teachers' mathematics-related beliefs and enactment of CDTs is contingent on the use of supportive curriculum materials. While this contingency on the relationship between teachers' beliefs and the enactment of CDTs has not been investigated, there is one study that provides initial support for this conjecture. Remillard (1999) found that when the curriculum did not address the less visible aspects of teaching (e.g., deciding what tasks to pose to students), teachers were left to interpret how that work should be carried out. In other words, without explicit support from the curriculum, beliefs about teaching and learning were increasingly related to teachers' instructional practice. The following relationship therefore might hold for enactment of CDTs: if the curriculum materials are not supportive of the enactment of CDTs then the relationship between mathematics-related beliefs and teachers' enactment of CDTs is likely to be weaker.

Lastly, one critical contingency associated with the relationship between beliefs about students' mathematical capabilities and enactment of CDTs is student characteristics. In particular, unproductive beliefs about specific students' mathematical capabilities only apply in cases when those specific students are present in the classroom. Although somewhat self-evident, it is important to account for this contingency in future large-scale studies.

Five studies have identified how the relationship between teachers' beliefs about students and teachers' instructional practice is contingent on characteristics of the students in the classroom. First, three studies suggest that teachers' beliefs vary for different groups of students (Anderson et al., 2005; Beswick, 2005; Cross, 2009). Next, Sztajn (2003) conducted case studies of two elementary teachers' implementation of reform recommendations. She found that both teachers conceptualized students' capabilities in terms of a deficit model, attributing a lack of success for students from certain socioeconomic groups to cultural or community characteristics. Because the two teachers taught students with different socioeconomic backgrounds, their beliefs influenced their teaching in different ways. Both teachers believed that students need basic skills in order to succeed, but because one teacher taught students who came from higher socioeconomic backgrounds and believed that those students could gain some of those skills at home, she did not feel the need to place such an emphasis on basic skills in the classroom, and therefore, had time for more mathematically rich activities. This was not the case for the other teacher. Although the two teachers' perceptions of students' needs were similar, the teachers' practices differed due to the relevance of those perceptions with the students they were teaching. Turner and colleagues (2011) similarly suggest that teachers' beliefs about students capabilities may have more of an impact on their practice in schools with low-achieving (and often high-minority, low-SES) populations. Therefore, there is some indication that the relationship between beliefs about students' mathematical capabilities and the enactment of CDTs may be contingent on the students in the classroom.

In sum, I have suggested several different contingencies associated with the relationships between mathematics teachers' knowledge and beliefs and the enactment of CDTs. This same approach of building on contingencies identified with small-scale studies and attending to the situated nature of these relationships should be followed for the other factors when conducting large-scale investigations.

Discussion

In this paper, I have attempted to synthesize the findings of existing research related to the enactment of CDTs in order to suggest directions for future research. I have proposed a set of potentially productive teacher and contextual factors to study and have suggested an approach for studying those factors on a large scale.

I have defined the enactment of CDTs as involving: 1) *selecting* such tasks; and 2) *maintaining the cognitive demand* of those tasks during classroom implementation. There is considerable evidence that enacting CDTs is very challenging work. Hence, teachers need significant support to do so. In order to understand the appropriate supports for teachers, we need to understand how particular teacher and contextual factors influence the enactment of CDTs.

All five of the studies that examined the enactment of CDTs from this perspective were conducted recently (Boston & Smith, 2009; Charalambous, 2010; Choppin, 2011; Son, 2008; Stein & Kaufman, 2010). The methodological approaches and findings from the studies are mixed and inconclusive. For example, Charalambous (2010) examined the instructional practices of two elementary teachers and described how the enactment of CDTs was related to the teachers' mathematical knowledge for teaching, but Stein and Kaufman (2010) studied a larger sample of 48 elementary teachers and found that teachers' MKT was not related to their enactment of CDTs. Similarly, results were mixed for the influence of curricular materials. I argue that these mixed findings are likely related to differences in measurement of the constructs of interest and a lack of attention to the contingencies associated with the relationships. Several factors have only been investigated in one study at this point so there has not been an opportunity for vetting the findings, but they appear to be promising factors to continue to explore. Those factors include: knowledge of students' thinking (Choppin, 2011), conceptions of knowing and doing mathematics (i.e., beliefs about teaching and learning mathematics), and teacher perceptions about student achievement (i.e., beliefs about students' mathematical capabilities) (Son, 2008).

Given the importance of supporting the enactment of CDTs at scale, we need to continue to study the influence of teacher and contextual factors on the enactment of CDTs. As an attempt to build on these five existing studies I have focused on two major goals: 1) identifying potentially important teacher and contextual factors that can be related to processes of enacting CDTs, and 2) describing an effective approach for studying these factors on a large scale.

In service of the first goal, I conducted a broad literature review of factors related to the nature of students' mathematical activity which identified a set of promising factors to investigate. I drew on evidence from the broad literature review and on an analysis of the processes associated with enacting CDTs to justify each potential factor's inclusion in the list of promising factors. Promising teacher factors included: (a) mathematical knowledge for teaching (including content knowledge and pedagogical content knowledge), (b) knowledge of students' mathematical thinking, (c) beliefs about mathematics, (d) beliefs about teaching and learning mathematics, (e) beliefs about students' capabilities, and (f) classroom management skills. Promising contextual factors included: (a) class time, (b) class size, (c) characteristics of students, (d) departmental culture, (e) school leaders' expectations for instructional improvement, (f) nature of the curricular materials, and (g) learning opportunities through interactions. The majority of the studies in the broad review focused on teacher factors and less attention has been given to contextual factors. More work is therefore needed on how contextual factors influence teachers' instructional practice- in particular, the enactment of CDTs. Given that teachers' work is situated in schools and districts and that there is considerable variation in those contexts, it is unreasonable to expect that we can understand the work of enacting CDTs at scale without considering the school and district contexts in which teachers work.

With regard to the second goal of understanding an effective approach for studying the promising factors on a large scale, I drew on results from small-scale studies to identify possible contingencies to investigate in large-scale studies. In doing so, I identified several potentially important contingencies associated with the relationship between teachers' knowledge and the enactment of CDTs: (a) beliefs about teaching and learning mathematics, (b) beliefs about students' mathematical capabilities, and (c) curriculum supports. Similarly, potentially important contingencies associated with the relationships between mathematics teachers' beliefs and the enactment of CDTs include: (a) teachers' knowledge, (b) curriculum supports, and (c) student characteristics. Recall that these hypothesized contingencies are based on small-scale studies and that relatively few small-scale studies have investigated relationships between contextual factors and teachers' instructional practice. Therefore, there is a pressing need for small-scale studies that investigate how contextual factors influence teachers' instructional practices.

Given the set of promising teacher and contextual factors along with potential contingencies between some of those factors and the enactment of CDTs, the next steps are to build on existing work to develop reliable measures of these different factors and to design large-scale studies to explore the important relationships. Once we have an understanding of how different factors influence the enactment of CDTs, we will be able to design and adjust supports for teachers as they work to enact CDTs, and thus promote efforts for all students to learn mathematics.

CHAPTER III

MIDDLE SCHOOL MATHEMATICS TEACHERS' ENACTMENT OF COGNITIVELY DEMANDING TASKS: INVESTIGATING LINKS TO TEACHERS' MATHEMATICAL KNOWLEDGE FOR TEACHING AND BELIEFS ABOUT TEACHING AND LEARNING MATHEMATICS

Introduction

New reform goals and standards for students' mathematical learning have been put in place over the past two decades (e.g., see National Council of Teachers of Mathematics [NCTM], 1989, 2000). These goals for students' mathematical learning also imply new expectations for mathematics teachers' work in their classrooms. The *Curriculum and Evaluation Standards* and *Principles and Standards for School Mathematics* documents published by the National Council of Teachers of Mathematics (1989, 2000) reflect a consensus within the mathematics education research community for comprehensive reforms to traditional mathematics instruction. Two fundamental aspects of high quality mathematics instruction proposed in these documents are the use of challenging mathematical tasks and discussions of students' solutions to such tasks that focus on key mathematical ideas. These aspects have clear implications for the role of the teacher (Hiebert et al., 1997). For example, the teacher is expected to choose and set up the challenging tasks for students and to orchestrate productive discourse within the classroom (Stein et al., 2008).

The level of challenge of the tasks students solve and discuss is the foundation for students' learning opportunities. For example, it would require considerable teacher expertise to provide conceptual learning opportunities for students based on a task that requires students to reproduce memorized facts. On the other hand, cognitively demanding, or challenging, tasks are much more likely to set the stage for conceptual conversations about mathematics. The use of cognitively demanding tasks (CDTs) is a critical aspect of high-quality mathematics instruction that requires further investigation.

Developing instruction centered around CDTs requires considerable learning on the part of most U.S. teachers, requiring changes in knowledge and beliefs along with related changes in practice (Ball & Cohen, 1999; Kazemi & Franke, 2004; Stein et al., 1999; C. L. Thompson & Zeuli, 1999). More generally, in order to improve mathematics teachers' instructional practices, we need to understand more about what influences different aspects of teachers' instructional practice.

One way to better understand teachers' instructional practice is to examine actions that occur in the classroom and try to discern reasons for those actions. For example, Stein, Grover, and Henningsen (1996) found that teachers changed the nature of tasks by stressing more or less challenging aspects of the tasks. While this information helps explicate how teachers' actions change the cognitive demand of tasks, it is not clear *why* teachers might stress different aspects of the tasks or engage in other actions. Further, knowing how particular actions influence the cognitive demand of tasks in the classroom allows us to recommend or discourage particular teacher actions, but it does not allow us to attend to underlying supports for the development of those actions. In order to understand how to support teachers to enact instruction centered on CDTs, we need to study what influences such actions and decisions. In particular, we need to look beyond actions that occur in the classroom to teacher and school context factors (e.g., teachers' knowledge and beliefs, interactions with colleagues, principal expectations, and formal professional development) to examine what influences teachers' practice. In this study, I investigate the following research questions:

1) How are mathematics teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics related to the cognitive demand of the tasks they select?
2) How are mathematics teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics related to maintenance of the cognitive demand of high-level tasks?

Conceptual Framework

The cognitive demand of a task refers to "the cognitive processes students are required to use in accomplishing it" (Doyle, 1988, p. 170). Stein, Grover, and Henningsen (1996) classified mathematical tasks into those with low and high cognitive demand. Mathematical tasks with low cognitive demand require students to memorize or reproduce facts, or perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks with high cognitive demand (or CDTs) require students to make connections to the underlying mathematical ideas. In addition, students are asked to engage in disciplinary activities of explanation, justification, and generalization, or to use procedures to solve tasks that are open with regard to which procedures to use. While implied in the definition, it is important to emphasize that the distinctions between high and low cognitive demand are relative to students' current understanding and, thus, are situation-dependent.

There is evidence that CDTs can provide critical learning opportunities for all students. Stein and Lane (1996) found that the use of tasks with high cognitive demand was related to greater student gains on an assessment requiring high levels of mathematical thinking and reasoning. In particular, the greatest gains occurred when teachers assigned tasks that were initially of high cognitive demand, and teachers and students maintained the cognitive demand throughout the lesson, rather than engaging the students in a procedural activity. Further, there is evidence that CDTs afford valuable learning opportunities for all students, not just previously high-achieving students (Zohar & Dori, 2003). The enactment of CDTs in the classroom therefore appears to be important in supporting all students' learning.

There is also evidence that in classrooms in the U.S., CDTs are not often posed, and when they are posed the cognitive demand of the tasks are not maintained. In attempting to understand more about changes in cognitive demand during a lesson, Stein, Grover, and Henningsen (1996) documented the initial cognitive demand of mathematical tasks as written or verbally posed to students and examined whether teachers and students maintained, increased, or decreased the demand in different phases of a math lesson. They found that in classrooms where tasks with the potential for high levels of cognitive demand were used, teachers and/or students often decreased the cognitive demand during implementation of the tasks. The results from the 1999 TIMSS video study are consistent with those of Stein and colleagues in that they suggest that the mathematical activity in U.S. middle school classrooms tends to be procedural in nature, and when teachers do select high-level tasks they often implement them in low-level ways (Hiebert et al., 2003; Hiebert et al., 2005).

The Math Task Framework proposed by Stein, Grover, and Henningsen (1996) is useful when analyzing how teachers enact tasks. The Framework divides a lesson up into phases and transitions between phases of the lesson (see Figure 1). The squares denote different phases of the lesson and flow from left to right. In the subsequent paragraphs, I describe how my definition of the enactment of CDTs maps onto the Math Task Framework. In this analysis, I focus on the cognitive demand of the written task as selected from the curricular materials and then the changes in the cognitive demand from how it is written to the implementation by the teacher and students (denoted in the second and third squares). Both of these aspects of cognitive demand influence whether students engage in cognitively demanding mathematical activity in the classroom.

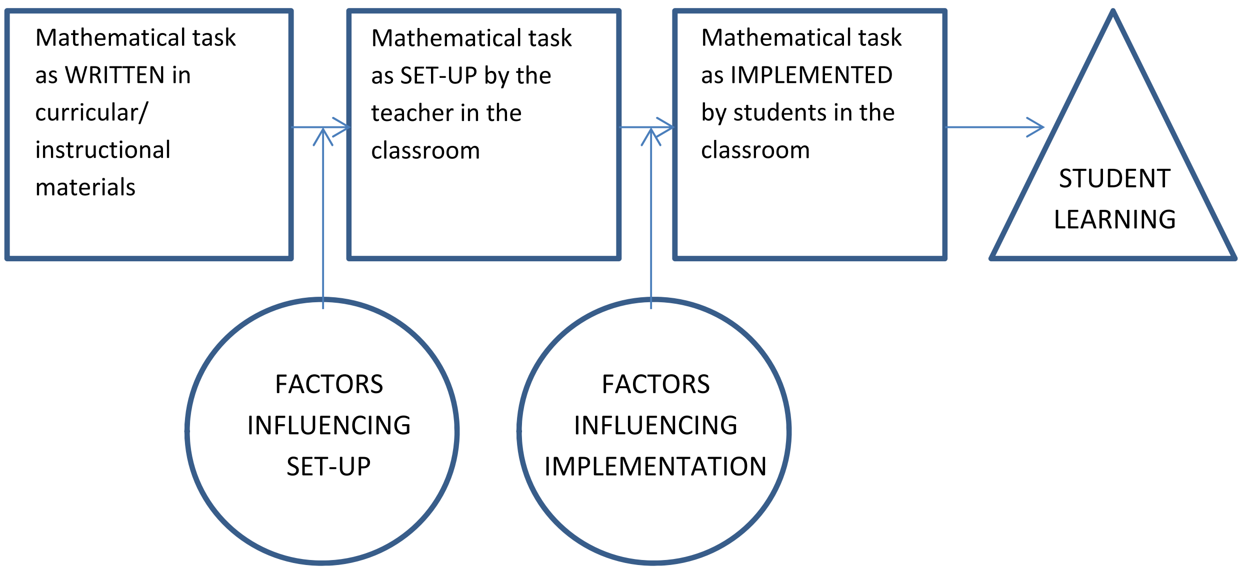


Figure 1. Modified "Math Tasks Framework" (Stein et al., 1996, p.459)

First, teachers select a task from the curriculum materials (in Figure 1, the leftmost square). In selecting a task to pose to students, teachers might choose to use a task directly from the curriculum as suggested by the district pacing guide, use other tasks from the curriculum, adjust tasks from the curriculum, use tasks from supplementary materials, or create their own tasks. In doing any one of these options, the teacher may have to balance a range of issues including: teaching goals, class time, knowledge of the students, what he or she feels capable of, and others' expectations. For example, a teacher who only has 45 minutes for each class period might decide that she does not have enough class time to engage students in a high cognitive demand task. It is unlikely that teachers consider each of these concerns individually, but instead they weigh them against each other when deciding which to prioritize and how they fit with their goals for the class period (Remillard, 1999). The cognitive demand of the selected task sets the stage for the cognitive demand over the course of the lesson.

Once a CDT has been selected, maintenance of the cognitive demand concerns both the task set-up and implementation (in Figure 1, the second and third squares). In the first of those phases, the task is set up for the students in the classroom (in Figure 1, the second square from the left). In other words, the teacher explains what students are expected to do to complete the task. During this phase a teacher can alter the cognitive demand of the task by clarifying or changing the expectations set out in the written task. For example, a teacher might tell students to complete only part of the written task or might go through a series of examples that change the nature of the task in which students will engage.

The next phase (in Figure 1, the third square from the left) is the implementation of the task by the students in the classroom. This phase includes all of the remaining class time spent on the task. For example, it might include both student work time and a concluding whole-class discussion. The cognitive demand of the task can also change within this phase of instruction, depending on teacher and student actions as they carry out the task. There are several reasons why cognitive demand might decrease: 1) the teachers' expectations for students' work might be unclear, 2) the classroom environment might not be conducive to engaging in challenging mathematical activity (e.g., poor classroom management, unproductive classroom norms), 3) the task might not be appropriate for students given their current mathematical knowledge, or 4) the appropriate level of scaffolding or teacher support for students to productively engage in high-level ways might not occur (Henningsen & Stein, 1997; Stein et al., 1996). Given the range of the reasons why cognitive demand might decrease, it is clear that maintaining the cognitive demand of CDTs is quite demanding.

In the following section, I describe the existing research on how teachers' knowledge and beliefs are related to their enactment of CDTs and describe the particular aspects of teachers' knowledge and beliefs that I investigate in this study.

Mathematical Knowledge for Teaching and Beliefs about Teaching and Learning Mathematics

The question of which aspects of teachers' knowledge and beliefs are related to teachers' enactment of CDTs has just begun to be investigated. In recent years, two studies have investigated how mathematics teachers' knowledge is related to their enactment of CDTs, with conflicting results. Both of these studies examined one aspect of mathematics teachers' knowledge: mathematical knowledge for teaching. Within the mathematics education community, mathematical knowledge for teaching is generally conceptualized as a combination of subject matter knowledge and pedagogical content knowledge (PCK) (Hill, Ball, et al., 2008). In the view of Hill and colleagues, subject matter knowledge is conceptual knowledge of mathematics that is necessary for solving mathematics problems but that is not specific to the work of teaching. On the other hand, PCK is specific to the work of teaching. Shulman (1986, 1987) introduced the notion of PCK and suggested that it "represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction" (1987, p. 228). Hence, mathematical knowledge for teaching (MKT) goes beyond pure mathematical content knowledge to also include PCK for mathematics teachers.

Charalambous (2010) examined the relationship between teachers' mathematical knowledge for teaching and changes in the cognitive demand over the course of a lesson for two elementary teachers who differed dramatically in their mathematical knowledge for teaching. He found that mathematical knowledge for teaching was related to maintenance of the cognitive demand of CDTs and as a result of his case analyses, he proposed several hypotheses about how high levels of mathematical knowledge for teaching are integral to the work of teaching. For example, his first hypothesis was that "strong mathematical knowledge for teaching supports teachers in using representations to attach meaning to mathematical procedures rather than to simply show answers" (Charalambous, 2010, p. 273). These hypotheses suggest mechanisms by which mathematical knowledge for teaching is integral to teacher practices.

Instead of focusing only on mathematical knowledge for teaching, Stein and Kaufman (2010) focused on curricular materials and teacher capacity, conceptualized as teacher education, experience, professional development, and mathematical knowledge for teaching, as they related to inquiry-oriented elementary mathematics curriculum implementation (with one of three foci being maintaining high levels of cognitive demand). To investigate the enactment of CDTs, they created a total cognitive demand score that was the sum of the cognitive demand of the task as selected and as implemented. In particular, if high-level tasks were selected and the cognitive demand was maintained, then the total cognitive demand score would be at its maximum value. Although they expected that teacher knowledge, education, and experience would be related to curriculum implementation, they did not find any significant relationships between those teacher factors and instruction. In fact, they found that mathematical knowledge for teaching was not significantly related to the cognitive demand of the tasks in the classroom, directly contradicting Charalambous' findings. This suggests that the relationship between mathematical knowledge for teaching and the cognitive demand of mathematical tasks might not be straightforward, and should be explored further.

According to A. G. Thompson (1992) teachers' beliefs about mathematics teaching include: "what a teacher considers to be desirable goals of the mathematics program, his or her own role in teaching, the students' role, appropriate classroom activities, desirable instructional approaches and emphases, legitimate mathematical procedures, and acceptable outcomes of instruction" (p.135). Only one study has examined how teachers' beliefs about teaching and learning mathematics are related to their enactment of CDTs. Son (2008) found that several aspects of teachers' beliefs about teaching and learning mathematics were related to teachers' enactment of CDTs.

Son (2008) examined elementary teachers' mathematics textbook use with an emphasis on the patterns of cognitive demand in their classrooms using survey data from 169 teachers, with observational data and artifacts from eight of those teachers. In the observational analysis Son focused on three different patterns of cognitive demand in the problems used from the textbook and the types of questions asked of the students over the course of the lesson: high-level problems to high-level questions, high-level problems to low-level questions, and low-level problems to low-level questions. This approach is similar to examining the changes in cognitive demand from the task as selected to the task as implemented. She found that teachers' beliefs about teaching and learning mathematics were related to the cognitive demand patterns. Aspects of beliefs about teaching and learning mathematics that she named included conceptions of knowing and doing mathematics, goals for student learning, and perceptions about student achievement (described as "students' diversity in terms of mathematics ability" (Son, 2008, p.219)). Another way to conceptualize perceptions about student achievement is as beliefs about students’ mathematical capabilities (e.g., Stipek et al., 2001; Stodolsky & Grossman, 2000; A. G. Thompson, 1984). Son’s results give some indication that teachers’ beliefs about teaching and learning mathematics, and, specifically, beliefs about students’ mathematical capabilities are related to their enactment of CDTs, yet they are limited to elementary teachers’ described practices with regard to the enactment of CDTs, which she found to be inaccurate representations of teachers’ actual classroom practice. This leaves open the question of whether teachers’ beliefs about teaching and learning mathematics are related to their observed enactment of CDTs.

One dimension of beliefs about teaching and learning mathematics that has not yet been investigated but might be integral to teachers’ enactment of CDTs is teachers’ beliefs about supporting struggling students (which is related to teachers’ beliefs about students’ mathematical capabilities). For example, some teachers believe that CDTs should be simplified for struggling students, while others believe that teachers should use CDTs with multiple entry-points to support struggling students. In fact, there is evidence that many teachers believe that high cognitive demand tasks are not appropriate for currently low-achieving students (Zohar, Degani, & Vaaknin, 2001). If we are interested in enabling all students to engage in cognitively demanding mathematical activity, then a particular set of beliefs about supporting students are more productive: the belief that struggling students should be supported to participate in rigorous mathematical activity. Hence, teachers’ beliefs about supports for struggling students is one aspect of beliefs about teaching and learning mathematics that might be integral to enacting CDTs, especially in large, urban districts where many of the students have been identified as “struggling.”

Contingencies Associated with Knowledge and Beliefs and the Enactment of CDTs

The relationships between knowledge and beliefs and the enactment of CDTs are often more complex than the relationships assumed in simple regression analysis. In other words, the relationships between teachers' knowledge and beliefs and the enactment of such tasks might have some associated contingencies. Further, the presence of unexamined contingencies could account for the mixed results in recent studies of how mathematical knowledge for teaching is related to the enactment of CDTs. As a part of this analysis, I draw on evidence from several small-scale studies suggesting one key set of contingencies: an interaction between mathematical knowledge for teaching and beliefs about teaching and learning math. Interactions between mathematical knowledge for teaching and beliefs about teaching and learning math can be interpreted in two different ways: (a) the relationship between mathematical knowledge for teaching and the enactment of CDTs is contingent on beliefs about teaching and learning mathematics, or (b) the relationship between beliefs about teaching and learning math and the enactment of CDTs is contingent on a teachers' mathematical knowledge for teaching. Evidence from small-scale studies supports both of these interpretations.

On the one hand, there is evidence that the relationship between mathematics teacher knowledge and the enactment of CDTs is contingent on teachers' beliefs about teaching and learning mathematics (Ball, 1991; Schoenfeld, 2011; Turner et al., 2011). The studies indicating this contingency suggest that an inquiry-orientation in teachers' beliefs about teaching and learning mathematics is necessary for teachers who have developed sufficient mathematical knowledge for teaching to support conceptually-rich mathematical activity in the classroom. In particular, several studies suggest that beliefs about teaching and learning mathematics can limit teachers' ability to enact mathematical tasks in conceptual ways. Schoenfeld (2011) demonstrated that for one teacher who seemed to have developed relatively sophisticated mathematical knowledge, his belief that "what he said should be an elaboration or clarification of what a student had said" (p.82) limited the nature of the mathematical activity in the classroom. Ball (1991) described a similar phenomenon and argued that teachers' beliefs about teaching and learning mathematics might interfere with whether or not they draw on their mathematical knowledge.

There are also indications that the relationship between mathematics teachers' beliefs about teaching and learning mathematics and the enactment of CDTs might be contingent on teachers' knowledge. Putnam et al. (1992) studied the beliefs and mathematics instructional practices of four fifth grade teachers and reported such a contingency: They describe cases of two teachers who, despite their inquiry-oriented beliefs about teaching mathematics, were limited by their mathematical knowledge for teaching. In these situations, it appears that a basic level of mathematical knowledge for teaching is necessary for the relationship between mathematics teachers' beliefs and the enactment of CDTs to hold. In sum, there is evidence suggesting complex interrelationships between mathematical knowledge for teaching, beliefs about teaching and learning mathematics, and the enactment of CDTs, with several different interpretations. I modeled these complexities as I investigated the relationships between teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics and their enactment of CDTs.

Method

In this study, I investigated the relationships between aspects of teachers' knowledge and beliefs and their enactment of CDTs. In doing so, I considered both the cognitive demand of the tasks selected and the extent to which the cognitive demand of high-level tasks is maintained, separately, and I also explored potential contingencies in the relationships.

Sample

I drew on data collected in the course of a four-year study that sought to address the question of what is needed to improve the quality of middle-grades mathematics teaching, and thus student achievement, at the scale of a large urban district (Cobb & Jackson, 2011; Cobb & Smith, 2008). The research team collaborated with the leaders of four large, urban districts that were attempting to achieve a vision of high-quality mathematics instruction that was compatible with the National Council of Teachers of Mathematics' (2000) *Principles and Standards for School Mathematics*. In each of the four districts, the research team selected a sample of 6 to 10 middle-grades schools that reflected variation in student performance and in capacity for improvement in the quality of instruction across the district. Within each school, up to five mathematics teachers were randomly selected to participate in the study, for a total of approximately 30 teachers per district. Although we tried to retain as many teachers as possible throughout the study, the sample varies from year to year as we recruited replacements when teachers left schools or changed teaching assignments.

The four collaborating school districts were typical of large, urban districts in that they had limited resources, large numbers of traditionally low-performing students in mathematics, high teacher turnover, and disparities among subgroups of students in their performance on state standardized tests (Darling-Hammond, 2007). The districts were atypical in their response to high-stakes accountability pressures: they responded by focusing on improving the quality of instruction rather than focusing exclusively on student test scores. Consistent with this response, three of the four districts (which I will call Districts A, B, and D) adopted the Connected Mathematics Project 2 (CMP2) as their primary curriculum. District C adopted a more traditional curriculum but encouraged teachers to supplement with CMP2 and another internally-developed inquiry-oriented curriculum. District B, C, and D began implementation of their respective curricula in Year 1 of the study. In contrast, District A began implementation of CMP2 in Year 2 of the study, but their district had a 10-year history of using the original Connected Mathematics Project Curriculum, prior to the adoption of CMP2. Additionally, each district attempted to implement a number of strategies to support teachers in improving their instruction (e.g., curriculum frameworks, coaching, regularly scheduled time to collaborate with colleagues on issues of instruction, or professional development for instructional leaders).

In each of the four years of the study (2007-2011), we collected several types of data to test and refine a set of hypotheses and conjectures about district and school organizational arrangements, social relations, and material resources that might support mathematics teachers' development of high-quality instructional practices at scale. The primary data sources used for my analysis were video-recordings of teachers' classroom instruction, an assessment of teachers' mathematical knowledge for teaching, and interviews with teachers. My primary analytic sample was 214 middle-school mathematics teachers pooled over the four years of the study, with 406 lessons in total (24 teachers with 4 years of data, 40 teachers with 3 years, 41 teachers with 2 years, and 108 teachers with 1 year of data).

Focal Measures

I first describe the primary outcome measures of teachers' enactment of CDTs and then describe the measures that I use to assess teachers' knowledge and beliefs. Lastly, I describe the measures I used to control for teachers' experience of teaching and with the curriculum, the prior mathematics achievement of their students, the number of students in their class, and the duration of their class period.

Enactment of CDTs. I constructed measures of teachers' enactment of CDTs by drawing on the measures of the quality of teachers' instructional practice used in the larger research project: the Instructional Quality Assessment (IQA, Boston & Wolf, 2006; Matsumura et al., 2006). We used this instrument to code video-recordings of the participating teachers' classroom instruction. In each of the four years of the study, we video-recorded two (ideally consecutive) mathematics lessons conducted by each of the 120 teachers in the study in late winter. Teachers were asked to engage students in a problem-solving lesson with a related whole-class discussion.

The IQA was developed by a team of researchers at the University of Pittsburgh, and the larger study used eight of their developed rubrics to assess the quality of teachers' instruction. I focus on two of those rubrics: Task Potential and Task Implementation. The Task Potential and Implementation rubrics were based on the earlier work by Stein and colleagues (e.g., see Stein et al., 1996; Stein & Lane, 1996), described above. These rubrics were designed to measure the cognitive demand of the task posed to students (Task Potential) and the cognitive demand of the task as implemented by the students and the teacher in the classroom (Implementation). Both rubrics use the same scale with five levels of cognitive demand. A task is coded as 0 if it is not mathematical in nature. Tasks coded as levels 1 and 2 are low in cognitive demand, with a level 1 task requiring only memorization or the reproduction of facts and a level 2 task requiring students to perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks coded at levels 3 and 4 represent tasks of high cognitive demand. A level 3 task requires students to make connections to underlying mathematical ideas, but tasks do not include explicit requests for generalization or justification. At the highest level, a level 4 task asks students to engage in the disciplinary activities of explanation, justification and generalization, or to use procedures to solve tasks that are somewhat open-ended in nature. One critical distinction is between low-level tasks (level 2 or less) and high-level tasks (level 3 or 4).

Recall that my definition of teachers' enactment of CDTs includes: 1) the *choice* of such tasks; and 2) the *maintenance of the cognitive demand* of those tasks during classroom implementation. The cognitive demand of the selected task is measured by the Task Potential rubric. The maintenance of the cognitive demand is derived by comparing the Task Potential and Implementation scores; maintenance of the cognitive demand of a task is a measure of whether the score for Implementation is at least as high as the score for the Task Potential. Maintenance is generally a measure of whether the score for Task Potential is equal to the score for Implementation because in this sample tasks rarely[2] increase from selection to implementation.

Table 1

Reliability Information for Task Potential and Implementation

	Y1		Y2		Y3		Y4	
	Potential	Impl	Potential	Impl	Potential	Impl	Potential	Impl
% Agreement	59.4	78.1	56.9	78.5	75	89.3	59.1	63.6
kappa	0.37	0.51	0.29	0.37	0.63	0.75	0.36	0.29

Using the IQA requires experience with mathematics and teaching. Therefore, each year we recruited doctoral students in education and master's students in mathematics education, to serve as IQA coders. An IQA developer trained coders in each year of the study. Coders were required to achieve 80 percent agreement with previously consensus-coded videos during the training reliability phase and inter-rater agreement was assessed every other week over the course of the 10 weeks of coding (resulting in double-coding of about 15% of the video sample).

[2] 16 times in the sample, with the majority in district A

Table 1 gives the reliability information for each rubric in each study year because coding the video-recordings from each study year were coded each summer, which produced four sets of reliability information. Ongoing reliability was calculated with percent exact agreement and kappa scores. A kappa score is a measure of reliability based on percent exact agreement that is adjusted for the chance agreement based on the actual distribution of the data (J. Cohen, 1960). The exact agreement percentages for the Task Potential rubric were between 56.9% and 75% and kappa scores were between 0.29 and 0.63. Task Implementation reliability was slightly higher with exact agreement ranging from 63.6% to 89.3% and kappa scores between 0.29 and 0.75. Hartmann, Barrios, and Wood (2004) suggest that appropriate agreement rates are between 80 and 90 percent, but that for more complex instruments 70% could be sufficient. The kappa scores are at worst “Fair” agreement and at best “Substantial” agreement (Landis & Koch, 1977), There is some evidence that percent agreement is not the best measure of inter-rater reliability, and that kappa scores are more accurate in measuring inter-rater agreement. Unfortunately, there is evidence that kappa scores are often negatively skewed when the actual scores are not well distributed (Gwet, 2010). Therefore, given the complex nature of this instrument and the imperfection in the measures of inter-rater agreement, a case can be made that these inter-rater reliability scores are sufficient.

Descriptive statistics for Task Potential and Task Implementation, by District, are given in Table 2. While scores for Task Potential had the potential to range from 0 to 4, only 1 of observation received a Task Potential score less than 2. Therefore, for this set of analyses, I considered all tasks with Task Potential less than or equal to 2 as tasks with “Low Task Potential.” Therefore, there were 3 primary categories of interest for Task Potential: Low, 3, and 4.

Table 2

Descriptive Statistics of Task Potential and Implementation, By District

	District A		District B		District C		District D	
	Potential	Impl	Potential	Impl	Potential	Impl	Potential	Impl
Mean	3.08	2.64	3.03	2.32	2.51	2.17	3.03	2.38
Score=1	0	1	1	2	0	2	0	2
Score=2	15	36	28	81	49	80	31	66
Score=3	47	38	60	39	43	14	41	33
Score=4	22	9	34	1	3	1	34	5
N	84		123		95		106	

Mathematical knowledge for teaching. In March of each of the four years of the larger study, we assessed all participating teachers' mathematical knowledge for teaching (MKT) by using a pencil-and-paper instrument developed by the Learning Mathematics for Teaching project at the University of Michigan (Hill et al., 2004). The instrument has a reliability index of .70 or above and can be used to assess teachers' knowledge with respect to two dimensions: number concepts and operations (NCOP); and patterns, functions and algebra (PFA). For each of the two subtests (NCOP and PFA), raw scores were translated into IRT (item response theory) scale scores (provided by MKT developers), the determination of which was based on results from a pilot administration of the assessment to a national sample of approximately 640 practicing middle school teachers. To investigate how teachers' mathematical knowledge for teaching is related to their enactment of CDTs, I used a combined average of these two scale scores to form a single MKT score for each participant in each year. The use of IRT scores based on the national sample allows me to interpret the MKT scores of the teachers in our sample to the national average and distribution (i.e., a mean score of 0 and standard deviation of 1). Descriptive statistics for MKT are given in Table 3. In addition to the inclusion of the IRT score as a continuous factor, I examined differences by score quartile to try to understand differential effects for different categories of quartile membership.

Table 3

Descriptive Statistics of Independent Variables, By District

	Yrs Exp	New to CMP	# Stud	Class Time	Stud Prior M	Stud Prior SD	MKT	VHQMI	BSSS
Dist A N=82									N=47
Mean	15.4	0.037	20.1	65.4	-0.54	0.71	0.44	2.74	1.47
SD	8.58	0.19	6.06	22.3	0.65	0.2	0.96	0.52	0.75
Min	1	0	6	40	-1.82	0.21	-1.47	1	0
Max	40	1	36	120	1.66	1.57	2.01	3.83	2
Dist B N=124									N=54
Mean	7.63	0.30	18.2	85.0	-0.70	0.68	-0.19	2.34	0.93
SD	8.61	0.46	4.84	24.8	0.55	0.18	0.68	0.61	0.91
Min	1	0	7	39	-1.88	0.19	-1.85	0.5	0
Max	40	1	31	184	1.35	1.21	1.53	3.67	2
Dist C N=99									N=33
Mean	8.71	0	18.1	74.0	-0.44	0.75	-0.32	2.15	0.85
SD	7.79	0	5.19	33.9	0.52	0.20	0.7	0.71	0.91
Min	1	0	4	40	-1.65	0.27	-1.81	0	0
Max	37	0	29	200	1.22	1.09	1.40	3.4	2
Dist D N=109									N=44
Mean	7.83	0.42	21.5	65.8	-0.52	0.75	-0.15	2.16	0.89
SD	8.51	0.50	4.62	19.8	0.45	0.13	0.57	0.73	0.84
Min	1	0	7	43	-1.42	0.40	-1.83	0	0
Max	34	1	32	130	0.79	1.14	1.35	3.71	2

Vision of high quality mathematics instruction. Several of the measures pertaining to teachers' beliefs about teaching and learning mathematics that I utilized are derived from interviews conducted with the 120 participating teachers in January of each year of the larger study. In general, in the interviews we ask about the school and district settings in which teachers work as well as their vision of high quality math instruction and to what they attribute the lack of success of particular groups of students. To understand teachers' Visions of High Quality

Mathematics Instruction (VHQMI; Munter, Under review), teachers were asked what they would look for when observing another mathematics teacher's instruction to determine if the instruction was of high quality. Depending on the breadth of their responses, teachers were then asked a series of probes (see Appendix A for details).

Each year, teacher interviews were transcribed and then coded. Teachers' responses to the interview question were coded on several different dimensions: the role of the teacher, mathematical tasks, classroom activity, and discourse (including the structure, the nature of talk, teacher questions, student questions, and student explanations) (see Appendix A for details). For each rubric, scores range between 0 and 4. Teachers who describe more traditional instruction are at the bottom of the scale and the top of the scale is inquiry-oriented instruction that includes CDTs, rich whole-class discussions, and a proactive role of the teacher in guiding these activities. Coders were trained by the developer of the measure and expected to reach an 80% agreement level prior to beginning coding. Overall, the ongoing reliability percent exact agreement between coders was 80%.

To estimate teachers' VHQMI, I used standardized mean scores that are the mean across the scored dimensions (i.e., if only two dimensions received scores, then the mean would be calculated across those two dimensions). Descriptive statistics for VHQMI scores are provided in Table 3. In the models, the scores are standardized based the sample of 406 lessons for ease of interpretation of interaction effects. As with MKT, I examined quartile membership within the sample of teachers to investigate contingencies and differential effects at different levels of VHQMI.

Beliefs about supporting struggling students. The second measure derived from the interviews was developed to assess the extent to which a teacher believes that all students can be supported to participate in high cognitive demand activity. Teachers were asked how they adjust their instruction for different groups of students. In coding for this measure, coders examined each interview transcript to look for instances when teachers described what they view as appropriate supports for students who struggle with mathematics. Segments of talk were coded as Unproductive, Mixed, or Productive with regard to whether the teachers' descriptions of supports for struggling students are aimed at enabling them to participate in rigorous mathematical activity. Productive views are indicated by accounts of supports that allow all students to participate in rigorous mathematical activity. Unproductive views are indicated by accounts of supports that diminish the cognitive demand of the activity for struggling students. For the purposes of this analysis, I focus on three categories of Beliefs about Supporting Struggling Students (BSSS): Productive, Not-Productive (i.e., Mixed or Unproductive), and Un-coded. The Un-coded category is included because in this sample of teachers, there were 236 yearly interviews for which we were unable to code the productivity of teachers' beliefs about supporting struggling students. Our inability to code is likely a result of interviewing (e.g., a lack of probing for details or reasoning following a particular response) rather than any characteristics of the teachers, but we include a dummy variable to account for the un-coded interviews to test that hypothesis empirically. Coders were trained and required to achieve 80% exact agreement with previously coded transcripts before beginning coding. Ongoing reliability was also assessed and coders achieved 64% exact agreement with corresponding kappa score of 0.47. This kappa score falls into the range of "Moderate" agreement (Landis & Koch, 1977). Given the complexity of this instrument these reliability scores are acceptable. Descriptive statistics for beliefs about struggling students are given in Table 3.

Control Measures

I included a set of control variables to account for other factors that, according to previous research, might influence the enactment of CDTs. The measures corresponding to these control variables are described below.

Teacher experience. Several studies have noted the potential importance of teaching experience with regard to teachers' instructional practices (Charalambous, 2010; Escudero & Sánchez, 2007; Remillard & Bryans, 2004). Also, from the perspective of reform implementation, there is evidence that experience with a particular curriculum matters in that it takes time to implement a new program effectively (Fullan, 2000). To control for the possibility of teachers' experience contributing to relationships between teachers' knowledge and beliefs and their instructional practice, I included two measures of teachers' experience: their years of experience teaching mathematics (Yrs Exp) and whether or not they are new to Connected Mathematics Project 2 (New to CMP2). Recall that three of the districts (Districts A, B, and D) were using CMP2 at the start of this study- with two of them beginning their use in the first year of the study and District A having a history with the Connected Mathematics program that began long before the start of this study. District C officially adopted a more traditional curriculum but encouraged teachers to supplement with CMP2 and another internally-developed inquiry-oriented curriculum. Descriptive statistics for these variables are given in Table 3.

Students' prior achievement. To control for differences in the students' prior achievement, I included measures of students' mean prior achievement (STUD M), and the standard deviation of students' prior achievement (STUD SD), representing the degree to which the class is made up of "struggling" students and the heterogeneity of the prior achievement of student in the class, respectively. Descriptive statistics for this variable are provided in Table 3.

Class time. Several studies suggest that the amount of class time available to work on challenging tasks influences teachers' instructional practices (Barrett Paterson, 2009; Cooney, 1985; Raymond, 1997; Son, 2008). In particular, the teachers in Son's study of enacting CDTs who did not use CDTs in their classroom explained that they felt they had limited time to cover the content and that using CDTs would take too much time. Specific to the enactment of CDTs in the classroom, Henningsen and Stein (1997) reported that time allocated to work on the task (either too much or too little) was a common reason for decline in cognitive demand of high-level tasks. Although teachers generally see a lack of time as the primary constraint, too much time was also a reason for decline in cognitive demand in Henningsen and Stein's study. Because the enactment of CDTs may depend on the amount of class time (CLASS TIME), I controlled for it in the subsequent analyses. Descriptive statistics for this variable are given in Table 3 and a standardized version of this variable is used in the models.

Class size. Another structural factor that might be an important consideration in teachers' enactment of CDTs is the size of the class (e.g., the number of students in the class). Rousseau and Powell (2005) studied four secondary teachers' mathematics reform efforts and found that the teachers mentioned the size of their classes as a reason why enacting CDTs was difficult. Teachers who had large classes explained that they did not engage students in collaborative activity because that type of activity would be difficult to manage with a large class. Productive learning opportunities for students with challenging tasks often involve students working in groups (E. G. Cohen, 1994). The large number of students in the classroom might make group activities more difficult due to space constraints or the challenge of keeping all groups engaged. Although the size of the class is not prohibitive of enacting high cognitive demand tasks in the classroom, it is likely to make it more challenging. Therefore, I controlled for the number of students in the class (# STUD) in each model. Descriptive statistics for this variable are given in Table 3.

Hypothesized Contingencies

In addition to looking for relationships between teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics and their enactment of CDTs, I also investigated the possible contingencies in these relationships. Recall that the literature suggests that the relationship between VHQMI and the enactment of CDTs might depend on MKT, and that the relationship between MKT and the enactment of CDTs might depend on VHQMI. Statistically, there is one test for both of the contingencies: a statistical interaction between MKT and VHQMI. However, the qualitative evidence in the literature suggests that considering an interaction of continuous variables might not be the most precise way to represent or interpret the expected contingencies. Instead, grouping teachers by score categories improved precision and interpretation. More specifically, the hypothesis that teachers need to have achieved a particular level of sophistication in their VHQMI in order for MKT to be related to the enactment of CDTs, suggests a threshold effect rather than a continuous interaction. The same possible threshold effect applies for the influence of MKT on the relationship between VHQMI and the enactment of CDTs. As a consequence, beyond examining a continuous interaction, I also considered interactions between the continuous variables and quartile score category representations of the variables. Those continuous by categorical interactions attempt to model the following two additional specifications of the interactions between MKT and VHQMI:

Specification (a): The relationship between MKT and the enactment of CDTs is likely to vary for different levels of VHQMI. In particular, the positive relationship between MKT and the enactment of CDTs will not hold if teachers have not developed an inquiry-oriented VHQMI.

Specification (b): The relationship between VHQMI and the enactment of CDTs is likely to vary for different levels of MKT. In particular, the positive relationship between VHQMI and the enactment of CDTs will not hold if teachers MKT is undeveloped.

Analyses

I used multi-level logistic regression models to model two different aspects of the enactment of CDTs: (a) the cognitive demand of tasks posed to students (Task Potential), and (b) maintenance of the cognitive demand of tasks. Logistic regression models are appropriate because the scores for Task Potential and maintenance of the cognitive demand are ordinal (rather than continuous). Given the distribution of Task Potential scores and my primary interest in the enactment of high-level tasks, the 3 categories for Task Potential within the models will be Low, 3, and 4. Because there are 3 categories of interest, I use multinomial logistic regression models so that I can consider differences between categories: Low compared with 3 and 3 compared with 4. The use of multinomial logistic regression treats each comparison separately and does not assume any consistency in how the variables are related between the difference comparisons. In practice, for example, this means that knowledge and beliefs might be related to the choice between a low and a level 3 task differently from how they are related to the choice between a level 3 and a level 4 task. Multi-level models are used because of the structure of the data: there are multiple observations for some teachers and teachers are nested within schools.[3] To control for unmeasured differences between school districts in the sample, I included district-level dummy variables.

For the models of maintenance of the cognitive demand, the sample is limited to teachers who initially posed a high-level task (i.e., Task Potential of 3 or 4), and I consider whether the cognitive demand of task stayed at the same level (i.e. Task Implementation greater than or equal to Task Potential) or decreased (i.e., Task Implementation less than Task Potential). For these models, the sample is limited to 171 teachers and 286 total observations because the remaining observations in the primary sample involve tasks with low Task Potential.

For both Task Potential and Maintenance of the Cognitive Demand of High-Level Tasks, I examined several models. First, I considered the simplest model with mathematical knowledge for teaching, vision of high-quality mathematics instruction, and beliefs about supporting struggling students. The level 1 equations were of the following form:

$$
\begin{aligned}
&(TASK\ POTENTIAL\ or\ MAINTENANCE)_{ijk} \\
&\quad = \pi_{0jk} + \pi_{1jk}(New\ to\ CMP2)_{1ijk} + \pi_{2jk}(Yrs\ Exp)_{2ijk} + \pi_{3jk}(\#\ STUD)_{3ijk} \\
&\quad + \pi_{4jk}(CLASS\ TIME)_{4ijk} + \pi_{5jk}(STUD\ M)_{5ijk} + \pi_{6jk}(STUD\ SD)_{6ijk} \\
&\quad + \pi_{7jk}(MKT)_{7ijk} + \pi_{8jk}(VHQMI)_{8ijk} + \pi_{9jk}(BSSS)_{9ijk} + \varepsilon_{ijk}
\end{aligned}
$$

With level 2 equations of the form:

$$\pi_{pjk} = \beta_{p0k} + r_{pjk} \text{ for } p = 0 \text{ to } 9.$$

[3] I used the HLM software to estimate models of Task Potential and the GLLAMM software package (Rabe-Hesketh, Skrondal, & Pickles, 2005) in STATA to estimate models of Maintenance of the Cognitive Demand

And, level 3 equations of the following form:

$$\beta_{00k} = \gamma_{000} + \gamma_{001}(DIST\ A)_k + \ \gamma_{002}(DIST\ C)_k + \ \gamma_{003}(DIST\ D)_k + u_{00k}$$

$$\beta_{pjk} = \gamma_{pjk} + u_{pjk} \text{ for } p\text{= 1 to 9.}$$

Before examining the contingencies, I examined the possibility of differential effects in the relationships between mathematical knowledge for teaching and beliefs about teaching and learning mathematics and the enactment of CDTs. Lastly, I examined the hypothesized contingencies in the relationships, with several different model specifications. Those model specifications included: 1) a continuous interaction between MKT and VHQMI, 2) an interaction between a MKT and quartile membership categories for VHQMI, and 3) an interaction between VHQMI and quartile membership categories for MKT.

Results

Prior to modeling the enactment of CDTs, I examined correlations among the variables to be investigated to check for potential sources of multicollinearity and to document basic relationships between variables. Table 4 shows that only two pairs of variables were correlated above .40: Task Potential and Implementation (r=0.42, p<.05), and BSSS-P and BSSS-Un-coded (r=-0.54, p<.05). Neither of these relationships is surprising: Task Potential serves as the starting value for Implementation; and, teachers who espouse productive beliefs about supporting struggling students were able to be coded (so they are, by default, not un-coded). Also, several other variables were correlated at or above the 0.20 level: MKT and Implementation (r=0.22, p<.05), VHQMI and Implementation (r=0.21, p<.05), MKT and VHQMI (r=0.24, p<.05), VHQMI and being brand new to the curriculum (r=-0.22, p<.05), years of experience teaching math and being brand new to the curriculum (r=-0.22, p<.05), years of experience teaching mathematics and students' mean prior math achievement (r=0.25, p<.05), and the number of students in the class and students' mean prior math achievement (r=0.25, p<.05). The first few correlations suggest that MKT and VHQMI are positively related to each other and to implementation scores. The positive correlation between years of experience teaching mathematics and students' mean prior math achievement is consistent with the notion that teachers who are more experienced often teach the highest achieving students (Darling-Hammond, 2007). Also, the positive correlation between the number of students in the class and students' mean prior math achievement suggests that the classes with higher achieving students tend to be the ones with a larger number of students in them. Overall, there were some modest correlations, but none that suggest that multicollinearity would be a problem in modeling relationships between these factors and the enactment of CDTs.

With regard to the models of the enactment of CDTs, I first report the results of the analyses conducted for Task Potential and then Maintenance of the Cognitive Demand of High-Level Tasks (which I refer to as "Maintenance"). As described above, in all of the models I included the same set of control variables, including district dummy variables (with District B as the reference because of its typicality), years of experience teaching math, a dummy variable indicating whether it was a teachers' first year using CMP2, the number of students in the class, the class time, the mean prior mathematics achievement of the students in the class, and the heterogeneity of the prior mathematics achievement of the students in the class.

Table 4

Correlations between Variables

	Task Pot	Impl.	Yrs Exp	New to CMP	# Stud	Class Time	Stud Prior M	Stud Prior SD	MKT	VHQMI	BSSS-P	BSSS-Un-coded
Task Potential	1											
Implementation	0.42*	1										
Yrs Exp	-0.06	0.07	1									
New to CMP	0.06	-0.10*	-0.22*	1								
# Stud	0.09*	0.10*	0.04	0.02	1							
Class Time	0.04	-0.08	0.004	-0.08	-0.08	1						
Stud Prior M	-0.05	0.04	0.25*	-0.03	0.25*	0.03	1					
Stud Prior SD	0.04	-0.01	-0.07	0.01	0.05	-0.06	-0.06	1				
MKT	0.13*	0.22*	0.09*	-0.08	-0.01	-0.03	0.04	-0.04	1			
VHQMI	0.13*	0.21*	0.05	-0.22*	0.02	0.08	-0.06	-0.11*	0.24*	1		
BSSS-P	0.09*	0.18*	0.08	-0.06	-0.03	0.02	-0.05	-0.01	0.09*	0.11*	1	
BSSS-Un-coded	-0.04	-0.05	0.01	0.01	0.03	-0.002	0.06	0.02	-0.11*	-0.02	-0.54*	1

Table 5

Task Potential Models, MKT and VHQMI/VHQMI-S

	MKT, VHQMI, BSSS				MKT and VHQMI (Categories)			
	Low v 3		4 v 3		Low v 3		4 v 3	
	Coef (SE)	RRR	Coef (SE)	RRR	Coef (SE)	RRR	Coef (SE)	RRR
District A	-.46 (.45)	0.63	-.44 (.43)	0.64	-.50 (.45)	0.61	-.47 (.44)	0.63
District C	1.15** (.35)	3.16	-1.85** (.66)	0.16	1.08** (.36)	2.94	-1.90** (.66)	0.15
District D	.35 (.36)	1.42	.29 (.35)	1.34	.39 (.36)	1.48	.29 (.36)	1.34
Yrs Exp	.04** (.01)	1.04	.02 (.02)	1.02	.04** (.02)	1.04	.02 (.02)	1.02
New to CMP	.75** (.35)	2.12	.86** (.35)	2.36	.70** (.35)	2.01	.84** (.36)	2.32
# Students	-.001 (.02)	1.00	.03 (.03)	1.03	-.002 (.02)	1.00	.02 (.03)	1.02
Class Time	-0.05 (.12)	0.95	.21 (.15)	1.23	-.05 (.13)	0.95	.19 (.15)	1.21
Stud Prior M	.10 (.23)	1.11	.12 (.27)	1.13	.13 (.24)	1.14	.15 (.27)	1.16
Stud Prior SD	-.88 (.69)	0.41	.79 (.78)	2.20	-.90 (.70)	0.41	.92 (.79)	2.51
MKT	-.04 (.17)	0.96	.24 (.19)	1.27				
MKT (Q2)					-.41 (.33)	0.66	-.21 (.42)	0.81
MKT (Q3)					-.33 (.35)	0.72	.33 (.40)	1.39
MKT (Q4)					-.25 (.36)	0.78	.19 (.43)	1.21
VHQMI	.10 (.13)	1.11	.34** (.15)	1.40				
VHQMI (Q2)					.14 (.34)	1.15	.19 (.44)	1.21
VHQMI (Q3)					.06 (.36)	1.06	.42 (.44)	1.52
VHQMI (Q4)					.31 (.38)	1.36	1.13** (.44)	3.10
BSSS-P	-.47 (.40)	0.63	-0.002 (.41)	1.00	-.48 (.41)	0.62	-.07 (.42)	0.93
BSSS-Un-coded	.08 (.28)	1.08	.11 (.32)	1.12	-.08 (.29)	0.92	.05 (.33)	1.05
Constant	-.63 (.75)		-2.1** (.90)		-.40 (.86)		-2.62** (1.02)	
N	414				414			

Note. RRR= Relative Risk Ratios.

* p<.1. **p<.05

Task Potential

Tables 5, 6, and 7 contain results for the set of models of Task Potential, or teachers' choice of tasks. First, results from the model examining the relationship between MKT, VHQMI, BSSS, and Task Potential (see Table 5) suggest that when controlling for MKT, BSSS, and the other factors listed above, a teacher's VHQMI is significantly related to his or her choice of tasks between a level 3 task and a level 4 task ($B=.34$, $p<.05$), but not between low-level and level 3 tasks ($B=.10$, $p=.45$). In other words, the relative risk of selecting a level 4 task over a level 3 task for a teacher who had a VHQMI score that was one standard deviation higher than the mean, or more inquiry-oriented in their beliefs, is 1.40 times the relative risk for a teacher who had a VHQMI score that was equal to the sample mean. In this model, MKT was not significantly related to teachers' choice of tasks between a low-level task and a level 3 task or between tasks of levels 3 and 4. Also, productive or un-coded beliefs about supporting struggling students were not significantly related to teachers' choice of tasks. Another significant result is that the greater a teacher's years of experience teaching, the greater the relative risk that he or she would pose a low-level task over a level 3 task ($B=.04$, RRR= 1.04, $p<.05$). While this is a relatively small relative risk ratio, the influence is more substantial than it initially appears. For example, for every 5 additional years of experience, a teacher is 1.22 times more likely to have posed a low-level task over a level 3 task. Also, teachers who were new to the curriculum are over 2 times more likely to have posed a low-level task or level 4 task, rather than a level 3 task (RRR=2.12 and RRR=2.36, respectively, $p<.05$). One explanation for this result is that teachers who were new to the curriculum might have been more likely to use a task directly from the curricular materials without altering it dramatically, or might have chosen a task from a curricular resource with more procedural tasks because it was more aligned with their prior experience. Lastly, the results for district fixed effects suggest that differences in MKT, VHQMI, and BSSS do not account for all of the significant differences between teachers in Districts B, the reference district, and District C. Teachers from District C are over 3 times more likely to have posed a low-level task over a level 3 task and over 6 times more likely to have posed a level 3 tasks over a level 4 task (RRR=3.06 and RRR=.16, respectively, $p<.05$). In other words, when comparing levels of Task Potential, teachers in District C were more likely to choose lower level tasks, possibly attributable to curricular differences.

I examined the possibility of differences in the relationships between VHQMI and Task Potential and MKT and Task Potential by score quartile. The results from the analysis with categories of quartile membership are given in the second set of columns in Table 5. Because this model only includes changes with regard to the variables MKT and VHQMI, I focus on those results here. Results from this model indicated that there are no differential relationships by categories of quartile membership between MKT and Task Potential, but they do suggest differential relationships by categories of quartile membership for VHQMI (see Figure 2). In particular, it appears that having a VHQMI score in the top quartile is significantly related to teachers' choices of a level 4 task over a level 3 task (RRR= 3.10, $p<.05$): if teachers scored in the top quartile for VHQMI, they are over 3 times more likely than a teacher who scored in the first quartile to have chosen a level 4 task over a level 3 task. In Figure 2 this increased probability of selecting a level 4 task is demonstrated by the longer green section of the bar representing VHQMI Quartile 4—the right-most bar. In other words, when comparing teachers who espoused inquiry-oriented beliefs about teaching and learning mathematics with the teachers who espoused traditional beliefs, the teachers who espoused inquiry-oriented beliefs were more likely to choose a level 4 task over a level 3 task. The results also suggest that there were not significant differences in the likelihood of choosing a level 4 task over a level 3 task for teachers in the first and second or first and third quartiles of VHQMI scores. Thus, the critical difference in this sample is between teachers who espoused inquiry-oriented beliefs and teachers who espoused traditional beliefs or were transitioning toward inquiry-oriented beliefs, but had not yet fully developed inquiry-oriented beliefs. There are not significant differences by VHQMI score quartile for the choice of a low-level task over a level 3 task.

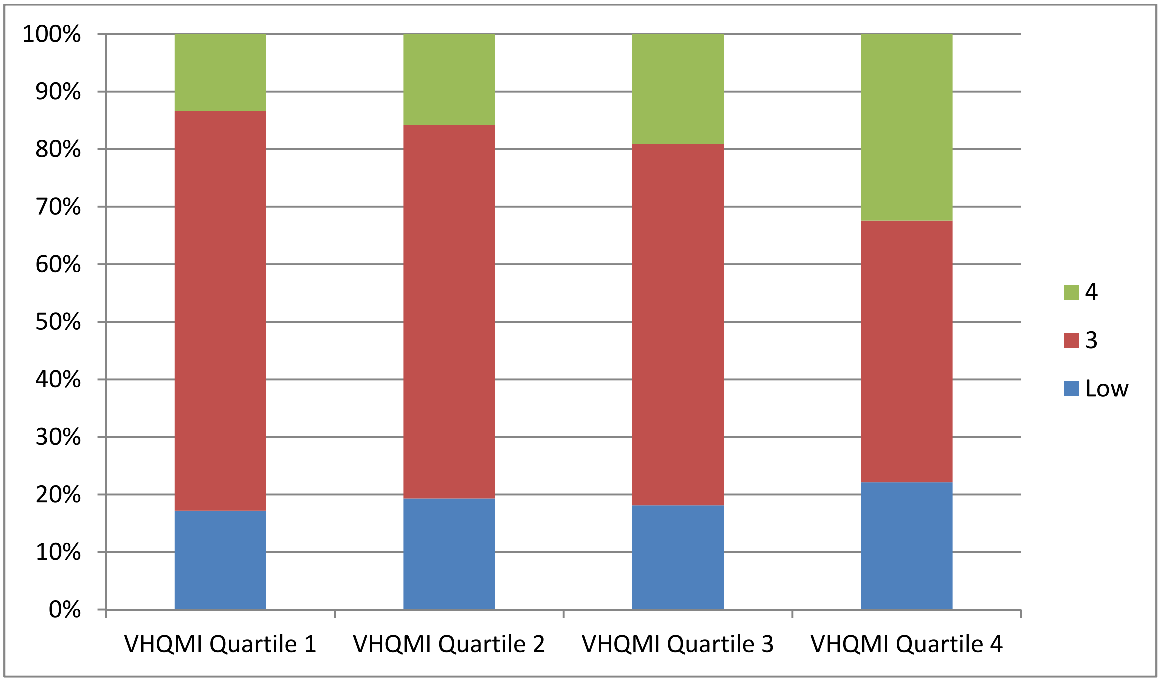


Figure 2. Task Potential by VHQMI Score Quartile

Task Potential and Hypothesized Contingencies

Following examination of simple relationships between MKT, VHQMI, and the selection of tasks, I modeled the hypothesized contingencies in several different ways. First, I modeled the existence of a continuous interaction between MKT and VHQMI. Those results are given in Table 6. The results do not suggest a significant interaction between MKT and VHQMI for the choice between a level 3 and level 4 task, but there is a marginally significant interaction between MKT and VHQMI for the choice between a low-level and level 3 task (B=-.26, RRR=0.77, p=.088). For the model specifications assuming simple relationships (given in Table 5), there is no indication that MKT and VHQMI are significantly related to the choice between a low-level task and a level 3 task. The marginally significant statistical interaction between MKT and VHQMI is a sign that the prior non-significant results might have been due to an interaction between MKT and VHQMI.

To more fully investigate specification (a) of the hypothesized contingencies, I examined the interaction of VHQMI by quartile with MKT (see model (a) in Table 7). Again, none of the interaction terms are statistically significant, but the marginally significant effects in the model of low-level tasks v. level 3 tasks gives some indication of differential effects of MKT on teachers' choices of tasks when teachers had VHQMI scores in the third quartile. Further, the model suggests that when teachers had VHQMI scores in the third quartile, compared with teachers in the first quartile of scores for VHQMI, increases in MKT correspond to decreased likelihood of choosing a low-level task over a level 3 task. In other words, when teachers had VHQMI scores in the third quartile, or espoused beliefs about teaching and learning mathematics that were approaching inquiry-oriented, the more developed their MKT, the more likely they were to choose a high-level task.

Table 6

Task Potential Model, Continuous MKT/VHQMI Interaction

	MKT/VHQMI Interaction			
	Low v 3		4 v 3	
	Coef (SE)	RRR	Coef (SE)	RRR
District A	-.34 (.45)	0.71	-.44 (.45)	0.64
District C	1.17** (.36)	3.22	-1.85** (.66)	0.16
District D	.36 (.36)	1.43	.28 (.35)	1.32
Yrs Exp	.05** (.02)	1.05	.02 (.02)	1.02
New to CMP	.74** (.35)	2.10	.85** (.35)	2.34
# Students	-.003 (.02)	1.00	.03 (.03)	1.03
Class Time	-.05 (.13)	0.95	.20 (.15)	1.22
Stud Prior M	.09 (.24)	1.09	.11 (.27)	1.12
Stud Prior SD	-.89 (.69)	0.41	.81 (.78)	2.25
MKT	-.05 (.17)	0.95	.22 (.19)	1.25
VHQMI	.04 (.13)	1.04	.32** (.15)	1.38
VHQMI * MKT	-.26* (.15)	0.77	-.01 (.19)	0.99
BSSS-P	-.49 (.41)	0.61	0.002 (.41)	1.00
BSSS-Un-coded	.08 (.28)	1.08	.11 (.41)	1.12
Constant	-.60 (.75)		-2.12** (.91)	
N	414			

Note. RRR= Relative Risk Ratios.

* p<.1. **p<.05, ***p<.001

Table 7

Task Potential Models, Hypothesized Contingency: Specifications a and b

	(a): VHQMI Moderates MKT					(b): MKT Moderates VHQMI			
	Low v 3		4 v 3			Low v 3		4 v 3	
	Coef (SE)	RRR	Coef (SE)	RRR		Coef (SE)	RRR	Coef (SE)	RRR
District A	-.50 (.46)	0.61	-.64 (.46)	0.53	District A	-.53 (.46)	0.59	-.41 (.44)	0.66
District C	1.10** (.36)	3.00	-1.89** (.66)	0.15	District C	1.15** (.36)	3.16	-1.92** (.66)	0.15
District D	.33 (.36)	1.39	.28 (.36)	1.32	District D	.40 (.36)	1.49	.28 (.36)	1.32
Yrs Exp	.05** (.02)	1.05	.02 (.02)	1.02	Yrs Exp	.05** (.02)	1.05	.02 (.02)	1.02
New to CMP	.68* (.35)	1.97	.82** (.35)	2.27	New to CMP	.65* (.36)	1.92	.78** (.35)	2.18
# Students	-.003 (.03)	1.00	.03 (.03)	1.03	# Students	-.0004 (.02)	1.00	.03 (.03)	1.03
Class Time	-.04 (.13)	0.96	.19 (.15)	1.21	Class Time	-.05 (.13)	0.95	.23* (.15)	1.26
Stud Prior M	.11 (.39)	1.12	.17 (.28)	1.19	Stud Prior M	.07 (.24)	1.07	.10 (.27)	1.11
Stud Prior SD	-.78 (.69)	0.46	1.05 (.80)	2.86	Stud Prior SD	-.74 (.70)	0.48	.90 (.80)	2.46
MKT	.50 (.35)	1.65	.66 (.50)	1.93	MKT				
VHQMI					VHQMI	.54** (.23)	1.72	.28 (.33)	1.32
VHQMI (Q2)	-.09 (.36)	0.91	-.12 (.47)	0.89	MKT (Q2)	-.56 (.34)	0.57	-.23 (.43)	0.79
VHQMI (Q3)	-.21 (.39)	0.81	.33 (.45)	1.39	MKT (Q3)	-.47 (.36)	0.63	.36 (.40)	1.43
VHQMI (Q4)	.11 (.39)	1.12	1.06** (.45)	2.89	MKT (Q4)	-.38 (.37)	0.68	.13 (.45)	1.14
MKT* VHQMI (Q2)	-.73 (.48)	0.48	-1.23* (.66)	0.29	MKT (Q2)* VHQMI	-.65* (.33)	0.52	.27 (.45)	1.31
MKT* VHQMI (Q3)	-.91* (.49)	0.40	-.47 (.63)	0.63	MKT (Q3)* VHQMI	-0.88** (.34)	0.41	-.29 (.41)	0.75
MKT * VHQMI (Q4)	-.49 (.47)	0.61	-.20 (.58)	0.82	MKT (Q4) * VHQMI	-.38 (.37)	0.68	.26 (.45)	1.30
BSSS-P	-.44 (.41)	0.64	.07 (.42)	1.07	BSSS-P	-.42 (.40)	0.66	0.003 (.41)	1.00
BSSS-Un-coded	.09 (.29)	1.09	.12 (.33)	1.13	BSSS-Un-coded	.10 (.29)	1.11	.15 (.33)	1.16
Constant	-.59 (.83)		-2.60** (1.00)		Constant	-.42 (.81)		-2.36** (.96)	
N	414					414			

Note. RRR= Relative Risk Ratios.
* p<.1. **p<.05, ***p<.001

With regard to the alternate specification of the interrelationship, specification (b), I examined whether MKT moderates the relationship between VHQMI and teachers' choice of tasks. Analogous to the previous specification, I examined the interaction between the quartile category version of MKT and VHQMI scores (see model (b) of Table 7). In this case, there are significant results for the model examining the choice of a low-level task over a level 3 task and those results are represented in Figure 3. In particular, results suggest that when teachers scored in the first quartile for MKT, a higher VHQMI score increases the likelihood of posing a low-level task over a level 3 task, in fact a teacher with a VHQMI score 1 standard deviation higher is 1.72 times more likely to pose a low-level task over a level 3 task (RRR= 1.72, $p<.05$). But, that result does not hold for teachers in the second and third quartiles for MKT; the interaction effect is in the opposite direction: the higher the VHQMI score the more likely they were to pose a level 3 task over a low-level task (RRR=.52, $p<.10$, RRR=.41, $p<.05$). The latter result is more in line with what one would expect—teachers with more sophisticated VHQMI are more likely to pose higher level tasks—but the former result suggests a need for more investigation of the relationship between teachers' beliefs about teaching and learning and their choice of tasks when their MKT is relatively undeveloped.

In sum, results from the investigation of specifications (a) and (b) of the hypothesized contingencies suggest that there is a significant interaction between VHQMI and MKT with regard to teachers' choice of low-level tasks over level 3 tasks, but no significant interaction between VHQMI and MKT for teachers' choice of level 4 tasks over level 3 tasks. Further, the original results for the choice of low-level over level 3 tasks (given in column 1 of Table 5) did not show any indication of significant linear relationships between MKT or VHQMI and teachers' choice of tasks. Yet, the examination of the statistical interactions suggests that the moderating effects at different levels of VHQMI and MKT may have been the reason for the lack of general significant effects when considering the teachers' choice of low-level or level 3 tasks.

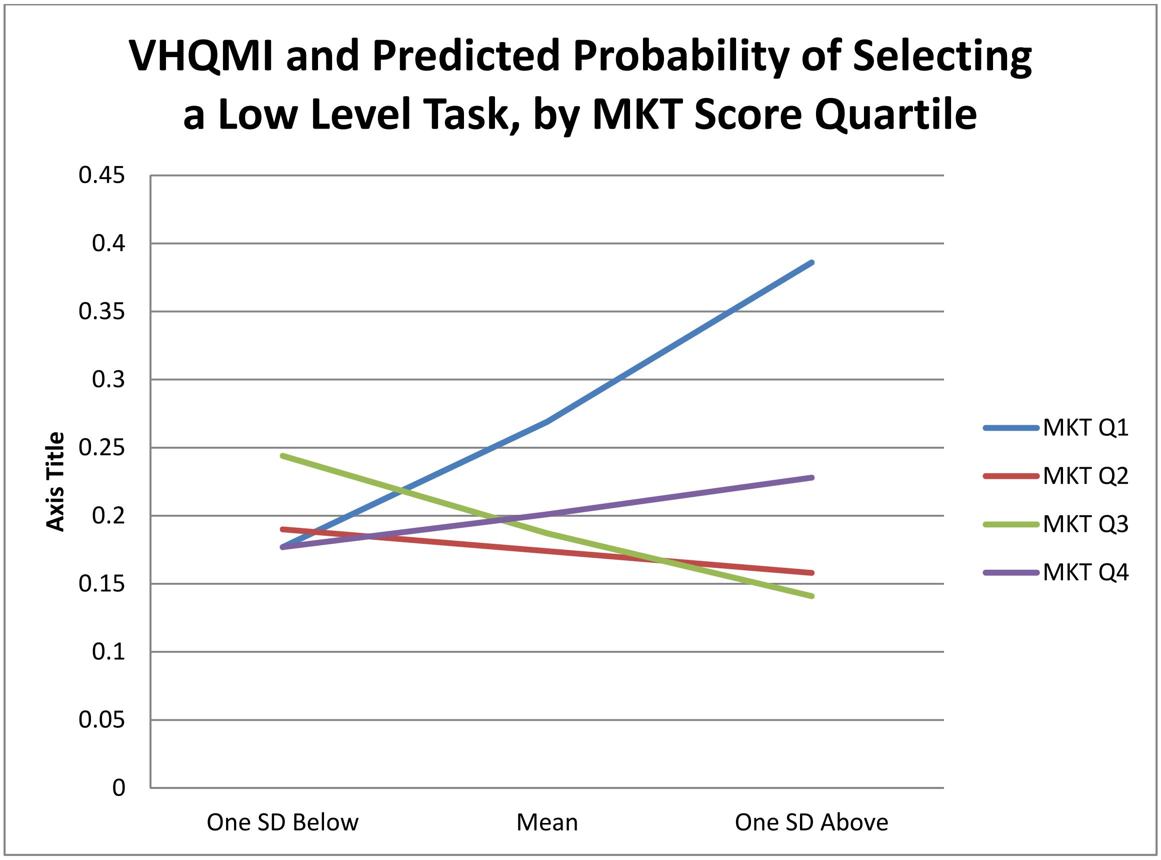


Figure 3. VHQMI and Predicted Probability of Selecting and Low-level Task, by MKT Score Quartile

Maintenance of the Cognitive Demand of High-Level Tasks

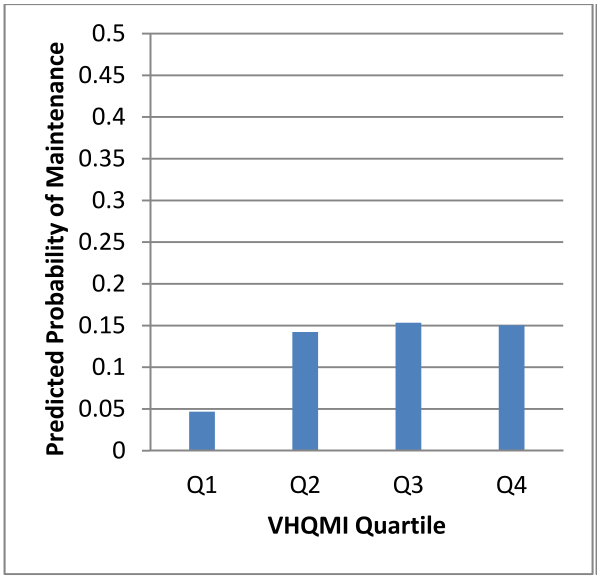
The models of maintenance of the cognitive demand of high-level tasks are similar to those of task potential, especially with regard to the independent variables included in each model. One key difference is that these models are limited to the lessons with tasks that were initially of high cognitive demand, or level 3 or 4 tasks. The outcome of interest is whether the cognitive demand of the task was maintained or decreased, a dichotomous outcome. For example, did a level 3 task stay at a 3 in implementation (maintained) or was it proceduralized to a 2 (decreased)? The first model with results in Table 8 uses the same control variables as the previous set of models and includes MKT, VHQMI, and BSSS. From the results in the first column of Table 8, we can see that MKT, VHQMI, and BSSS are significantly related to maintenance of the cognitive demand. In particular, the model suggests that the odds of maintaining the cognitive demand of a high-level task they pose for a teacher with an MKT score that was one standard deviation above the mean is 1.52 times the odds for a teacher who had the mean MKT score (B=.42, OR=1.52, p<.05). A similar result holds for VHQMI (B=.45, OR= 1.57, p<.05). In other words, the odds of maintaining the cognitive demand for teachers with VHQMI scores one standard deviation above the mean, or more inquiry-oriented VHQMI, are 1.57 times the odds of maintaining the cognitive demand for a teacher who had the sample mean VHQMI score. In addition, the odds of maintaining the cognitive demand of a high-level task for teachers who espoused productive beliefs about supporting struggling students are 2.92 times the odds for teachers who described unproductive or mixed beliefs about supporting struggling students to maintain the cognitive demand of high-level tasks (B=1.07, p<.05). Lastly, all else equal, teachers who were new to CMP2 were less likely to maintain the cognitive demand of a high-level task (B=-.91, OR=.40, p<.10). Unlike the models for Task Potential, there are no significant differences between districts for maintenance of the cognitive demand of high-level tasks, after controlling for experience, classroom and student characteristics, teachers' knowledge and beliefs, and limiting the sample to teachers who selected CDTs.

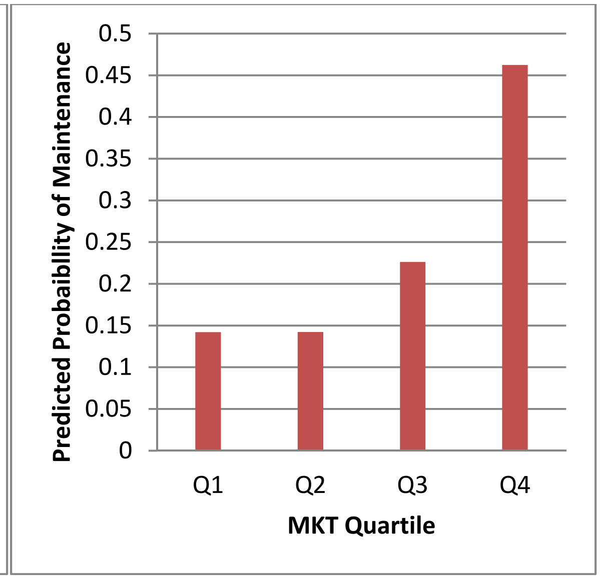
Table 8

Maintenance of the Cognitive Demand of High-Level Tasks, Models

	MKT and VHQMI		MKT and VHQMI (Categories)		MKT/VHQMI Interaction	
	Coef (SE)	OR	Coef (SE)	OR	Coef (SE)	OR
District A	.10 (.49)	1.11	.12 (.52)	1.13	.08 (.50)	1.08
District C	.15 (.52)	1.16	.14 (.55)	1.15	.14 (.52)	1.15
District D	.36 (.47)	1.43	.44 (.50)	1.55	.36 (.47)	1.43
Yrs Experience	.01 (.02)	1.01	.02 (.02)	1.02	.01 (.02)	1.01
New to CMP	-.91* (.47)	0.40	-0.99 (.49)	0.37	-.91* (.47)	0.40
# Students	.01 (.03)	1.01	.02 (.03)	1.02	.01 (.03)	1.01
Class Time	-.15 (.16)	0.86	-.11 (.16)	0.90	-.15 (.16)	0.86
Stud Prior M	.08 (.30)	1.08	.03 (.31)	1.03	.07 (.30)	1.07
Stud Prior SD	-.55 (.83)	0.58	-.73 (.85)	0.48	-.54 (.83)	0.58
MKT	.42** (.20)	1.52			.41* (.21)	1.51
MKT (Q2)			.002 (.51)	1.00		
MKT (Q3)			.57 (.49)	1.77		
MKT (Q4)			1.08** (.49)	2.94		
VHQMI	.45** (.18)	1.57			.45** (.18)	1.57
VHQMI (Q2)			1.22** (.57)	3.39		
VHQMI (Q3)			1.31** (.58)	3.71		
VHQMI (Q4)			1.28** (.58)	3.60		
BSSS-P	1.07** (.44)	2.92	1.07** (.45)	2.92	1.07** (.44)	2.92
BSSS-Un-coded	.39 (.37)	1.48	.48 (.38)	1.62	.39 (.37)	1.48
MKT * VHQMI					0.04 (.20)	1.04
Constant	-1.36 (.93)		-3.02** (1.18)		-1.37 (.93)	
N	286		286		286	

Note. OR= Odds Ratio
* p<.1. **p<.05.





Figures 4 and 5. Predicted Probability of Maintaining the Cognitive Demand, by VHQMI Score Quartile and MKT Score Quartile

The results from the model with categorical specifications for VHQMI and MKT, based on quartile membership, are given in the second set of columns in Table 8. Because this model only includes changes to the specification of MKT and VHQMI, I focus on those results here. Results from this model suggest differential effects for categories of VHQMI score quartile membership and MKT score quartile membership (see Figures 4 and 5). In particular, it appears that teachers with VHQMI scores in the first quartile were significantly less likely to maintain the cognitive demand of the task. As demonstrated in Figure 4, for teachers' with VHQMI scores in second, third and fourth quartiles, there are not significant differences in their predicted probability of maintaining the cognitive demand of high-level tasks. With regard to MKT, having an MKT score in the top quartile is significantly related to teachers' maintenance of the cognitive demand: if teachers scored in the top quartile for MKT, they were nearly 3 times more likely than teachers with scores in the first quartile to maintain the cognitive demand of the task (B=1.08, RRR=2.94, $p<.05$). The results also suggest that there are not significant differences in the likelihood of maintaining the cognitive demand for teachers in the first and second or first and third quartiles for MKT. Therefore, there is evidence of non-linear relationships between MKT, VHQMI, and the likelihood of maintaining the cognitive demand.

Table 9

Maintenance of the Cognitive Demand of High-Level Tasks, Models Continued

	VHQMI Moderates MKT			MKT Moderates VHQMI	
	Coef (SE)	OR		Coef (SE)	OR
District A	.11 (.50)	1.12	District A	.02 (.51)	1.02
District C	.15 (.52)	1.16	District C	.07 (.54)	1.07
District D	.45 (.47)	1.57	District D	.39 (.49)	1.48
Yrs Experience	.01 (.02)	1.01	Yrs Experience	.01 (.02)	1.01
New to CMP	-.98** (.48)	0.38	New to CMP	-1.10** (.50)	0.33
# Students	.02 (.03)	1.02	# Students	.02 (.03)	1.02
Class Time	-.09 (.16)	0.91	Class Time	-.14 (.16)	0.87
Stud Prior M	.04 (.31)	1.04	Stud Prior M	.09 (.30)	1.09
Stud Prior SD	-.52 (.85)	0.59	Stud Prior SD	-.53 (.86)	0.59
MKT	.66 (.63)	1.93	VHQMI	.51 (.40)	1.67
VHQMI (Q2)	1.15* (.57)	3.16	MKT (Q2)	-.05 (.52)	0.95
VHQMI (Q3)	1.30** (.57)	3.67	MKT (Q3)	.65 (.48)	1.92
VHQMI (Q4)	1.16 (.58)	3.19	MKT (Q4)	0.95* (.50)	2.59
MKT* VHQMI (Q2)	-.30 (.73)	0.74	VHQMI* MKT (Q2)	.26 (.54)	1.30
MKT* VHQMI (Q3)	-.51 (.75)	0.60	VHQMI * MKT (Q3)	-.55 (.49)	0.58
MKT* VHQMI (Q4)	.04 (.71)	1.04	VHQMI* MKT (Q4)	0.13 (.52)	1.14
BSSS-P	1.09** (.46)	2.97	BSSS-P	1.14** (.45)	3.13
BSSS-Un-coded	.44 (.37)	1.55	BSSS-Un-coded	.47 (.38)	1.60
Constant	-2.68** (1.13)		Constant	-1.88* (1.03)	
N	286		N	286	

Note. OR= Odds Ratio
* p<.1. **p<.05

Maintenance and Hypothesized Contingencies

Following an examination of the simple relationships, I then tested the hypothesized contingency of an interaction between MKT and VHQMI. First, I examined a continuous interaction between MKT and VHQMI, which was not statistically significant ($B=0.04$, $OR=1.04$, $p=.845$, see the last set of columns in Table 8). To investigate specification (a) and (b) of the contingency, I examined the interaction of VHQMI by quartile with MKT and the interaction of MKT by quartile with VHQMI, respective (see Table 9). None of the interaction terms are statistically significant. Given the results of these three models, I conclude that there is not a statistically significant interaction between VHQMI and MKT as they relate to maintenance of the cognitive demand. In other words, the hypothesized contingency does not hold for maintenance of the cognitive demand for this sample of teachers.

Discussion and Conclusion

There is evidence that CDTs provide all students with important learning opportunities (Stein & Lane, 1996; Zohar & Dori, 2003). Unfortunately, CDTs are very challenging for teachers to enact (Hiebert et al., 2003; Stein et al., 1996). Only a few studies have examined the relationships between teachers' knowledge and beliefs and the enactment of CDTs, and findings from those studies have been mixed. The results from this study shed light on middle-school mathematics teachers' enactment of CDTs in several ways. First, they confirm the findings that mathematical knowledge for teaching and beliefs about teaching and learning mathematics are related to teachers' enactment of CDTs. Second, they suggest that the relationships are nuanced both with regard to the specification of the outcome of interest and the nature of the relationships. In the following paragraphs, I highlight the key findings pertaining to mathematical knowledge for teaching and beliefs about teaching and learning mathematics and compare them to previous findings, and then I discuss implications of those findings.

The findings of this investigation suggest that MKT is related to the enactment of CDTs, even when controlling for other potentially related factors including teachers' experience and beliefs. Yet, MKT was only significantly related to some aspects of the enactment of CDTs, and the nature of the relationship varied for different levels of VHQMI. First, teachers' MKT was significantly related to maintenance of the cognitive demand of high-level tasks. Further, teachers with scores in the top quartile for MKT were better able to maintain the cognitive demand of a high-level task. Overall, MKT was not consistently and significantly related to teachers' task selection. Hence, teachers' MKT seems to feature more prominently in the actual enactment in the classroom, rather than the selection of the tasks. The variation in how MKT is related to the enactment of CDTs may be the reason for the surprisingly non-significant relationship between MKT and the enactment of CDTs in Stein and Kaufman's (2010) study where they combined task potential and implementation to create a score for the enactment of CDTs rather than examining the different score for enactment separately and examined MKT as a continuous measure. Another possibility is that controlling for teachers' visions of high-quality mathematics instruction increased the precision in the model. Neither of the previous studies of MKT and the enactment of CDTs has included measures of beliefs about teaching and learning mathematics in their examination of the relationship between knowledge and the enactment of CDTs.

Only one study has investigated how teachers' beliefs about teaching and learning mathematics are related to their enactment of CDTs. Son's (2008) study used survey measures of beliefs about teaching and learning mathematics, whereas my measures of beliefs come from interviews. My findings suggest that beliefs about teaching and learning mathematics are an important area to continue to investigate. In particular, teachers' visions of high quality mathematics instruction are related to both teachers' task selection and the subsequent maintenance of the cognitive demand of high-level tasks. With regard to teachers' task selection, I found that VHQMI was related to teachers' choice of a level 4 task over a level 3 task. Further, being in the top quartile of VHQMI, or describing an inquiry-oriented vision of high-quality mathematics instruction, was particularly important. This suggests that choosing a level 4, "doing mathematics" task over a level 3, "procedures with connections" task requires a particular level of sophistication with regard to beliefs about teaching and learning mathematics.

Beliefs about students' mathematical capabilities is an aspect of teachers' beliefs about teaching and learning mathematics that has received relatively little attention in the literature. My findings suggest that this aspect is worthy of more attention, especially as it relates to the enactment of CDTs. I found that teachers who espoused productive beliefs about supporting struggling students in interviews were more likely to maintain the cognitive demand of high-level tasks. Further, this is in addition to the increased likelihood predicted by MKT and VHQMI. Therefore, beliefs about supporting struggling students are an important factor influencing the enactment of CDTs that needs to be investigated further. However, beliefs about supporting struggling students were not significantly related to teachers' task selection. Because the measure used in this study was developed as a part of the larger study and has not been previously examined, it is important that more work is done to confirm these findings and further investigate the relationship between teachers' beliefs about students' mathematical capabilities and the enactment of CDTs.

While the focus of this analysis was on teachers' knowledge and beliefs, several findings pertain to measures included as controls. First, there is evidence that in this sample, teachers with more teaching experience were more likely to pose a low-level task over a level 3 task, but experience teaching mathematics was not significantly related to the enactment of CDTs in any of the other analyses. This is consistent with the literature suggesting that teachers who are more experienced tend to be less likely to implement new reform curricula in the intended ways (Remillard & Bryans, 2004). Also, teachers who were new to the curriculum were more likely to choose either a low-level task or level 4 task, over a level 3 task, and more likely to decrease the cognitive demand of a high-level task. There is evidence that teachers use new curricular materials in different ways (Lambdin & Preston, 1995; Remillard, 2005; Stein et al., 2007). Further, the different ways that teachers use curricular materials could result in a variety of adaptations. For example, Lambdin and Preston (1995) describe one teacher who adapted the Connected Mathematics curricular materials by: 1) demonstrating an investigation for the students to watch; and 2) incorporating more practice with procedural problems. The former adaptation would result in a Task Potential score of a 4, while the later adaptation would likely result in a Task Potential score of a 2. The case of the demonstrated investigation would be a high-level task in which the cognitive demand was not maintained. This adaptation maps onto the findings for teachers' who are new to the curriculum. Combining the findings pertaining to task selection and maintenance of the cognitive demand of high-level tasks when teachers are new to the curriculum leads to the findings that teachers who are in their first year of using the curriculum generally either pose a low-level task or decrease the cognitive demand of the high-level task they pose. This suggests a need for further research in how to support teachers to enact CDTs in their first year of curriculum use.

Another finding was that teachers in District C were significantly less likely to select tasks of high cognitive demand; given the choice between tasks of low-level and level 3, or tasks of level 3 and level 4, teachers in District C were more likely to select the lower level task. Further, since the models controlled for differences in knowledge and beliefs, this district difference was not attributed to differences in teachers' mathematical knowledge for teaching or beliefs about teaching and learning mathematics. One possible explanation is that Districts A, B, and D, all adopted the same inquiry-oriented curriculum—CMP2–but District C did not. Therefore, teachers in District C did not have the same access to CDTs as teachers did in the other districts. Lastly, despite evidence from other studies of the potential importance of the other control variables, the enactment of CDTs was not significantly related to the time allocated for the lesson, the number of students in the class, the mean prior mathematics achievement of the students in the class, or the heterogeneity of the prior achievement of students in the class.

Several qualitative studies of the relationship between teachers' knowledge and beliefs and their practice suggest that the relationships between knowledge and practice and between beliefs and practice are not necessarily straightforward. I investigated possible interrelationships between teachers' knowledge and beliefs and their enactment of CDTs by considering different hypotheses based on previous findings. I found that relationships between knowledge and beliefs and the enactment of CDTs are indeed complex. The most convincing evidence for this claim is the significant relationships between MKT and VHQMI and teachers' selection of low-level tasks over level 3 tasks when contingencies were examined. The initial lack of significant findings between mathematical knowledge for teaching and beliefs about teaching and learning mathematics and the choice between low-level tasks and level 3 tasks was disconcerting because this choice between low-level tasks and level 3 tasks is considered especially critical. While there are added benefits of engaging students in level 4 tasks, the choice between of a level 3 task over a low-level task is much more critical because level 3 tasks allow students access to the concepts underlying the mathematics, whereas low-level tasks do not.

The findings of this investigation suggest that teachers' choice between low-level and level 3 tasks is related to MKT and VHQMI, but the relationships are not straightforward. For example, in different score quartiles for MKT, increases in VHQMI had differential effects on the likelihood of choosing between a low-level and level 3 task. To some extent, these results are consistent with specification (b) which was derived from the literature: Teachers need a certain level of mathematical knowledge for teaching for their beliefs about teaching and learning mathematics to have the expected positive effect on their enactment of CDTs (Putnam et al., 1992). For teachers whose MKT scores were in the second and third quartiles, compared with those in the first quartile, having a higher VHQMI score increased the chances of choosing a level 3 task over a low-level task, but this relationship was not statistically significant for teachers with MKT scores in the fourth quartile. The lack of statistical significance at this level appears to be due to the large standard error, implying that there is more variation in this moderating relationship for teachers with MKT scores in the fourth quartile.

Across the different models, there are several results suggesting that teachers who scored in the top quartile for VHQMI or MKT were the most likely to enact CDTs, but the result from the examination of the hypothesized contingency and teachers' choice of tasks suggests that there is variation in how VHQMI and MKT are related for teachers who scored in the top quartile for MKT. This suggests that a more in-depth analysis of teachers within these top quartiles is warranted, but that we also need to consider the remaining majority of teachers as we work to support teachers to enact CDTs.

My results indicate the value of examining contingencies in large-scale studies: Without considering contingencies, I would have concluded that MKT and VHQMI are not related to the choice between low-level and level 3 tasks. Although there is still much to be learned about selecting high rather than low cognitive demand tasks, the results from investigating contingencies suggest that MKT and VHQMI are related to that choice of tasks, but that they are interrelated in complicated ways.

A second methodological implication is the importance of modeling different phases of the enactment of CDTs separately. There is good reason to believe that teachers' knowledge and beliefs are not related to teachers' choice of tasks and to the maintenance of the cognitive demand of high-level tasks in the same ways. For example, it is somewhat intuitive that teachers' beliefs about supporting struggling students might have a greater impact of their maintenance of the cognitive demand of high-level tasks, than with the choice of tasks, assuming that the district has adopted a rigorous curriculum. Without modeling choice of tasks and maintenance of the cognitive demand separately, that distinction would have been impossible to investigate. Further, Stein and Kaufman's (2010) combining of these measures might be the reason why they did not find a statistically significant relationship between MKT and the enactment of CDTs.

The finding that sophisticated mathematical knowledge for teaching, vision of high quality mathematics instruction, and beliefs about supporting struggling students are related to the enactment of CDTs indicate the importance of continuing to investigate ways to support the development of teachers' knowledge, beliefs, and practice. First, by better understanding how teachers' mathematical knowledge for teaching and beliefs about teaching and learning mathematics are related to their enactment of CDTs, teachers can be better supported to enact CDTs. Given the statistical significance and lack of prior research, more investigation is warranted with regard to teachers' beliefs about supporting struggling students and the enactment of CDTs. Also, further investigation of mathematical knowledge for teaching and beliefs about teaching and learning mathematics is warranted with particular attention to variations in effects at different levels of sophistication and interrelationships between knowledge and beliefs.

The findings of this study suggest that teachers' knowledge, beliefs, and practice are interconnected and all related to their enactment of CDTs. Therefore, as we work to support teachers' enactment of CDTs it is important that we work on developing the many facets of teachers' knowledge, beliefs, and practice simultaneously. Since the enactment of CDTs is just one dimension of ambitious teaching practice (Lampert & Graziani, 2009), it will be important to situate the enactment of CDTs within the larger activity structure that makes up ambitious teaching practice. Effective supports will center on teachers' problems of practice and push teachers to discuss and develop their mathematical knowledge for teaching and beliefs about teaching and learning mathematics. Given the complexity of these aims, it will likely be important for there to be someone or something that guides and focuses conversations (e.g., Borko, 2004; Borko, Jacobs, Eiteljorg, & Pittman, 2008; Elliott et al., 2009; Remillard & Kaye, 2002). Whether it is a mathematics coach, PD leader, or instructional tool that helps to focus teachers' conversations, the most important aspect will be intentional foci on developing teachers' knowledge, beliefs, and practice in concert.

In service of this goal, future research should examine existing supports and their effects on teachers' development of ambitious instructional practice. Hopefully this research will lead to new ideas about minor adjustments to existing supports that will make those supports more productive as well as more dramatic innovation as we work to support teachers' development of ambitious instructional practice. Given the complexity of the enactment of CDTs and other aspects of ambitious instructional practice, coordinated supports for pre- and in-service teachers' development of mathematical knowledge for teaching, beliefs about teaching and learning mathematics, and practice are critically important.

CHAPTER IV

MIDDLE SCHOOL MATHEMATICS TEACHERS' ENACTMENT OF COGNITIVELY DEMANDING TASKS: INVESTIGATING TEACHER DEVELOPMENT THROUGH INTERACTIONS WITH COLLEAGUES

Introduction

New reform goals and standards for students' mathematical learning have been put in place over the past two decades (e.g., see National Council of Teachers of Mathematics [NCTM], 1989; National Council of Teachers of Mathematics [NCTM], 2000; National Governors Association for Best Practices & Council of Chief State School National Governors Association for Best Practices & Council of Chief State School Officers, 2010). These goals for students' mathematical learning also imply new expectations for mathematics teachers' work in their classrooms. The *Curriculum and Evaluation Standards* and *Principles and Standards for School Mathematics* documents published by the National Council of Teachers of Mathematics (1989, 2000) and the more recent *Common Core State Standards* (National Governors Association for Best Practices & Council of Chief State School National Governors Association for Best Practices & Council of Chief State School Officers, 2010) reflect a consensus within the mathematics education research and policy communities for comprehensive reforms to traditional mathematics instruction. A fundamental aspect of high-quality, inquiry-oriented mathematics instruction proposed in these documents is the use of challenging, or cognitively demanding, mathematical tasks. In particular, the level of challenge of the tasks students solve and discuss is the foundation for students' mathematical learning opportunities (Doyle, 1988; Hiebert & Grouws, 2007; Stein et al., 2007). Challenging mathematical tasks support students in developing conceptual understanding (Stein & Lane, 1996). Teachers in the U.S. generally do not use cognitively demanding mathematical tasks in their classrooms, and when they do, they often enact them in ways that make them less challenging for students (Hiebert et al., 2003; Hiebert et al., 2005; Stein et al., 1996).

There is considerable evidence that it is challenging for teachers to develop the types of instructional practices described in the *Standards* documents (e.g., Ball & Cohen, 1999; Elmore et al., 1996; Lambdin & Preston, 1995; C. L. Thompson & Zeuli, 1999), and even when they believe they are teaching in a manner consistent with the reforms, practices frequently diverge from the reforms (e.g., D. K. Cohen, 1990). While researchers have begun to identify characteristics of effective teacher professional development (Desimone, Porter, Garet, Yoon, & Birman, 2002; Wilson & Berne, 1999), we still need to understand more about the design and implementation of formal professional development and about other opportunities for mathematics teacher learning (Borko, 2004). Further, what we know about professional learning (e.g., see Bruner, 1996; Lave & Wenger, 1991) suggests that interactions with colleagues in several different settings (e.g., formal professional development, teacher collaborative meetings, work with a math coach, and informal interactions with colleagues) have some potential to serve as productive sites for learning as teachers work to enact CDTs.

This study investigates changes in teachers' enactment of cognitively demanding tasks (CDTs) over time and seeks to understand whether teachers' interactions with colleagues in different settings (e.g., formal professional development, teacher collaborative meetings, work with a math coach, and informal interactions with colleagues) are related to change in their enactment of CDTs.

Conceptual Framework

The Cognitive Demand of Mathematical Tasks

The cognitive demand of a task refers to "the cognitive processes students are required to use in accomplishing it" (Doyle, 1988, p. 170). Stein, Grover, and Henningsen (1996) built on Doyle's work as they classified tasks into those with low and high cognitive demand. Tasks with low cognitive demand require students to memorize or reproduce facts, or perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks with high cognitive demand or cognitively demanding tasks (CDTs) require students to make connections to the underlying mathematical ideas. In addition, students are asked to engage in mathematical activities of explanation, justification, and generalization, or use procedures to solve tasks that are open with regard to which procedures to use. It is important to emphasize that the distinctions between high and low cognitive demand are relative to students' current understanding and, thus, are situation-dependent.

There is evidence that CDTs can provide critical learning opportunities for all students. Stein and Lane (1996) found that the use of tasks with high cognitive demand was related to greater student gains on an assessment requiring high levels of mathematical thinking and reasoning. In particular, the greatest gains occurred when teachers assigned tasks that were initially of high cognitive demand, and teachers and students maintained the cognitive demand throughout the lesson. Further, there is evidence that cognitively demanding tasks afford valuable learning opportunities for all students, not just previously high-achieving students (Zohar & Dori, 2003). The enactment of high cognitive demand tasks in the classroom therefore appears to be important in supporting all students' learning.

Unfortunately, it is clear that CDTs are not often enacted in U.S. classrooms. In attempting to understand more about changes in cognitive demand during a lesson, Stein, Grover, and Henningsen (1996) documented the initial cognitive demand of mathematical tasks as written or verbally posed to students and examined whether teachers and students maintained, increased, or decreased the demand in different phases of a math lesson. They found that in classrooms where tasks with the potential for high levels of cognitive demand were used, teachers and/or students often decreased the cognitive demand during implementation of the tasks. The point is not whether the teacher or students decreased the cognitive demand but that the cognitive demand decreased during the interactions between the students and the teacher (e.g., the students pressed the teacher to demonstrate a solution method). The results from the 1999 TIMSS video study are consistent with those of Stein and colleagues in that they suggest that the mathematical activity in U.S. middle school mathematics classrooms tends to be procedural in nature, and when teachers do select high-level tasks they often implement them in low-level ways (Hiebert et al., 2003; Hiebert et al., 2005).

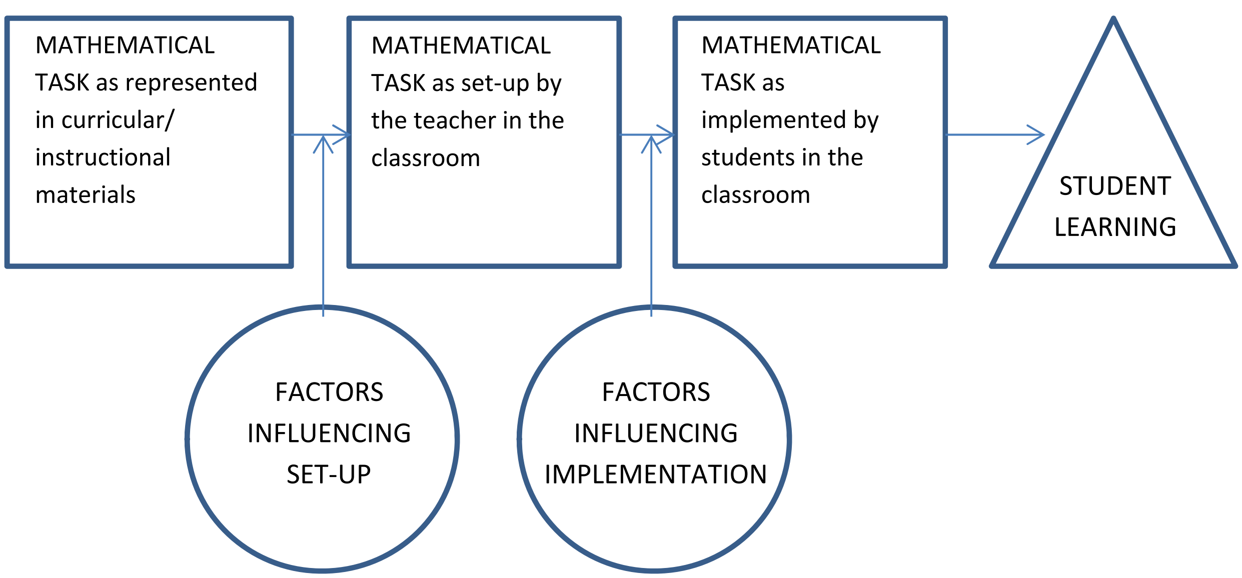


Figure 1. Modified "Math Tasks Framework" (Stein et al., 1996, p.459)

The Math Tasks Framework

Of central importance to analyzing how teachers enact tasks is a framework for examining the nature of classroom activity over the course of a lesson. The Math Tasks Framework proposed by Stein, Grover, and Henningsen (1996) is useful in thinking about changes in cognitive demand over different phases of a lesson (see Figure 1). When examining the use of CDTs in the classroom, it is important to consider both the cognitive demand of the task in particular phases of the lesson and changes in the cognitive demand from one phase to another. In this analysis, I focus on the cognitive demand of the task as selected by the teacher (corresponding to the phase in the left-most square in Figure 1) and the transition from that phase to the task as implemented by the teacher and students in the classroom (the third square from the left). This transition from selection to implementation corresponds to the maintenance of the cognitive demand over the course of the lesson. Both task selection and maintenance of the cognitive demand influence whether students engage in cognitively demanding mathematical activity in the classroom.

Supporting Teachers' Enactment of Cognitively Demanding Tasks

Only one study has investigated change in teachers' enactment of CDTs. Boston and Smith (2009) studied 18 secondary mathematics teachers' enactment of CDTs before, during, and after their participation in the Enhancing Secondary Mathematics Teacher Preparation (ESP) professional development project that focused specifically on selecting and enacting challenging mathematical tasks. They contrasted the participating teachers with 10 secondary math teachers who did not participate in the professional development program and found that through their participation in the professional development program, teachers improved in their selection and implementation of cognitively demanding tasks. Therefore, there is evidence that teachers can improve their enactment of CDTs and that focused professional development is one effective support for such change.

Teachers' Learning Opportunities through Interactions with Colleagues

In addition to professional development, interactions with colleagues in other settings (e.g., collaborative teacher meetings, informal interactions with colleagues) might be potential sites for teacher learning. Studies of professional learning indicate that co-participation in activities that approximate the targeted practices with more accomplished others is critical for the learning of complex practices (Bruner, 1996; Lave & Wenger, 1991). Further, these activities typically occur over an extended period of time (Lave & Wenger, 1991). Therefore, there are three crucial criteria to consider when investigating the potential learning opportunities that might arise for teachers through interacting with colleagues: 1) whether there are ongoing opportunities to work with others, 2) whether they involve activities that approximate targeted practices, and 3) whether they occur with someone who is more accomplished. The goal of this study was to identify a set of potentially productive interactions and empirically test whether they supported teachers' development with regard to enacting CDTs. In the following paragraphs I begin by reviewing the evidence that clarifies the importance of opportunities for interaction in ongoing, close-to-practice activities with someone who is relatively accomplished. Then, I use these three criteria to assess a priori teachers' interactions with colleagues within several different settings for their potential in supporting teachers to enact CDTs.

Opportunities for ongoing interaction. There is evidence that effective supports for improvement in teachers' instructional practice include interactions with other individuals (Frank, Zhao, & Borman, 2004; Penuel, Riel, Krause, & Frank, 2009; Putnam & Borko, 1997).

Further, several studies of effective professional development programs suggest that teachers' co-participation in PD sessions with colleagues influenced teachers' learning from the PD programs (Franke & Kazemi, 2001; Kazemi & Franke, 2004). There is also evidence that other forms of ongoing interactions with colleagues (e.g., collaborative teacher meetings, interactions with a math coach) might support teachers' development of their practice (e.g., Bruce & Ross, 2008; Diaz, 2004; Neuberger, 2010; Smith, 2000; Wang & Paine, 2003). In addition, several studies demonstrate that content of the ongoing interactions matters for supporting teachers' development (Horn & Little, 2010; Kruse & Louis, 1995). Therefore, the evidence suggests that interactions with colleagues might support teachers' development.

Activities close to practice. The findings of a number of studies indicate that engaging in activities close to practice is an effective way to support teachers' development (Ball & Cohen, 1999; Franke & Kazemi, 2001; Kazemi & Franke, 2004; Wilson & Berne, 1999). Grossman and colleagues (2009; 2008) built on these findings to suggest that activities that are close to practice should consist of both pedagogies of investigation and pedagogies of enactment. Pedagogies of investigation are more common in teacher education and professional development, and involve analyzing and critiquing practice (e.g. reviewing student work, watching and critiquing classroom video) (Kazemi, Franke, & Lampert, 2009). On the other hand, pedagogies of enactment are less common in teacher education and professional development and involve teachers actually practicing and receiving feedback on aspects of teaching. Specific to the enactment of CDTs, pedagogies of enactment could involve making decisions about tasks to be used in the classroom or rehearsing specific questions to ask students that maintain the cognitive demand, whereas pedagogies of investigation could involve looking over student work to determine whether the teacher and students maintained the cognitive demand.

Relatively accomplished colleagues. The findings of several studies indicate that co-participation with relatively accomplished colleagues is critical for teacher learning through interaction (Kruse & Louis, 1995) and that interactions with such colleagues are linked to instructional improvement (Frank et al., 2004; Penuel et al., 2009). Teachers who are attempting to develop inquiry-oriented instructional practices need support from people who are already relatively accomplished mathematics teachers, with a range of knowledge and skills including: using curriculum materials effectively to support students' attainment of ambitious mathematical learning goals, having deep mathematical knowledge for teaching, and having a vision of high-quality inquiry-oriented instruction (Hill, Schilling, & Ball, 2004). Therefore, it is important to consider the expertise of the individuals within the interactions to understand the potential for supporting teachers' development.

In sum, in considering teachers' potential learning opportunities through interactions with colleagues, it is critical to examine whether there are opportunities for ongoing interaction, whether those opportunities involve activities that approximate targeted practices, and whether they occur with someone who is more accomplished instructionally.

Possible Settings for Potentially Productive Interactions with Colleagues

The most common opportunities for mathematics teachers to interact with colleagues include: formal pull-out professional development, collaborative mathematics teacher meetings, interactions with a math coach, and informal interactions with colleagues. In this section, I assess the extent to which each of these types of interactions has the potential to support teachers' enactment of CDTs by determining the extent to which they meet the criteria described above.

Formal Pull-Out Professional Development

The three criteria described above are consistent with the characteristics of effective professional development (PD) described in the literature. In particular, the criterion that teachers have ongoing opportunities to co-participate in activities close to practice implies that they 1) involve active learning, 2) are grounded in teachers' practice 3) are coherent with other learning opportunities, 4) focus on content, 5) involve collective participation of teachers from the same school or grade, and 6) are ongoing in duration (Desimone et al., 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Horn & Little, 2010; Putnam & Borko, 1997; Wilson & Berne, 1999). The last criterion that PD be ongoing in duration is especially important because many formal PD sessions provide recommendations for practice but do little to support teachers in incorporating them into their practice (Wilson & Berne, 1999). In evaluating the potential of professional development programs, it will be important to consider the three criteria described above.

The one PD program that supported teachers in enacting CDTs, the ESP project (Boston & Smith, 2009), involved opportunities for co-participation in activities close to practice with people who are relatively accomplished. In particular, the ESP project PD gave teachers ongoing opportunities to interact with peers and mathematics education experts to work closely on the enactment of CDTs. The PD program took place over the course of two years with 11 days of professional development sessions in each the first year and 5 half-day sessions in the second year. The PD program was close to the practice of enacting CDTs in that it was centered on a framework and tools specific to the enactment of CDTs, which were used to analyze instructional episodes and written work produced by students and designed to allow teachers to use them flexibly in their classrooms (Boston & Smith, 2009). In addition, teachers participated in cycles of planning, enacting tasks in the classroom, and then reflecting on the lesson. Therefore, there is evidence that the PD program involved both pedagogies of investigation and pedagogies of enactment which may have contributed to its success in supporting teachers' development. In sum, teachers who co-participated in ongoing activities specific to the enactment of CDTs with peers and mathematics education experts generally improved their enactment of CDTs. It is unclear whether PD programs that are less specifically focused on the enactment of CDTs have the potential to support teachers in enacting CDTs.

Collaborative Mathematics Teacher Meetings

Collaborative mathematics teacher meetings might provide teachers with opportunities to interact with one another. Meetings of this sort are theoretically promising because they have the potential to involve activities that are close to teachers' practice. For example, teachers might use this time to plan lessons together, which is central to their practice. Yet, empirical results regarding the effect of teachers meeting together on their practice are mixed. Several studies found that teacher meetings (especially about the curriculum) support teachers' development of inquiry-oriented instructional practices (Diaz, 2004; Smith, 2000; Wang & Paine, 2003).

On the other hand, there is evidence that the presence of collaborative teacher meetings is a necessary but not sufficient condition in supporting teachers' development. It is also important to consider whether activities are close to practice and involve co-participation with relatively accomplished colleagues. For example, Peterson, McCarthey, and Elmore's (1996) study of three elementary schools undergoing restructuring found that allocating time for teacher collaboration was a necessary but not a sufficient condition for instructional improvement. This finding is broadly consistent with Horn and Little's (2010) findings from a study of two teacher work groups in the same high school. Horn and Little found dramatic differences in teachers' opportunities to develop their practice within interactions with colleagues in the two work groups and attributed those differences to contrasting conversational routines and the leadership within the groups. Therefore, it may also be important to account for the expertise within the meetings. The leaders in the more effective work group had a learning-centered vision for the work group which seemed to influence the conversational routines. Therefore, collaborative mathematics teacher meetings involving relatively accomplished colleagues, effective leaders, and activities that are close to practice might be potentially productive interactive settings for supporting teachers' enactment of CDTs.

Interactions with a Mathematics Coach

It is becoming increasingly common for districts to provide district- or school-based mathematics coaches to support teachers' improvement of their instructional practices (Coburn & Russell, 2008). Interactions with a mathematics coach are theoretically likely to provide productive learning opportunities for teachers provided the coach is relatively accomplished and the interactions involve activities that are close to practice. No empirical studies have investigated the influence of interactions with a coach on teachers' enactment of CDTs. However, there is some empirical evidence that interactions with a coach can support teachers' development of inquiry-oriented instructional practices. As mentioned above, Diaz (2004) found in her cross-site comparison that interactions with a content specialist (i.e., coach) supported teachers' implementation of an inquiry-oriented curriculum. In addition, Neuberger (2010) found that work with a mathematics coach assisted an elementary teacher in developing her mathematics instructional practice. Further, there is evidence that interactions with a coach can influence the nature of a teacher's informal interactions with other colleagues (Coburn & Russell, 2008), which might in turn support the development of teachers' instructional practice, as described in the next section. Hence, both theoretical considerations and empirical evidence suggests that interactions with a coach have the potential to support teachers' enactment of CDTs.

Informal Interactions with Colleagues

In addition to collaborative teacher meetings, more informal interactions with colleagues might also give rise to learning opportunities. In particular, certain advice-seeking interactions with colleagues theoretically have the potential to support teacher learning because they are likely to be closely related to teachers' instructional practice. It is not clear whether advice-seeking interactions will necessarily involve relatively accomplished colleagues. For example, a teacher might not have any colleagues who are relatively accomplished teachers at his or her school. There is some evidence that teachers' informal interactions with someone who is relatively accomplished positively influence their practice. For example, Frank, Zhao, and Borman (2004) found that teachers' informal interactions with technologically expert peers increased their use of technology in the classroom. Similarly, Sun, Garrison, Larson and Frank (Under review) found that mathematics teachers' advice seeking interactions with colleagues influenced their instructional practice. Despite the fact that no studies have specifically focused on the influence of teachers' informal interactions with colleagues on their enactment of CDTs, it is reasonable to conjecture that teachers' informal, advice-seeking interactions with relatively accomplished colleagues might support change in their enactment of CDTs.

In sum, there is evidence that several types of common interactions with colleagues have some potential to serve as productive sites for learning as teachers work to enact CDTs. The extent to which these interactions are productive is likely to depend on whether they involve relatively accomplished colleagues and the activities within in the meetings are related to the enactment of CDTs. My goal was to investigate empirically whether each of the interactional settings discussed above is associated with improvement in enacting CDTs. As I clarify below, the data set that I analyzed allowed me to investigate whether interactions in the various types of settings with colleagues who are relatively accomplished is associated with change in enacting CDTs. However, I was not able to investigate whether engaging in specific activities pertaining to the enactment of CDTs in those settings is associated with change in enacting CDTs.

Method

The two objectives of the study are as follows: 1) To describe change in teachers' enactment of cognitively demanding tasks over time, and 2) to investigate how interactions with colleagues and the expertise of colleagues within those interactions are related to any change in teachers' enactment of CDTs.

Sample

I drew on data collected in the course of a four-year study that sought to address the question of what it takes to improve the quality of middle-grades mathematics teaching, and thus student achievement, at the scale of a large urban district (Cobb & Jackson, 2011; Cobb & Smith, 2008). The research team collaborated with the leaders of four large, urban districts that were attempting to achieve a vision of high quality mathematics instruction that was compatible with the National Council of Teachers of Mathematics' (2000) *Principles and Standards for School Mathematics*. In each of the four districts, the research team selected a sample of 6 to 10 middle grades schools that reflected variation in student performance and in capacity for improvement in the quality of instruction across the district. Within each school, up to six mathematics teachers were randomly selected to participate in the study, for a total of approximately 30 teachers per district. In general, the three to six teachers in our sample from each school are just a subset of all of the mathematics teachers in the school. Further, although we tried to retain as many teachers as possible throughout the study, the sample varies from year to year as we recruited replacements when teachers left schools or changed teaching assignments.

The four collaborating school districts (Districts A, B, C and D) were typical of large, urban districts in that they had limited resources, large numbers of traditionally low-performing students in mathematics, high teacher turnover, and disparities amoung subgroups of students in their performance on state standardized tests (Darling-Hammond, 2007). For example, on average 75.7% of the students in the study schools were eligible for free or reduced price lunch, ranging from 17% to 96.3%. There were significant differences ($p<.05$) between districts in the mean percentage of students within the school eligible for free or reduced price lunch, with a District A mean of 67.4% (SD=25.1%) and a District C mean of 88.9% (SD=6.7%).

The districts were atypical in their response to high-stakes accountability pressures: they responded by focusing on improving the quality of instruction rather than teaching for the test. Consistent with this response, three of the four districts (which I will call Districts A, B, and D) adopted the Connected Mathematics Project 2 (CMP2) as their primary curriculum. The CMP2 curriculum is consistent with an inquiry-oriented approach to teaching mathematics and includes a high proportion of cognitively demanding tasks (Choppin, 2011). District C adopted a more traditional curriculum but encouraged teachers to supplement it with CMP2 and another internally-developed inquiry-oriented curriculum. District B, C, and D began implementing their respective curricula in Year 1 of the larger study. District A began implementing CMP2 in Year

2 of the larger study but had a 10 year history of using the original Connected Mathematics Project Curriculum, prior to adopting CMP2.

Each district implemented a number of strategies that were designed to support teachers in improving their instruction (e.g., curriculum frameworks, mathematics coaching, regularly scheduled time to collaborate with colleagues on issues of instruction, professional development). With respect to the types of interactions on which this study focused, there was considerable variation in strategies. Initially District A had no math coaches but subsequently assigned math coaches to some schools and also increasingly tried to encourage collaborative teachers meetings. District B hired school-based math coaches who were expected to support curriculum implementation by working with all of the teachers in the school, often in groups. District C assigned district-based math coaches to the lowest performing schools where they were expected to work with teachers in whatever way the coach and principal decided were appropriate. District C also mandated daily collaborative teacher meetings during the school day. Lastly, District D assigned district-based math coaches to all schools, with the amount of time that they spent at each school depending on student achievement. The coaches were expected to work with the weakest teachers and build capacity with groups of teachers during collaborative teacher meetings. In comparing the four districts with regard to coaching, District A, C, and D employed math coaches to varying degrees but they all encouraged coaches to work with weaker teachers, while coaches in District B were expected to work with all teachers. The emphasis on collaborative teacher meetings also varied across the districts with District C requiring daily collaborative meetings and District B not even making teacher meetings a specific strategy (although there were regular mathematics teacher meetings in some schools). All four districts offered math-specific PD on curriculum use or high quality mathematics instruction (including the use of rigorous math tasks). Although the four districts attempted to support teachers in improving their instruction, there was not an explicit focus on the enactment of CDTs in any of the districts.

In each of the four years of the study (2007-2011), we collected several types of data to test and refine a set of hypotheses and conjectures about district and school organizational arrangements, social relations, and material resources that might support mathematics teachers' development of high-quality instructional practices at scale. The primary data sources on which I drew for this study were video-recordings of teachers' classroom instruction, an assessment of teachers' and coaches' mathematical knowledge for teaching, and an online teacher survey and teacher interviews focused on the school and district settings in which teachers work (e.g., working with a coach and time allocated for collaboration with colleagues). In order to avoid omitted variable bias when accounting for colleagues' expertise, I used an expertise score from the prior year (Allison, 2005). For example, expertise scores for colleagues in Year 2 came from Year 1 of the study, and there are no prior expertise scores for Year 1. For this reason, my primary analytical sample of teachers for whom I am investigating change in their enactment of CDTs is limited to teachers in Years 2 through 4. Also, because the study sample varied each year, a cost of using prior scores is that there are missing prior scores in any given year. I used multiple imputation[4] to estimate prior scores for participants in Years 2 through 4 who were missing prior scores (i.e., scores in Years 1 through 3).

[4] I imputed several different measures of expertise (described below) using chained equations (ICE) in STATA by using information from other years of study to impute expertise information for years in which that information was missing for participants. I also took into account that participants were nested within schools by including information dummy variables for school membership in the model. For 98% of the teachers in the sample, some measure of colleagues' expertise includes imputed data to some extent, the majority of which were for some advice-seeking interaction. In comparison, only about 31% of scores were imputed for coach MKT, for example.

Measures of the Enactment of Cognitively Demanding Tasks and Other Expertise

I first describe the primary outcome measures of teachers' enactment of CDTs and the additional measures that I use to assess colleagues' expertise. Then, I describe the measures I used to represent different types of interactions with colleagues and explain how I created different aggregate measures of expertise within those interactions. Descriptive statistics for all of the measures are given in Table 2.

Enactment of CDTs. I constructed measures of teachers' enactment of CDTs by drawing on the measures of the quality of teachers' instructional practice used in the larger research project: the Instructional Quality Assessment (IQA, Boston & Wolf, 2006; Matsumura et al., 2006). We used this instrument to code video-recordings of the participating teachers' classroom instruction. In each of the years of the study, we video-recorded two (ideally consecutive) mathematics lessons conducted by each of the 120 teachers in the study in late winter. Teachers were asked to engage students in a problem-solving lesson with a related whole-class discussion.

The IQA was developed by a team of researchers at the University of Pittsburgh, and the larger study used eight of their developed rubrics to assess the quality of teachers' instruction. I focus on two of those rubrics: Task Potential and Implementation. Recall that my definition of teachers' enactment of cognitively demanding tasks includes: 1) *selecting* such tasks; and 2) *maintaining the cognitive demand* of the high-level tasks during classroom implementation. The cognitive demand of the selected task is measured by the Task Potential rubric. Maintenance of the cognitive demand of high-level tasks is measured by comparing the Task Potential and Implementation rubrics. The Task Potential and Implementation rubrics were based on the earlier work by Stein and colleagues (e.g., see Stein et al., 1996; Stein & Lane, 1996), described above.

These rubrics were designed to measure the cognitive demand of the task posed to students (Task Potential) and the cognitive demand of the task as implemented by the students and the teacher in the classroom (Implementation). Both rubrics use the same scale with five levels of cognitive demand embedded within one larger distinction: low cognitive demand or high cognitive demand. In an effort to reduce the complexity of investigating change in teachers' practice over time, in this study I focus on the crucial distinction between low cognitive demand and high cognitive demand to describe teachers' practice. Tasks coded as levels 1 and 2 are low in cognitive demand, with a level 1 task requiring only memorization or the reproduction of facts and a level 2 task requiring students to perform relatively routine procedures without making connections to the underlying mathematical ideas. Tasks coded at levels 3 and 4 represent tasks of high cognitive demand. A level 3 task requires students to make connections to underlying mathematical ideas, but tasks do not include explicit requests for generalization or justification. At the highest level, a level 4 task asks students to engage in the disciplinary activities of explanation, justification and generalization, or to use procedures to solve tasks that are somewhat open-ended in nature.

I use the Task Potential rubric to determine whether the teacher selected a high or low level task. My examination of maintenance of the cognitive demand is limited to teachers who selected a high-level task. If a teacher's scores for Task Potential and Implementation are both of high cognitive demand, then the cognitive demand of the high-level task was "maintained." If the score for Task Potential is high but the score for Implementation is low, then the cognitive demand of the high-level task "decreased."

Table 1

Reliability Information for Task Potential and Implementation

	Y1		Y2		Y3		Y4	
	Potential	Impl	Potential	Impl	Potential	Impl	Potential	Impl
% Agrmt	59.4	78.1	56.9	78.5	75	89.3	59.1	63.6
kappa	0.37	0.51	0.29	0.37	0.63	0.75	0.36	0.29

Each year we recruited doctoral students in education and master's students in mathematics education, to serve as IQA coders. An IQA developer trained coders in the summer of each year of the study. Coders were required to achieve 80 percent agreement with previously consensus-coded videos during the training reliability phase and inter-rater agreement was assessed every other week over the course of the 10 weeks of coding (resulting in double-coding of about 15% of the video sample). Table 1 gives the reliability information for each rubric in each study year because coding the video-recordings from each study year were coded each summer, which produced four sets of reliability information. Ongoing reliability was calculated with percent exact agreement and kappa scores. There is some evidence that percent agreement is not the best measure of inter-rater reliability, and that kappa scores are more accurate in measuring inter-rater agreement. A kappa score is a measure of reliability based on percent exact agreement that is adjusted for the chance agreement based on the actual distribution of the data (J. Cohen, 1960). Unfortunately, there is evidence that kappa scores are often negatively skewed when the actual scores are not well distributed (Gwet, 2010). Despite the fact that both percent agreement and kappa are imperfect measures of reliability, there are not commonly used measures that are less imperfect for measures of this complexity. Therefore, I report both percent agreement and kappa to justify the reliability of the data.

The exact agreement percentages for the Task Potential rubric were between 56.9% and 75% and kappa scores were between 0.29 and 0.63. Task Implementation reliability was slightly higher with exact agreement ranging from 63.6% to 89.3% and kappa scores between 0.29 and 0.75. There are no hard rules about sufficient reliability, but, instead, several rules of thumb. Hartmann, Barrios, and Wood (2004) suggest that appropriate agreement rates are between 80 and 90 percent, but that for more complex instruments 70% could be sufficient. The kappa scores indicate at worst “Fair” agreement and at best “Substantial” agreement (Landis & Koch, 1977), Therefore, given the complex nature of this instrument and the imperfection in the measures of inter-rater agreement, there is some indication that these inter-rater reliability scores are sufficient.

I created one set of scores for each teacher in each study year by choosing the better set of scores across the two days of coded instruction. Because teachers knew we were coming to video-record and we asked them to engage students in a problem-solving lesson, I do not consider our sample of their instruction to be representative of their typical classroom practice. Instead, the better of the two sets of scores represents the teachers’ best shot at enacting CDT in their current school context. Across the three years, 67% of teachers chose a high-level task (see Table 2). For the imputed sample across all four years, 68.0% of participants chose a high-level task (see Appendix B for a table comparing expertise in the original sample and the imputed sample).[5] In addition, across the three years, 45.0 % of teachers who posed high-level tasks implemented them in high-level ways (e.g., maintained the cognitive demand of the tasks). For the imputed sample across all four years, 42.1% of teachers who posed high-level tasks maintained the cognitive demand as high. Because understanding change in teachers’ enactment

[5] Imputed values for Task Potential and Implementation were only used to assess colleagues’ prior expertise and not with the outcome variable in the models of change in enactment of CDTs.

of CDTs is one of the objectives of this study, detailed analysis of change in the enactment of CDTs is discussed in the results section.

Mathematical knowledge for teaching. In March of each of the four years of the larger study, we assessed all participating teachers' and mathematics coaches' mathematical knowledge for teaching (MKT) by using a pencil-and-paper instrument developed by the Learning Mathematics for Teaching project at the University of Michigan (Hill et al., 2004). The instrument has a reliability index of .70 or above and can be used to assess teachers' knowledge with respect to two dimensions: number concepts and operations (NCOP); and patterns, functions and algebra (PFA). For each of the two subtests (NCOP and PFA), raw scores were translated into IRT (item response theory) scale scores (provided by MKT developers), the determination of which was based on results from a pilot administration of the assessment to a national sample of approximately 640 practicing middle school teachers. I used a combined average of these two scale scores to form a single MKT score for each participant in each year. The use of IRT scores based on the national sample allows me to relate the MKT scores of the teachers in our sample to the national average and distribution (i.e., a mean score of 0 and standard deviation of 1). For the imputed sample, the MKT scores range from -3.84 to 3.56 with a mean of -0.18 and a standard deviation of 0.95 (see Appendix B for a table comparing expertise in the original sample and the imputed sample).

Vision of high quality mathematics instruction. The Vision of High Quality Mathematics Instruction (VHQMI; Munter, Under review) measure pertains to teachers' beliefs about teaching and learning mathematics and is derived from interviews conducted with the participating teachers and coaches in January of each year of the larger study. In the interviews we asked teachers what they would look for when observing a mathematics teacher's instruction to determine if the instruction was of high quality. Depending on the breadth of their responses, participants were then asked a series of probes (see Appendix A for details).

Nine doctoral students in education and three post-doctoral education researchers were trained by the developer to code participants' VHQMI and began coding when they reached an 80% agreement level. Transcribed responses to the interview question were coded on eight different dimensions: the role of the teacher, mathematical tasks, classroom activity, and discourse (including the structure, the nature of talk, teacher questions, student questions, and student explanations) (see Appendix A for details). For each of these dimensions, scores range between 0 and 4. Participants who described more traditional instruction are at the bottom of the scale and the top of the scale is inquiry-oriented instruction that includes cognitively demanding tasks, rich whole-class discussions, and a proactive teacher who guides these activities. It was common for participants to receive scores for only some of the dimensions because there was not always enough information for coders to assign a score. Overall, the ongoing reliability percent exact agreement between coders was 80%.

To estimate participants' Visions of High Quality Mathematics Instruction (VHQMI), I used standardized mean scores that are the mean across the scored dimensions (i.e., if only two dimensions received scores, then the mean would be calculated across those two dimensions). For the imputed sample, VHQMI scores range from -0.95 to 5.15 with a mean of 2.42 and a standard deviation of 0.72 (see Appendix B for a table comparing expertise in the original sample and the imputed sample). In the models, the scores are standardized based the model sample for ease of interpretation of interaction effects. Descriptive statistics for specific measures of expertise are described below.

Measures Associated with Interactions and Contingencies

District professional development. In the spring of each year of the larger study, a survey was administered to teachers. The survey included items that asked teachers to report the extent to which the use of challenging, problem-solving tasks was addressed in professional development sessions. I used these self-reports regarding professional development as proxies for the amount of formal pull-out professional development (PD) teachers had received that related to cognitively demanding tasks. Unfortunately, I do not have information about whether the professional development they reported was specific to the enactment of CDTs and whether it involved pedagogies of investigation and enactment. Also, I do not have information about the expertise of individuals present within the PD sessions. Scores range from a 0 indicating no focus on use of challenging, problem solving tasks within professional development to a 3 indicating that professional development sessions addressed the use of challenging, problem-solving tasks to a great extent. The sample mean is 2.08 (with a standard deviation 0.90), corresponding to an average teacher reporting that district PD addressed the use of challenging, problem-solving tasks to a moderate extent.

Teacher collaborative time. In the interviews conducted in January of each year of the larger study, teachers were specifically asked about the time provided for mathematics teachers to collaborate. Teachers described the amount of time and the typical activities within the meetings. Teachers' responses to these questions were triangulated across the 3-6 interviewed teachers in each school. Because teacher meetings are only likely to support teachers' in developing their practice if they focus on problems of practice, I limited my measure of teacher collaboration (TCT) to the time in which the primary focus was on activities closely linked to teachers' practices (e.g., lesson planning). For example, I excluded meetings that were primarily administrative in nature (e.g., scheduling teachers to chaperone Saturday school, working on school improvement plan). Due to the nature of the data, I was able to discern whether the activities were generally related to teaching or more administrative in nature, but I was not able to identify whether this time focused on the enactment of CDTs. I estimated the number of hours per month of teacher collaborative time for each teacher. In some cases, the number of hours of teacher collaborative time was consistent across all math teachers in a school, whereas in other cases, the number of hours of teacher collaborative time varied by grade-level within a school. In the primary analytical sample, the mean is 6.06 (SD=5.27), indicating that the average teacher met with other math teachers to work on activities that are related to teaching for about 6 hours each month. The range is from 0 to 22 hours a month.

Table 2.

Descriptive Statistics for the Enactment of CDTs, Interactions in Different Settings, and Expertise within those Interactions

	Mean	SD	Min	Max	N
High Potential	0.67	0.47	0	1	381
Maintain	0.45	0.50	0	1	257
Work w/ Coach	13.60	15.96	0	100	377
Task PD	2.08	0.90	0	3	372
TCT	6.06	5.27	0	22	381
Advice In	68.80	109.55	0	776	378
Advice Out	13.89	44.53	0	360	378
Coach MKT	-0.003	0.85	-2.26	1.82	281
Coach VHQMI	2.53	0.71	0.68	3.85	281
TCT MKT	0.58	0.66	-1.49	1.72	381
TCT VHQMI	2.75	0.63	0	3.83	341
TCT High Pot.	0.64	0.35	0	1	378
TCT Maintain	0.17	0.33	0	1	381
MKT Advice	60.54	130.25	0	876.9	378
VHQMI Advice	77.19	168.52	0	1105.7	378
High Pot. Advice	18.38	53.31	0	360	378
Maintain Advice	8.95	34.27	0	360	378

Access to expertise in teacher collaborative time. I assessed the expertise available during teacher collaborative time ($EXPERTISE_{TCT}$) by using the maximum VHQMI and MKT for the other attendees of the meetings (including mathematics coaches). I also considered the percentage of teachers in attendance who selected high-level tasks and maintained the cognitive demand of high-level tasks, respectively.[6] The mean maximum available VHQMI score during teacher collaborative time is 2.75 (SD=0.63), with a range of 0 to 3.83. This mean indicates that on average there was someone in the meeting who described mostly inquiry-oriented instructional practices but did not consistently describe the function of inquiry-oriented forms of instruction (e.g., described that the teacher should facilitate but not the teacher's proactive role). The mean maximum available MKT score during teacher collaborative time is 0.58 (SD=.66), with a range of -1.49 to 1.72. This mean indicates the average teacher did have access to someone with more developed MKT during teacher collaborative time, but the range suggests that this was not always the case. The mean percentage of participants within teacher collaborative time who chose high-level tasks is 64% (SD=35%), and the mean percentage of participants who maintained the cognitive demand of those tasks is 17% (SD=33%). Teachers generally had access to people who chose high-level tasks and maintained the cognitive demand of those tasks, but not all of the teachers in the meetings choose high-level tasks and only about one third of teachers who chose high-level tasks maintained the cognitive demand.

Advice-seeking interactions. In the 2nd through 4th years of the larger study, the teacher survey administered in the spring included questions about teachers' advice-seeking interactions related to mathematics instruction. Teachers were asked "During this school year (including last summer), to whom have you turned for advice or information about teaching mathematics?"

[6] The fact that we did not collect MKT and VHQMI data for the entire mathematics department is a limitation on the precision of these measures of access to expertise

They were also asked about the frequency of those interactions. Choices for frequency were: (1) daily or almost daily, (2) once or twice a week, (3) once or twice per month, or (4) a few times per year. This information was re-coded to approximate the number of advice-seeking interactions teachers have with people both in (ADVICE-IN) and out (ADVICE-OUT) of their schools. District mathematics coaches who were assigned to a school are considered “in” that school. Because these advice-seeking interactions were teacher-initiated and pertaining to teaching mathematics, it is likely that they were generally close to practice, but it is not possible to infer whether they pertained to the enactment of CDTs. The mean number of times that teachers reported seeking advice from their colleagues within their schools is 68.8 (SD=109.6), with a range of 0 to 766 times. The mean number of times that teachers reported seeking advice from people outside the school is 13.9 (SD=44.5) with a range of 0 to 360 times.

Access to expertise in advice-seeking interactions. I assessed the expertise of the individuals with whom teachers interacted ($\text{EXPERTISE}_{\text{ADVICE-IN}}$) when possible by using the total available expertise in each teacher’s in-school network. The total available expertise was included as an exposure term, as is common in network influence models (Frank, Kim, & Belman, 2010). This means that the frequency of interactions was weighted by expertise and then summed across all interactions. In other words, expertise was integrated into the ADVICE-IN term rather than treating expertise as a separate main effect and indirect effect. As with collaborative teacher meetings, I considered colleagues’ VHQMI, MKT and their task selection and maintenance of the cognitive demand when applicable[7]. To avoid negative quantities of MKT advice, MKT scores were shifted by the minimum value to make all values positive. The mean MKT advice is 60.5 (SD=130.3), and the mean VHQMI advice is 77.2 (SD=168.5). Recall

[7] The fact that we did not collect MKT, VHQMI, and IQA data for the entire mathematics department is a limitation on the precision of these measures of access to expertise

that these are the number of interactions weighted by the expertise of the people from whom they report seeking advice. I also calculated the number of times they reported seeking advice from someone who posed a high-level task: mean=18.4 (SD=53.3). Similarly I calculated the number of times they reported seeking advice from a colleague who maintained the cognitive demand of a high-level task: mean=9.0 (SD=34.3). Using the mean of 68.8 times as a reference, this suggests that nearly one quarter of teachers' reported interactions were with someone who posed a high-level task and about one half of those interactions were with someone who maintained the cognitive demand of a high-level task.

Work with a mathematics coach. In the teacher survey administered every year of the larger study, teachers were asked "So far this school year, how often have the following events occurred?": 1) a mathematics coach observed my teaching (for at least 10 minutes); 2) a mathematics coach reviewed my students' work; 3) I discussed my teaching with a mathematics coach; and 4) I observed a mathematics coach demonstrate teaching in a classroom (for at least 10 minutes). The response options were: never, 1-2 times, 3-5 times, 6-10 times, 11-20 times, and more than 20 times. These results were recoded as counts of the number of times and summed across the four events (Work with Coach). Sums range from 0 to 100 with a mean of 13.6 and a standard deviation of 16.0. These particular interactions are likely to represent a subset of all of a teacher's interactions with a coach, but these activities are potentially productive ways to work with a coach because they are focused on teaching and learning (Gibbons, 2012), and are therefore a reasonable proxy for one-on-one work with a coach that might support teachers' development.

Access to expertise in working with a coach. I assessed the expertise of the mathematics coach ($\text{EXPERTISE}_{\text{Coach}}$), by using their MKT and VHQMI. Because the majority of the coaches in our sample were not concurrently teaching, I do not have information about their task selection and maintenance of the cognitive demand practices. In cases where more than one coach was assigned to a school[8], I consider the expertise of the coach with the greatest expertise because it is impossible to discern which coach the teachers referred to (or whether they referred to multiple coaches) when responding to the survey. The mean MKT score for coaches was -0.003 (SD=0.85). The mean VHQMI score for coaches was 2.53 (SD=0.71) with a range of 0.68 to 3.85. Compared to the sample means of -0.18 and 2.42 for MKT and VHQMI respectively these results suggest that, on average, the MKT or VHQMI of the coaches in the sample were not significantly more developed than those of the teachers in the schools that they served.

District fixed effects. As I have indicated above, District B and D began using an inquiry-oriented curriculum, CMP2, at the same time. In addition, District A had already been using a version of the inquiry-oriented curriculum for several years. District C only used this curriculum to supplement the more traditional curriculum it had adopted for middle-school mathematics. I use District B as the reference district and include district fixed effects (DIST A, DIST C, and DIST D) to control for differences in the adopted curriculum and the time of the adoption as well as other district contextual factors, such as district professional development and accountability climate, which are otherwise excluded from the model.

Analyses

My primary analytical sample was 195 teachers in Years 2 through 4 of the larger study, with 380 lessons in total (67 teachers with 3 years of data, 51 teachers with 2 years, and 77

[8] This was only the case for 6 schools in District D for some of Years 2-4 of the larger study: For three schools there were multiple coaches in two years and for three schools there were multiple coaches in one year.

teachers with 1 year). While teachers for whom there is only 1 year of data provide no information about change, they are included in the sample because they provide information that pertains to average behavior. The average teacher in the sample was fairly experienced in teaching mathematics when they entered the sample, with a mean of 8.12 years of experience and a range of 1 to 40 years of experience. Teachers in District A were significantly more experienced than teachers in the other 3 districts (mean=12.1 for District A, mean=7.1 for the other districts, $t=-3.3375$, $p<.001$).

I first set out to understand patterns of change in teachers' enactment of CDTs. To investigate this, I grouped teachers by patterns in their enactment across years and used t-tests and regression models to examine whether there were significant differences in their interactions with colleagues in different settings or available expertise within those interactions for the groups. The different task selection patterns include: 1) All high, 2) All low, 3) High then low, 4) Low then high, and 5) Mixed. The patterns for maintenance of the cognitive demand of high-level tasks are similar: 1) All maintain, 2) All decrease, 3) Maintain then decrease, 4) Decrease then maintain, and 5) Mixed.

As my goal was to understand how interactions with colleagues might support teachers' enactment of CDTs, I focused on teachers who had the potential for development: Teachers who initially chose low level tasks for the models of task selection and teachers who chose high-level tasks but decreased the cognitive demand of tasks when enacting them in their classroom for models of maintenance of the cognitive demand. For task selection, this includes teachers who were in the "All Low" group and teachers who were in the "Low then High" group. For maintenance of the cognitive demand, this includes teachers who were in the "All Decrease" group and teachers who were in the "Decrease then Maintain" group. There were a few teachers who started off with the less desirable outcome (e.g., selected a low level task) in the first year, then achieved the desired outcome (e.g., high-level task) in the second and third years. I excluded these teachers' third year from the models because they no longer had room for further development on the scale used in this analysis in the final year. The teachers who started similarly with the less desired outcome (e.g., selecting a low-level task) in the first year, improved in the second year (e.g., high-level task), but returned to the less desirable outcome (e.g., low-level task) in the third year were assigned the Mixed group and excluded from the models of development. Before modeling interactions with colleagues and change in the enactment of CDTs, I investigated differences between the primary sample (e.g., "All Low" and "Low then High") and the relatively accomplished group of teachers (i.e., "All High" or "All Maintain") with regard to the number of interactions and colleagues' expertise within those interactions. I assessed differences by using t-tests and regression models[9] to compare the means for the two groups. This set of analyses compared interactions and available expertise during interactions for teachers who are already relatively accomplished and the primary sample of weaker teachers for whom I then investigate improvement.

I investigated how interactions with colleagues and the expertise of colleagues involved in those interactions were related to any changes in teachers' enactment of CDTs by using multi-level logistic regression models that account for the clustering associated with multiple observations for some teachers who are nested within schools. I modeled both task selection and maintenance of the cognitive demand as dichotomous outcomes. For task selection the outcome was either high-level or low-level task. For maintenance of the cognitive demand of high-level

[9] I use regression instead of T-tests for measures of expertise because of the use of multiply imputed data. STATA does not have a routine for estimating t-tests with multiply imputed data, but using regression with the group as a predictor accomplishes the same thing (see http://www.stata.com/statalist/archive/2010-11/msg00235.html)

tasks the outcome was whether the cognitive demand was maintained or decreased, and was limited to teachers who selected high-level tasks. I investigated the effect of interactions with colleagues on task selection and on maintenance of the cognitive demand by using several rounds of models. The first round of models include only structural aspects of the interactions (e.g., how many times teachers worked with a coach, how much time each month they spent in collaborative teacher meetings), and exclude the expertise of the participants:

$$(HIGH\ POTENTIAL\ or\ MAINTENANCE)$$
$$= \beta_1(PD) + \beta_2(ADVICE - IN) + \beta_3(ADVICE - OUT) + \beta_4(TCT)$$
$$+ \beta_5(Work\ with\ Coach) + \beta_6(Year) + \beta_7(DIST\ A) + \beta_8(DIST\ C)$$
$$+ \beta_9(DIST\ D) + \varepsilon$$

The second, third, and fourth rounds of models include the expertise of colleagues within those meetings. For each round, models include a different type of expertise (VHQMI, MKT, and Task Selection or Maintenance, respectively), and I first model the direct effect of the expertise, followed by a model that includes direct and interaction effects for each measure of expertise. Because coach expertise only applies when there was actually a coach in the school and this further reduces the sample of teachers, I examined the effects of coach expertise as it pertains to working with a coach separately from the effects of other expertise. If coaches were present in collaborative teacher meetings or were indicated as people from who teachers seek advice, their expertise was also included in the measures associated with those settings. Models of coach expertise were of the form:

$$(HIGH\ POTENTIAL\ or\ MAINTENANCE)$$
$$= \beta_1(PD) + \beta_2(ADVICE - IN) + \beta_3(ADVICE - OUT) + \beta_4(TCT)$$
$$+ \beta\ \ (Work\ with\ Coach) + \beta\ \ (EXPERTISE_{Coach}) + \beta_7(Time) + \beta_8(DIST\ A)$$
$$+ \beta_9(DIST\ C) + \beta_{10}(DIST\ D) + \varepsilon$$

Models of expertise within other interactions were of the form:

$$(HIGH\ POTENTIAL\ or\ MAINTENANCE)$$
$$= \beta_1(PD) + \beta_2(ADVICE - IN) + \beta\ \ (EXPERTISE_{ADVICE-IN})$$
$$+ \beta_4(ADVICE - OUT) + \beta_5(TCT) + \beta_6(EXPERTISE_{TCT})$$
$$+ \beta\ \ (Work\ with\ Coach) + \beta\ \ (Time) + \beta_9(DIST\ A) + \beta_{10}(DIST\ C)$$
$$+ \beta_{11}(DIST\ D) + \varepsilon$$

By first examining the effect of participation in different types of interactions with colleagues that have the potential to be close to practice, then including the expertise, and finally including statistical interactions between participation and expertise, I considered the incremental effect of several forms of expertise over and above the sole effect of participation and how the availability of expertise might moderate the effects of participation. Lastly, in the cases where the models produced significant results that were difficult to interpret, I conducted additional analyses to aid in interpretation. For example, I estimated additional models to better understand possible district differences with regard the effect of working with a coach on task selection. When comparing models that did not involve multiple imputation, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) model fit statistics are provided.

Results

Recall that my two objectives were to characterize change in the enactment of CDTs and investigate how interactions with colleagues in different settings and the availability of expertise within those interactions were related to change in the enactment of CDTs. In addressing these questions, I examined task selection and maintenance of the cognitive demand of high-level tasks separately. In presenting the results for each aspect of enacting CDTs, I first describe patterns of change, then examine differences in interactions and expertise for different patterns of change, and finally model relationships between interactions and change in the enactment of CDTs for teachers who had the potential for development (i.e, “All Low” and “Low then High” for task selection or “All Decrease” and “Decrease then Maintain” for maintenance).

Table 3

Teachers and Number of Observations for Task Selection, by Change Pattern

Trend	1 observation	2 observations	3 observations	Total
All High	56	21	25	102
Mixed	---	---	9	9
High then Low	---	9	13	22
Low then High	---	10	15	25
All Low	21	11	5	37
Total	77	51	67	195

Patterns of Change in Teachers’ Task Selection

To understand change in teachers’ enactment of CDTs, I first classified teachers by their patterns of change in task selection across Years 2 through 4 of the larger study. Table 3 summarizes the distribution of teachers by change pattern and number of observations. As mentioned above, teachers who were observed in only one year are included in this analysis because they provide information about average behavior. There are two groups that make up the primary sample of 62 teachers for further investigation of change in Task Selection: the 25 teachers classified as "Low then High" and the 37 teachers classified as "All Low." All of these teachers started with a low-level task and some of them eventually selected a high-level task whereas others did not. The "Low then High" group is made up of the 25 teachers who selected a low-level task the first time they were observed, but in Year 3 or 4 selected a high-level task (and then did not drop back down to a low level task in the subsequent year). The "All Low" group includes the 37 teachers who only posed low level tasks. There is only one observation for 21 of these teachers, who contribute no information about change. I will call the sample of teachers in the "Low then High" and "All Low" groups the primary task selection sample.

There were 102 teachers who posed high-level tasks in every year we observed them, and 46 of those teachers were observed for multiple years of the study. There were 22 teachers who selected a high-level task in the first year we observed them, but in Year 3 or 4 selected a low level task. Further, there are 9 teachers who fluctuated between high- and low-level tasks over the three years of the study and were classified as "Mixed": 7 teachers posed a high-level task, then a low-level task, then another high-level task, and 2 teachers began with a low-level task, then posed a high-level task, and then a low-level task. The 22 teachers classified as "High then Low" and the 9 teachers classified as "Mixed" were excluded from the analyses of interactions or expertise within interactions because they represent decline and inconsistency which are not the foci of this study. The teachers within the "All High" category are an interesting comparison group, but they provide no information about change.

Group Differences for Patterns of Change in Task Selection

Prior to investigating whether participation in different types of interactions is related to change in the primary task selection sample, I examined whether there were any descriptive differences between the primary sample (i.e., the teachers who had the potential for development) and the other teachers (e.g., the “All High” group). I first examined the distribution of teachers across the four school districts (see Table 4) for three categories of teachers: the primary sample (i.e., “All Low” or “Low then High”), teachers who always selected a high-level task (i.e., “All High”), and the other teachers (i.e., “Mixed” and “High then Low”). First, a Kolmogorov-Smirnov test for equality of distribution functions suggests that there are marginally significant differences in how the three categories of teachers were distributed across the districts (p=.069). In particular, over 60% of the teachers in district A and D, and nearly 60% of the teachers in District B were in the “All High” group. On the other hand, only 17.5% of the teachers from District C were in the “All High” group and 50% of the teachers from District C were in the primary sample. In contrast, 23.1%, 24.1%, and 32.8% of teachers from Districts A, B, and D, respectively, were in the primary sample. There is also a higher percentage of teachers in District C who fall into the “other” category. Given what we know about the curricular differences between District C and the other three districts, it is not surprising that more teachers in District C either entered the study selecting low level tasks or waivered between low and high cognitive demand tasks over time.

Table 4.

Teacher Distributions across three Categories of Task Selection Change Patterns and School District

District	Primary Sample		All High		Other		Totals
	#	% of District Total	#	% of District Total	#	% of District Total	
A	9	23.1	25	64.1	5	12.8	39
B	14	24.1	34	58.6	10	17.3	58
C	20	50.0	7	17.5	13	32.5	40
D	19	32.8	36	62.1	3	5.1	58

Table 5

Descriptive Statistics for Interactions in Different Settings and Expertise within those Interactions for the Primary Task Selection Sample and the "All High" Change Pattern

	Primary Task Selection Sample			All High		
	Mean	SD	N	Mean	SD	N
Work w/ Coach	14.0	15.8	109	14.0	17.0	171
Task PD**	2.0	1.0	109	2.2	0.9	168
TCT**	6.8	5.6	109	4.3	4.2	173
Advice In**	85.8	122.0	109	59.2	110.7	172
Advice Out**	3.4	9.4	109	17.6	47.1	172
Coach MKT	-0.0005	0.93	80	0.15	0.76	123
Coach VHQMI	2.53	0.74	80	2.55	0.70	123
TCT MKT	0.51	0.72	109	0.60	0.66	173
TCT VHQMI**	2.62	0.75	109	2.84	0.59	173
TCT High Potential	0.63	0.33	109	0.66	0.37	171
MKT Advice*	83.08	162.65	109	52.67	118.81	172
VHQMI Advice	89.11	172.59	109	75.49	180.52	172
High Potential Advice	23.63	63.96	109	18.84	52.80	172

*p< 0.1, **p<0.05.

To begin to understand how participation in different types of interactions might be related to teachers' task selection, I compared means and standard deviations of interactions with colleagues (i.e., interactions variables) and expertise available in those interactions (i.e., expertise variables) for the primary sample and the "All High" group (see Table 5). T-tests suggest that there are at least marginally significant differences between the primary sample and the "All High" group for all of the interactions variables except work with a coach. However, the direction of the difference is not consistent across different interactional settings. Teachers in the primary sample participated in less professional development that focused on enacting challenging, problem-solving tasks ($p<.05$), participated in more collaborative teacher meetings focused on practice ($p<.05$), reported more advice-seeking interactions within their schools ($p<.05$), and reported fewer advice-seeking interactions outside of their schools ($p<.05$).

There are two significant differences between the primary sample for task selection and the "All High" group with regard to access to expertise: colleagues' VHQMI in collaborative teacher meetings ($p<.05$), and MKT expertise through advice-seeking interactions ($p<.10$). Results suggest that teachers in the "All High" group had access to colleagues with more sophisticated VHQMI during collaborative teacher meetings and less access to MKT expertise through advice-seeking interactions than teachers in the primary task selection sample. Another relevant finding that is true for both the primary sample and the "All High" group is that the maximum available expertise in collaborative teacher meetings (i.e., TCT MKT and TCT VHQMI) was generally greater than the available coach expertise (i.e., coach MKT and coach VHQMI). These finding suggests that mathematics coaches in these districts were generally not the most expert colleagues with regard to MKT and VHQMI, and that collaborative teacher meetings usually gave teachers access to relatively accomplished colleagues.

Given the differences in district strategies related to these types of interactions (e.g., coaching and collaborative teacher meetings), I also investigate differences between the primary sample for task selection and the "All High" group within districts. Table 6 reports the means, standard deviations, and the number of observations for the primary sample and the "All High" group within districts, and identifies statistically significant differences. Several results shed additional light on the differences between the primary sample and the “All High” group. Although there were no significant overall differences between the samples with regard to working with a coach, there are important differences between samples when separated by districts. For District B, teachers in the “All High” group reported significantly more interactions with a coach ($t= 1.76$, $p<.05$). In the other districts, the differences between groups were not statistically significant, but the means tended in the opposite direction: teachers in the primary sample reported working more with a coach than did teachers in the “All High” group. These findings are consistent with the differences in district designs for coaching: in District B, coaches were expected to work with all teachers, whereas in the other districts, coaches were expected to work with the neediest teachers. These differences between districts also provide an explanation for why there is no significant overall difference with regard to working with a coach.

Another notable difference between districts with regard to task selection concerns the level of accomplishment of colleagues in collaborative teacher meetings. Although there was no significant overall difference between the primary sample and the “All High” group, there were significant differences between the samples within districts. Further, the differences were in opposite directions. In particular, in District B, the mean percentage of teachers in collaborative teacher meetings who selected high-level tasks was higher for the “All High” group than the primary task selection sample ($b=0.17$, $SE=0.08$, $p<.05$), but in District C, the mean percentage of teachers in collaborative teacher meetings who selected high-level tasks was higher for the primary sample than for the “All High” group ($b=-0.18$, $SE=0.10$, $p<.10$). These results seem to suggest that participants in District B teacher meetings had similar instructional practices, but that in District C there was less consistency in meeting participants’ instructional practices. This variation explains why there was no significant overall difference between the two samples.

There is no clear explanation for why more accomplished teachers in District B tended to participate in meetings with colleagues who were also more accomplished, while more accomplished teachers in District C tended to participate in meetings with teachers who were less accomplished. This gives some indication that schools in District B might be more homogeneous than schools in District C with regard to task selection.

Table 6

Descriptive Statistics for Interactions in Different Settings and Expertise within those Interactions for the Primary Task Selection Sample and the "All High" Change Pattern, By District

	District A				District B				District C				District D			
	P. Sample		All High		P. Sample		All High		P. Sample		All High		P. Sample		All High	
	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Work w/	9.4	19	6.9	51	8.7	23	14.6**	55	11.9	37	8.5	11	23.6	30	21.4	54
Coach	(11.4)		(17.1)		(9.7)		(14.6)		(15.1)		(10.3)		(18.8)		(17.2)	
Task PD	1.7	19	2.4**	51	1.7	23	2.0	53	2.3	37	2.5	11	2.0	30	2.1	53
	(0.7)		(0.7)		(1.0)		(0.9)		(0.8)		(0.8)		(1.1)		(0.9)	
TCT	1.9	19	1.6	52	5.0	23	5.9	55	13.3	37	12.0	12	3.3	30	3.5	54
	(1.8)		(1.8)		(3.7)		(3.0)		(2.9)		(4.7)		(2.8)		(4.1)	
Advice In	59.6	19	40.6	52	101.7	23	75.9	55	112.2	37	131.6	11	57.6	30	45.7	54
	(101.4)		(80.7)		(133.1)		(137.0)		(147.7)		(196.8)		(78.6)		(70.0)	
Advice	4.2	19	31.7**	52	4.3	23	10.7	55	5.1	37	17.5*	11	0	30	11.1**	54
Out	(8.8)		(63.0)		(10.4)		(34.7)		(12.2)		(54.0)		(0)		(35.9)	
Coach	0.38	7	0.72	7	-0.24	23	-0.11	55	-0.64	20	-0.77	7	0.52	30	0.47	54
MKT	(1.29)		(0.66)		(0.62)		(0.60)		(1.00)		(0.87)		(0.59)		(0.70)	
Coach	1.84	7	3.08	7	2.44	23	2.36	55	2.17	20	2.45	7	2.78	30	2.69	54
VHQMI	(0.57)		(0.75)		(0.66)		(0.68)		(0.84)		(0.82)		(0.63)		(0.63)	
TCT MKT	0.90	19	0.66	52	0.41	23	0.59	55	0.30	37	0.28	12	0.59	30	0.62	54
	(0.74)		(0.76)		(0.68)		(0.63)		(0.80)		(0.71)		(0.50)		(0.55)	
TCT	2.84	19	2.93	29	2.77	23	2.74	54	2.21	37	2.45	11	2.94	30	2.98	52
VHQMI	(0.66)		(0.55)		(0.48)		(0.51)		(0.91)		(0.88)		(0.50)		(0.57)	
TCT High	0.44	19	0.49	51	0.60	23	0.78**	55	0.61*	37	0.44	12	0.79	30	0.76	54
Pot.	(0.48)		(0.47)		(0.29)		(0.28)		(0.28)		(0.33)		(0.25)		(0.26)	
MKT	114.1	19	51.2	52	47.0	23	71.9	55	96.0	37	50.1	11	75.2*	30	35.0	54
Advice	(211.3)		(105.3)		(97.1)		(160.9)		(188.8)		(85.5)		(132.1)		(80.0)	
VHQMI	82.5	19	62.9	52	72.8	23	125.1	55	103.0	37	104.4	11	88.7**	30	31.2	54
Advice	(167.9)		(153.0)		(145.5)		(256.5)		(198.4)		(168.7)		(167.5)		(73.1)	
High Pot.	26.4	19	18.7	52	7.9	23	28.1	55	27.8	37	10.0	11	28.7	30	11.3	54
Advice	(61.8)		(37.6)		(37.5)		(77.2)		(71.0)		(23.2)		(72.7)		(36.0)	

*p< 0.1, **p<0.05.

Lastly, although there are not significant differences between teachers in the primary sample for task selection and teachers in the "All High" group with respect to VHQMI expertise through advice-seeking interactions across the four districts, there are significant differences between the two samples in District D. In particular, District D teachers in the primary sample have greater access to VHQMI expertise through advice-seeking interactions ($p<.05$).

Overall, there are some notable differences between teachers who initially posed low-level tasks and teachers who posed all high-level tasks. Teachers who posed all high-level tasks reported more of an emphasis on challenging, problem-solving tasks in PD, fewer hours of collaborative teacher meetings each month, fewer advice-seeking interactions within their schools, more advice-seeking interactions outside of their schools, had access to colleagues with more inquiry-oriented VHQMI in collaborative teacher meetings, and had access to greater MKT expertise through advice-seeking interactions[10]. These findings might be partially explained by the distribution of teachers across the districts, especially as only 9 teachers from District A are in the primary sample, and only 7 teachers from District C are in the "All High" group. Also, there are some notable differences by district: teachers in District B who posed all high-level tasks reported more interactions with a coach, teachers in District C who posed all high-level tasks had access to fewer colleagues who also posed high-level tasks, and teachers in District D who posed all high-level tasks had less access to VHQMI expertise through advice-seeking interactions. In sum, there is some indication that weaker teachers (i.e., teachers in the primary sample) participated in fewer PD sessions that focused on enacting challenging tasks, but participated in more collaborative teacher meetings and advice-seeking interactions within their

[10] While some of these behaviors might be consistent with teachers who are more experienced, these trends are not attributable to teachers' experience in this case: In fact, teachers in the primary sample for Task Selection were significantly more experienced in teaching mathematics (mean=10.9 for primary sample and mean=8.5 for All High Sample, $p<.05$).

schools[11]. Also, coaches in District B tended to work more with stronger teachers (i.e., "All High"), and coaches in the other districts tended to work more with the weaker teachers (i.e., the primary sample). The next step in the analysis was to investigate whether interactions with colleagues and the available expertise within those interactions were related to improvement in selecting high-level tasks.

Models of Interactions and Change in Teachers' Task Selection

In this section, I describe results from models of interactions and available expertise within those interactions on change in teachers' task selection in order to understand how different types of interactions might support teachers to improve their task selection. The primary sample for addressing these questions consists of 62 teachers in 25 different schools who initially selected low-level tasks. Recall that 25 of those teachers improved (i.e., eventually chose a high-level task), 16 teachers continued to select low level tasks, and 21 teachers were only in the study for one year and selected a low level task in that year. Results from the 11 models of time spent interacting with colleagues and the expertise of colleagues within those interactions are given in Tables 7 and 8. Because quite a few of the models only differ slightly in which variables were included, I first explain the initial model and then describe trends and notable results across the other models.

A first general finding is that the number of interactions that teachers reported with a coach is negatively related to the selection of high-level tasks. For example, results from model (1), the model of interactions without considering expertise, suggest that if a teacher worked with a coach one standard deviation more than the average (approximately 30 times, instead of 14

---

[11] This set of results is not significantly associated with the poverty level of the school (i.e., the percentage of students eligible for free or reduced price lunch).

times), then their odds of choosing a high-level task are 0.10 (i.e., approximately 1 to 10) the odds of a person who reported working with a coach 14 times this year ($p<.05$). In other words, the more a teacher worked with a coach, the less likely they were to choose a high-level task. There are two possible interpretations: 1) the more a teacher worked with a coach, the less likely they were to *improve* in their task selection, or 2) coaches tended to work with teachers who were less likely to select a high-level task. I investigate and discuss these different interpretations, below.

A second general finding is that the number of within-school, advice-seeking interactions teachers reported is positively related to the selection of high-level tasks (see models (1)-(3), (6), and (7)). For example, results from model (1) suggest that the odds of choosing a high-level task for teachers who reported one standard deviation more advice-seeking interactions with colleagues than the mean, or 206 total advice-seeking interactions, are 3.06 times the odds for a teacher who reported the mean of about 86 interactions. While 206 advice-seeking interactions might seem high, it is equivalent to interacting with one person daily and two other people once a month. Further, this finding holds when the number of advice-seeking interactions is weighted by the VHQMI and task selection expertise of the colleagues with whom the teachers interacted ($p<.1$ and $p<.05$, respectively, see models (8)-(11)), but it does not hold when advice-seeking interactions are weighted by their colleagues' MKT (see models (4) and (5)). One possible interpretation is that the effect of advice-seeking interactions is reduced by different forms of expertise to different degrees. I discuss this interpretation further below.

Table 7

Models of Interactions with Colleagues and Expertise within those Interactions on Improvement in Teachers' Task Selection: Interactions and MKT Expertise

Model	(1) Interactions with Colleagues		(2) Coach MKT	(3) Coach MKT 2	(4) MKT	(5) MKT 2
High Potential	Coef (SE)	Odds Ratio	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
Year	2.75**	15.64	2.77**	2.83**	2.04**	2.19**
	(0.64)		(0.82)	(0.86)	(0.55)	(0.59)
District A	1.07	2.92	1.09	1.18	0.54	0.72
	(1.00)		(1.17)	(1.19)	(0.93)	(0.96)
District C	-0.98	0.38	-0.28	-0.39	-0.98	-0.90
	(1.21)		(1.78)	(1.79)	(1.19)	(1.18)
District D	2.14**	8.50	2.08*	2.08*	1.46	1.60
	(1.07)		(1.21)	(1.22)	(0.96)	(0.99)
Task PD	-0.17	0.84	-0.35	-0.34	0.12	0.08
	(0.33)		(0.41)	(0.41)	(0.30)	(0.31)
TCT	-0.08	0.92	-0.15	-0.13	-0.03	-0.03
	(0.61)		(0.70)	(0.70)	(0.60)	(0.58)
TCT Expertise					0.17 (0.34)	0.02 (0.36)
TCT*Exp						-0.33
						(0.32)
Advice-In	1.12**	3.06	1.31**	1.28**		
	(0.43)		(0.53)	(0.54)		
Advice-In Expertise					0.20 (0.33)	0.26 (0.34)
Advice-Out	3.09*	21.98	1.25	1.09	2.31	2.48
	(1.76)		(2.52)	(2.51)	(1.65)	(1.73)
Work with Coach	-2.26** (0.69)	0.10	-2.63** (0.92)	-2.66** (0.98)	-1.90** (0.64)	-1.98** (0.66)
Coach Expertise			0.15 (0.55)	0.06 (0.66)		
WWC*Exp				-0.34		
				(0.87)		
Constant	-4.08**		-4.43**	-4.47**	-3.66**	-3.83**
	(1.20)		(1.54)	(1.59)	(1.19)	(1.23)
N	109		87	87	109	109

*p< 0.1, **p<0.05.

Table 8

Models of Interactions with Colleagues and Expertise within those Interactions on Improvement in Teachers' Task Selection: VHQMI and Task Selection Expertise

Model	(6) Coach VHQMI	(7) Coach VHQMI 2	(8) VHQMI	(9) VHQMI 2	(10) Task Selection		(11) Task Selection 2
High Potential	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Odds Ratio	Coef (SE)
Year	2.82**	2.84**	2.20**	2.23**	2.42**	11.24	2.52**
	(0.79)	(0.82)	(0.58)	(0.59)	(0.59)		(0.62)
District A	0.90	0.96	0.50	0.46	0.50	1.64	0.23
	(1.17)	(1.18)	(0.95)	(0.96)	(0.97)		(1.02)
District C	-0.24	-0.13	-1.14	-1.33	-1.16	0.31	-0.79
	(1.72)	(1.78)	(1.25)	(1.26)	(1.28)		(1.36)
District D	1.97	2.02	1.53	1.39	1.50	4.48	1.41
	(1.20)	(1.22)	(1.00)	(1.03)	(1.06)		(1.05)
Task PD	-0.36	-0.35	-0.02	-0.02	0.003	1.003	-0.03
	(0.39)	(0.40)	(0.32)	(0.32)	(0.32)		(0.32)
TCT	-0.19	-0.18	-0.6	0.04	-0.06	0.94	-0.25
	(0.67)	(0.68)	(0.61)	(0.59)	(0.68)		(0.73)
TCT			0.14	0.16	0.01	1.01	0.21
Expertise			(0.59)	(0.56)	(0.37)		(0.43)
TCT*Exp				-0.30			0.47
				(0.41)			(0.47)
Advice-In	1.28**	1.28**					
	(0.54)	(0.55)					
Advice-In			0.66*	0.68*	0.79**	2.20	0.69*
Expertise			(0.36)	(0.36)	(0.34)		(0.35)
Advice-Out	1.38	1.20	2.27	2.34	2.60	13.46	2.36
	(2.47)	(2.53)	(1.69)	(1.71)	(1.69)		(1.64)
Work with	-2.64**	-2.93**	-2.18**	-2.22**	-2.13**	0.12	-2.05**
Coach	(0.91)	(1.11)	(0.73)	(0.76)	(0.75)		(0.70)
Coach	0.24	0.45					
Expertise	(0.47)	(0.59)					
WWC*Exp		0.53					
		(0.85)					
Constant	-4.34**	-4.64**	-3.71**	-3.68**	-3.86**		-4.08**
	(1.50)	(1.66)	(1.21)	(1.21)	(1.28)		(1.30)
N	87	87	109	109	109		109

*p< 0.1, **p<0.05.

The third general finding is that the results from models which included measures of expertise (i.e., models (2)-(11)) suggest that colleagues' MKT, VHQMI, and task selection are mostly not significantly related to change in teachers' task selection, and there are no statistically significant indirect effects of expertise within the interactions on the effect of the interactions. In particular, coach MKT and VHQMI are not significantly related to teachers' task selection, nor do they moderate the relationship between working with a coach and task selection. Further, the MKT, VHQMI, and task selection expertise of colleagues within collaborative teacher meetings are not significantly related to teachers' task selection and they do not moderate the effect of collaborative teacher meetings on change in task selection. The only expertise measures that are significantly related to the selection of high-level tasks are VHQMI and task selection expertise through advice-seeking interactions (see table 8). These results are discussed further below.

Table 9

District differences in the Effect of Working with a Coach

Model	(1) Compared to District B	(2) Compared to District D
High Potential	Coef (SE)	Coef (SE)
Year	3.04** (0.72)	3.04** (0.72)
District A	0.84 (1.43)	-0.99 (1.19)
District B		-1.82 (1.32)
District C	-0.57 (1.64)	-2.39 (1.60)
District D	1.82 (1.32)	
Task PD	-0.13 (0.34)	-0.13 (0.34)
TCT	-0.29 (0.64)	-0.29 (0.63)
Advice-In	1.13** (0.44)	1.13** (0.44)
Advice-Out	2.94* (1.73)	2.94* (1.73)
Work w/ Coach	-2.07 (1.70)	-5.60** (2.53)
Work w/ Coach in A	-0.19 (2.10)	3.34 (2.73)
Work w/ Coach in B		3.53 (2.87)
Work w/ Coach in C	0.46 (1.88)	3.99 (2.59)
Work w/ Coach in D	-3.53(2.87)	
Constant	-4.46** (1.65)	-2.64** (1.36)
N	109	109

*p< 0.1, **p<0.05.

Understanding Findings: Work with a Coach and Task Selection

One significant finding that required further investigation was the negative relationship between teachers' selection of high-level tasks and their reported work with a coach. Given the differences in coaching designs across the four districts, I first investigated differences in the relationship between working with a coach and teachers' selection of high-level tasks by district. Results of several models that include statistical interactions between district membership and working with a coach suggest that there are not statistically significant differences between districts (see Table 9). Results of the model with District D as the reference district suggest that the odds of a teacher in District D who reports approximately 30 interactions with a coach selecting a low-level task are about 270 times the odds of a teacher in District D who reports only 14 interactions with a coach (see column (2) of Table 9, OR=1/270, $p<.05$). While there are not significant differences by district, the coefficients on the interactions between dummy variables for Districts A, B, and C, and work with a coach tend in the direction of a decrease in the magnitude of the coefficient, with larger standard errors. In other words, the negative relationship between working with a coach and improvement in selecting high-level tasks is most profound in District D and there is considerable variation in Districts A, B, and C. For none of the districts is the magnitude of the interaction coefficient large enough to offset the negative relationship between working with a coach and task selection in District D. Overall, the results suggest that teachers who reported more interactions with a coach were less likely to select a high-level task, but that the odds vary by school district. This result is consistent with the coaching designs in Districts A, C, and D where coaches were expected to work with the weakest teachers.

Understanding Findings: Advice-Seeking Interactions and Task Selection

The second set of findings that required additional investigation pertains to the relationship between teachers' advice-seeking interactions and their selection of high-level tasks. Recall that there is evidence of a positive relationship between teachers' reported number of advice seeking interactions within schools and their selection of high-level tasks (see Table 7). Further, the significant relationships persist even when the expertise of the colleagues from whom they seek advice is considered. In particular, I found that the expertise of colleagues from whom teachers sought advice (as measured by their VHQMI and task selection), weighted by the frequency of those interactions, is positively and significantly related to their selection of high-level tasks (see Table 8). However, the effect size was greatest for the number of advice-seeking interactions, regardless of expertise. These findings suggest that on average, colleagues' MKT, VHQMI, and task selection expertise might not be as critical as the number of advice-seeking interactions that teachers' have with colleagues.

This result and the fact that the coefficient for teachers' advice-seeking interactions outside of their schools is large, positive, and nearly significant in most of the task selection models (see Tables 7 and 8) raises the question of whether there are differences in the influence of interactions within schools and outside of schools. It is also not clear whether the influence of a large number of advice-seeking interactions is due to interacting with more people or to interacting with the same number of people more frequently. I investigated these questions by estimating several different models. First, I combined the number of advice-seeking interactions within schools with the number of advice-seeking interactions outside of schools. The mean for this variable for the primary task selection sample is 89 interactions and the standard deviation is 123.5 interactions. Recall that the mean for advice-seeking interactions within schools for this sample is 85.8 interactions and the standard deviation is 122.0 interactions. The results for the original model for task selection from Table 7, with advice-seeking interactions within the school and outside the school separated, are given in column (1) of Table 10 for reference. In the original model, the number of advice-seeking interactions within the school is significantly and positively related to teachers' selection of high-level tasks (OR=3.06, p<.05). The results for the model with advice-seeking interactions within the school and outside the school combined are given in column (2) of Table 10. These results suggest that the combined number of advice-seeking interactions is similarly related to teachers' selection of high-level tasks (OR=2.97, p<.05). Further, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) are similar for these two models. The AIC is slightly lower for the original model but the BIC is lower for the combined model. Therefore, neither of these models is necessarily a better fit. This result suggests that the distinction between interactions within school and outside of school is not critical; instead it might be the number of advice-seeking interactions in either setting that is important.

The next model investigated the question of whether the significance of advice-seeking interactions was due to the number of people that teachers turned to for advice or the frequency of the interactions with those people. I investigated this question by including a variable for the number of people who teachers turned to for advice (# Advice Givers) and a variable for the average number of interactions across those people (Avg. Freq. Advice). Results from this model suggest that the statistical significance of the number of advice-seeking interactions is due to the average frequency of advice-seeking interactions (OR=2.56, p<.05) rather than the number of people teachers turn to for advice. For teachers in the primary sample for task selection, the average frequency of advice-seeking interactions is about 49 (SD=60.3), which is more than once a week during the school year. The results from the model suggest that for teachers who averaged 109 advice-seeking interactions, or sought advice about 3 times a week, their odds of selecting a high-level task are 2.56 times the odds of teachers who average 49 advice-seeking interactions. The relatively higher AIC and BIC for this model suggest that modeling the number of advice givers and the average frequency of the advice is not as good of a fit as combining them into one measure of the number of advice-seeking interactions. However, it seems that the frequency of advice-seeking interactions is related to teachers' improvement in their selection of high-level tasks.

Lastly, the large and marginally significant effect for advice-seeking interactions outside the school is notable (OR=21.98, $p<.10$). What is notable is that the coefficient and standard error for advice-seeking interactions outside the school in the original model are both large. This means that for some teachers, there was a strong positive relationship between the number of advice-seeking interactions outside of school and their improvement in task selection, but for other teachers, that relationship did not hold. Given the size of this sample it is difficult to investigate district or other variation in that relationship, but it is worthy of future investigation. For example, it would be good to know whether it was advice-seeking interactions outside of teachers' schools with particular people (e.g., district math leaders) or in particular districts that supported teachers' improvement.

Table 10

Supplemental Analyses Pertaining to Advice-Seeking Interactions

Model	(1)		(2)		(3)	
High Potential	Coef (SE)	Odds Ratio	Coef (SE)	Odds Ratio	Coef (SE)	Odds Ratio
Year	2.75**	15.64	2.71**	15.06	2.61**	13.66
	(0.64)		(0.61)		(0.59)	
District A	1.07	2.92	0.83	2.30	1.04	2.82
	(1.00)		(0.95)		(0.98)	
District C	-0.98	0.38	-0.58	0.56	-0.41	0.66
	(1.21)		(1.17)		(1.19)	
District D	2.14**	8.50	1.82*	6.19	2.04*	7.72
	(1.05)		(1.03)		(1.06)	
Task PD	-0.17	0.84	-0.13	0.88	-0.12	0.88
	(0.33)		(0.33)		(0.33)	
TCT	-0.08	0.92	-0.33	0.72	-0.25	0.78
	(0.61)		(0.58)		(0.58)	
Advice-In	1.12**	3.06				
	(0.43)					
Advice-Out	3.09*	21.98				
	(1.76)					
Advice-Any			1.09**	2.97		
			(0.43)			
Avg. Freq. Advice					0.94** (0.38)	2.56
# Advice Givers					0.05 (0.25)	1.05
Work w/ Coach	-2.26** (0.69)	0.10	-2.33** (0.72)	0.10	-2.38** (0.75)	0.09
Constant	-4.08**		-4.89**		-5.06**	
	(1.20)		(1.22)		(1.26)	
N	109		109		109	
AIC	89.5		89.9		91.9	
BIC	121.8		119.5		124.2	

*p< 0.1, **p<0.05.

Patterns of Change in Maintenance of the Cognitive Demand of High-Level Tasks

The second aspect of the enactment of CDTs that I investigated is maintenance of the cognitive demand of high-level tasks. To understand change in maintenance of the cognitive demand of high-level tasks, I first classified teachers by their patterns of change across the years in which they selected high-level tasks. Table 11 summarizes the distribution of teachers by pattern of change in maintenance of the cognitive demand and number of observations. Recall that this set of analyses is limited to teachers who selected a high-level task. There are 153 teachers who posed at least one high-level task, with 249 total observations.

Similar to the analysis of task selection, my primary sample for maintenance of the cognitive demand is the set of teachers who initially did not maintain the cognitive demand of the high-level task they posed. There are two groups that make up the primary maintenance sample of 90 teachers who initially decreased the cognitive demand of the high-level task they posed: the 19 teachers classified as "Decrease then Maintain" and the 71 teachers classified as "All Decrease." Note that this sample of 90 teachers includes 48 teachers who posed a high-level task only one year, 23 teachers who continued to decrease the cognitive demand of the high-level tasks they posed in subsequent years, and 19 teachers who maintained the cognitive demand of the high-level task they posed at least one subsequent year.

Table 11

Teachers and Number of Observations for Maintenance of the Cognitive Demand of High-Level Tasks, by Change Pattern

Trend	1 observation	2 observations	3 observations	Total
All Maintain	33	5	9	47
Mixed	---	---	7	7
Maintain then Decrease	---	6	3	9
Decrease then Maintain	---	16	3	19
All Decrease	48	21	2	71
Total	81	48	24	153

As described in Table 11, 47 of the 153 teachers maintained the cognitive demand in every year in which they selected a high-level task, and they made up the "All Maintain" group. Only 14 of these 47 teachers posed high-level tasks and maintained the cognitive demand of those tasks in multiple years. There are 9 teachers who initially maintained the cognitive demand of the high-level task but later decreased the cognitive demand of a high-level task, and they made up the "Maintain then Decrease" group. Further, there are 7 teachers who fluctuated across years between maintaining and decreasing the cognitive demand of the high-level tasks they posed: 3 teachers maintained the cognitive demand, then decreased, and then maintained again, and 4 teachers initially decreased, then maintained, and then decreased again. These teachers comprise the "Mixed" group. The 9 teachers classified as "Maintain then Decrease" and the 7 teachers classified as "Mixed" are excluded from the analyses of interactions or expertise within interactions because they represent decline and inconsistency which are not the foci of this study. The teachers within the "All Maintain" category are an interesting comparison group because they are relatively accomplished in enacting CDTs.

Group Differences for Patterns of Change in Maintenance of the Cognitive Demand

Similar to the analysis of task selection, I first examined whether there were any differences between the primary sample for maintenance of the cognitive demand and the other teachers. I first inspected the distribution of teachers across the four school districts (see Table 12) for three categories of teachers: the primary sample, teachers who always maintained the cognitive demand of a high-level task (i.e., "All Maintain"), and the other teachers (i.e., "Mixed" and "Maintain then Decrease"). First, overall, there are fewer teachers from District C than from their other districts. Recall that this sample is related to the Task Selection sample because it only includes teachers who selected high-level tasks and there were fewer teachers in District C who selected a high-level task. Hence, it makes sense that it includes only 23 teachers from District C. Second, a higher percentage of teachers from District A fall into the "All Maintain" group than in the other districts, which suggests that teachers in District A are generally more likely to consistently maintain the cognitive demand of a high-level task. Relatedly, 60 of the 90 teachers in the primary sample for maintenance are from Districts B or D.

Table 12

Distribution of Teachers across three Categories of Maintenance Change Patterns and School District

District	Primary Sample		All Maintain		Other		Totals
	#	% of District Total	#	% of District Total	#	% of District Total	
A	14	38.9	15	41.7	7	19.4	36
B	29	59.2	15	30.6	5	10.2	49
C	16	69.6	4	17.4	3	13.0	23
D	31	68.9	13	28.9	1	2.5	45

Table 13

Descriptive Statistics for Interactions with Colleagues and Expertise within those Interactions for the Primary Maintenance Sample and the "All Maintain" Change Pattern

	Primary Maintenance Sample			All Maintain		
	Mean (SD)	SD	N	Mean (SD)	SD	N
Work w/ Coach**	15.0	15.9	135	10.8	15.6	90
Task PD	2.1	0.8	135	2.1	0.9	90
TCT**	6.6	5.3	135	3.9	3.8	91
Days In**	77.3	120.9	135	51.9	92.5	91
Days Out*	11.5	39.0	135	21.4	52.9	91
Coach MKT*	0.03	0.83	110	0.27	0.77	61
Coach VHQMI	2.47	0.68	110	2.56	0.68	61
TCT MKT	0.66	0.60	135	0.66	0.68	91
TCT VHQMI*	2.72	0.57	135	2.88	0.48	91
TCT Maintain	0.16	0.30	135	0.18	0.36	91
MKT Advice	57.71	131.54	135	56.61	126.32	91
VHQMI Advice	81.07	188.24	135	63.94	141.51	91
Maintain Advice	11.39	43.93	135	7.78	22.55	91

*p< 0.1, **p<0.05.

To begin to understand how participation in interactions might be related to change in maintenance of the cognitive demand, I compared means and standard deviations of interactions with colleagues (i.e., interactions variables) and expertise available in those interactions (i.e., expertise variables) for the primary maintenance sample and the “All Maintain” group (see Table 13). T-tests and regression models for expertise suggest that there are significant differences between the two groups for work with a coach, collaborative teacher meetings and the number of advice-seeking interactions inside the school ($p<.05$), and marginally significant differences for the number of advice-seeking interactions outside the school, for coach MKT, and for colleagues’ VHQMI in collaborative teacher meetings ($p<.10$). For the three significant interactions variables (i.e., working with a coach, collaborative teacher meetings, and within-school advice-seeking interactions), teachers in the primary maintenance sample reported more interactions than teachers in the “All Maintain” group. In contrast, teachers in the primary maintenance sample reported fewer advice-seeking interactions outside of their schools than teachers in the “All Maintain” group. These group differences in reported advice-seeking interactions within and outside of schools are similar to those for Task Selection. Lastly, for the two expertise variables (i.e., coach MKT and colleagues’ VHQMI in collaborative teacher meetings), access to expertise is higher for teachers in the “All Maintain” group.

I next investigated differences between the primary sample and the “All Maintain” group within the four school districts (see Table 14). There are several notable differences between the two groups within districts. First, while there was no significant overall difference between teachers’ reports of the extent to which PD emphasized challenging, problem-solving tasks, there are marginally significant differences for District A and District C. In particular, the task PD mean for teachers in the “All Maintain” group in District A is higher than the task PD mean for teachers in the primary maintenance sample ($t=1.30$, $p<.10$). In contrast, in District C, the task PD mean for teachers in the “All Maintain” group is lower than the task PD mean for teachers in the primary sample ($t=-1.65$, $p<.10$). This difference could explain why there is no significant overall difference between the primary sample and the “All Maintain” group for professional development on tasks. Further, it might suggest differences in the quality of professional development between District A and District C.

Another set of notable differences pertain to MKT expertise in District C. First, teachers in the “All Maintain” group worked with coaches with more developed MKT (b=0.74, SE=0.35, p<.10). Second, teachers in the primary maintenance sample had greater access to MKT expertise in collaborative teacher meetings (b=-0.66, SE=0.25, p<.05). Lastly, teachers in the “All Maintain” group had greater access to MKT expertise through advice-seeking interactions in their schools (b=110.5, SE=58.4, p<.10). In sum, the “All Maintain” group had greater available coach MKT and MKT within advice-seeking interactions however the primary sample had greater available MKT within collaborative teacher meetings. It is possible that these significant findings are attributable to the small samples of District C teachers in this comparison. In other words, with just four teachers in one sample, a statistical trend for particular variables (e.g., coach MKT), suggests that those four teachers had coaches with higher MKT, but that result might not generalize to a larger population of teachers.

In sum, there are significant differences between the primary maintenance sample and the “All Maintain” group with regard to working with a coach, time in collaborative teacher meetings, advice-seeking interactions within schools, coach MKT, and colleagues’ VHQMI in collaborative teacher meetings. In general, the weaker teachers (i.e., the primary sample) reported more interactions but had less access to expertise. Further, there are differences within districts pertaining to the extent that PD focused on challenging, problem-solving tasks. Lastly, there are significant differences in MKT expertise available in different interaction settings for the two groups of teachers in District C. The next step in the analysis was to investigate whether interactions with colleagues in different settings and the available expertise within those interactions were related to improvement in maintenance of the cognitive demand of high-level tasks.

Table 14

Descriptive Statistics for Interactions and Expertise for the Primary Maintenance Sample and the "All Maintain" Change Pattern, by District

	District A				District B				District C				District D			
	P. Sample		All Maintain		P. Sample		All Maintain		P. Sample		All Maintain		P. Sample		All Maintain	
	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N
Work w/ Coach	13.2** (23.9)	21	4.2 (8.6)	34	16.9** (17.7)	47	10.2 (12.0)	26	10.0 (12.3)	22	9.0 (8.4)	8	18.4 (12.4)	45	21.8 (22.8)	22
Task PD	2.2 (0.7)	21	2.4* (0.7)	34	2.0 (0.9)	47	1.7 (1.0)	26	2.5* (0.9)	22	1.9 (0.8)	8	2.0 (0.88)	45	2.2 (1.1)	22
TCT	2.0 (2.0)	21	2.2 (1.8)	35	6.6** (3.4)	47	4.2 (2.9)	26	13.5 (2.7)	22	12.9 (2.5)	8	4.2 (4.6)	45	3.1 (2.8)	22
Advice In	36.3 (46.8)	21	32.5 (68.6)	35	94.6 (151.0)	47	60.5 (84.3)	26	112.2 (148.3)	22	103.8 (171.3)	8	61.2 (84.8)	45	53.8 (95.3)	22
Advice Out	21.5 (53.9)	21	33.9 (65.7)	35	19.6 (53.3)	47	11.1 (35.8)	26	2.5 (8.0)	22	0	8	2.8 (8.5)	45	21.3** (54.0)	22
Coach MKT	0.63 (0.80)	4	0.69 (1.19)	9	-0.10 (0.58)	47	-0.07 (0.52)	26	-1.12 (0.82)	14	0.07* (0.27)	4	0.46 (0.67)	45	0.53 (0.65)	22
Coach VHQMI	3.05 (0.72)	4	2.97 (0.59)	9	2.28 (0.65)	47	2.23 (0.69)	26	2.39 (0.62)	14	2.65 (0.37)	4	2.66 (0.64)	45	2.78 (0.58)	22
TCT MKT	0.90 (0.75)	21	0.94 (0.71)	35	0.53 (0.56)	47	0.56 (0.63)	26	0.74** (0.62)	22	0.07 (0.53)	8	0.64 (0.52)	45	0.54 (0.57)	22
TCT VHQMI	2.80 (0.68)	21	2.97 (0.31)	24	2.70 (0.47)	47	2.73 (0.52)	26	2.44 (0.70)	22	2.54 (0.47)	8	2.86 (0.54)	45	3.08 (0.50)	22
TCT Maintain	0.22 (0.38)	21	0.26 (0.43)	35	0.18 (0.32)	47	0.11 (0.24)	26	0.04 (0.13)	22	0	8	0.16 (0.31)	45	0.20 (0.39)	22
MKT Advice	34.6 (67.3)	21	57.0 (112.6)	35	92.9 (192.2)	47	38.5 (83.6)	26	32.6 (55.6)	22	143.0* (265.9)	8	44.1 (92.0)	45	46.1 (112.3)	22
VHQMI Advice	43.5 (77.7)	21	55.9 (108.8)	35	134.9 (275.2)	47	51.3 (103.7)	26	80.2 (146.5)	22	169.3 (313.1)	8	42.8 (102.4)	45	53.3 (128.6)	22
Maintain Advice	7.4 (18.3)	21	11.6 (17.4)	35	21.6 (67.2)	47	2.5 (9.3)	26	9.8 (38.8)	22	22.5 (63.6)	8	3.3 (9.5)	45	2.5 (8.0)	22

*p< 0.1, **p<0.05.

Models of Interactions and Change in Maintenance of the Cognitive Demand

The analyses I conducted to investigate relationships between types of interactions and changes in maintenance of the cognitive demand draw on the primary maintenance sample described above which consists of 90 teachers in 26 different schools who initially decreased the cognitive demand of the high-level tasks they selected. Nineteen of those teachers improved in subsequent years (i.e., eventually maintained the cognitive demand), 23 teachers continued to decrease the cognitive demand of the tasks, and 48 teachers were only in the sample of teachers who selected a high-level task for one of Years 2, 3, or 4 and decreased the cognitive demand of the high-level task they selected in that year.

Results from the models of time spent interacting with colleagues and the expertise of colleagues within those interactions on change in maintenance of the cognitive demand are given in Tables 15 and 16. As can be seen, very few of the variables of interest are significantly related to changes in teachers' maintenance of the cognitive demand. Results from model (1) and (3) suggest that the number of interactions with a coach is significant and negatively related to the maintenance of the cognitive demand ($p<.10$). According to model (1), teachers who worked more with a coach were less likely to maintain the cognitive demand of the high-level task they posed ($OR=0.51$, $p<.10$). There is one other marginally significant relationship pertaining to coach expertise. Results from model (3) suggest that although working with a coach is negatively related to maintenance of the cognitive demand, if the coach's MKT was sufficiently developed, the direction of the result changes ($p<.10$). Given the magnitude of the coefficients ($b=-1.45$ and $b=0.97$, for working with a coach and the interaction between coach MKT and working with a coach, respectively), it would require working with a coach who had an MKT score two standard deviations above the mean to reverse the generally negative relationship associated with working with a coach. None of the other variables are significantly related to maintenance of the cognitive demand of high-level tasks. Potential reasons for this lack of significant results are discussed below.

Table 15

Models of Interactions with Colleagues and Expertise within those Interactions on Improvement in Maintenance of the Cognitive Demand: Participation in Interactions and MKT Expertise

	(1) Interactions with Colleagues		(2) Coach MKT	(3) Coach MKT 2	(4) MKT	(5) MKT 2
Maintenance	Coef (SE)	Odds Ratio	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
Year	1.27**	3.56	1.22**	1.43**	1.24**	1.28**
	(0.43)		(0.49)	(0.54)	(0.44)	(0.44)
District A	0.55	1.73	0.75	0.52	0.63	0.78
	(0.89)		(1.41)	(1.51)	(0.91)	(0.95)
District C	-0.23	0.79	0.43	0.02	-0.56	-0.43
	(1.10)		(1.49)	(1.59)	(1.14)	(1.14)
District D	1.12	3.06	0.96	1.22	1.11	1.20
	(0.70)		(0.76)	(0.83)	(0.72)	(0.74)
Task PD	0.13	1.14	0.02	0.09	0.15	0.15
	(0.34)		(0.36)	(0.38)	(0.34)	(0.34)
TCT	0.07	1.07	-0.16	-0.08	0.20	0.22
	(0.37)		(0.42)	(0.41)	(0.40)	(0.40)
TCT Expertise					0.37	0.29
					(0.30)	(0.31)
TCT*Exp						-0.20
						(0.33)
Advice-In	-0.36	0.70	-0.24	-0.20		
	(0.30)		(0.49)	(0.53)		
Advice-In Expertise					-2.67	-2.74
					(1.94)	(1.98)
Advice-Out	-0.30	0.74	-0.20	-0.32	-0.37	-0.38
	(0.31)		(0.32)	(0.35)	(0.30)	(0.32)
Work with Coach	-0.68*	0.51	-0.73	-1.45**	-0.64	-0.66
	(0.40)		(0.44)	(0.73)	(0.41)	(0.42)
Coach Expertise			0.19	0.34		
			(0.41)	(0.46)		
WWC*Exp				0.97*		
				(0.57)		
Constant	-4.18**		-3.91**	-4.82**	-4.89**	-5.05**
	(1.12)		(1.18)	(1.50)	(1.37)	(1.42)
N	135		110	110	135	135

*p< 0.1, **p<0.05.

Table 16

Models of Interactions and Expertise on Improvement in Maintenance of the Cognitive Demand: VHQMI and Maintenance of the Cognitive Demand Expertise

	(6) Coach VHQMI	(7) Coach VHQMI 2	(8) VHQMI	(9) VHQMI 2	(10) Maintain	(11) Maintain 2
Maintenance	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)	Coef (SE)
Year	1.13**	1.18**	1.09**	1.06**	1.54**	1.53**
	(0.51)	(0.52)	(0.53)	(0.52)	(0.50)	(0.45)
District A	0.61	0.63	0.75	0.77	0.91	0.87
	(1.45)	(1.46)	(0.91)	(0.91)	(0.95)	(0.96)
District C	0.16	0.12	-0.18	-0.28	-0.07	-0.09
	(1.40)	(1.41)	(1.13)	(1.15)	(1.16)	(1.15)
District D	0.88	0.91	0.78	0.71	1.01	0.98
	(0.77)	(0.78)	(0.77)	(0.77)	(0.76)	(0.74)
Task PD	0.02	0.02	0.12	0.08	0.15	0.16
	(0.36)	(0.37)	(0.34)	(0.35)	(0.35)	(0.35)
TCT	-0.17	-0.16	0.19	0.23	0.03	0.01
	(0.42)	(0.43)	(0.40)	(0.41)	(0.35)	(0.40)
TCT Expertise			0.39	0.45	0.45	0.43
			(0.43)	(0.42)	(0.28)	(0.29)
TCT*Exp				-0.25		-0.09
				(0.25)		(0.42)
Advice-In	-0.31	-0.25				
	(0.52)	(0.52)				
Advice-In Expertise			-2.69 (1.76)	-2.67 (1.75)	-5.44 (4.11)	-5.43 (4.11)
Advice-Out	-0.21	-0.27	-0.39	-0.38	-0.46	-0.45
	(0.33)	(0.35)	(0.32)	(0.32)	(0.35)	(0.33)
Work with Coach	-0.72 (0.44)	-1.06 (0.64)	-0.53 (0.41)	-0.52 (0.41)	-0.69 (0.42)	-0.67 (0.41)
Coach Expertise	0.32 (0.45)	0.44 (0.47)				
WWC*Exp		0.57				
		(0.56)				
Constant	-3.76**	-3.96**	-4.64**	-4.55**	-5.72**	-5.70**
	(1.22)	(1.31)	(1.38)	(1.37)	(1.75)	(1.60)
N	110	110	135	135	135	135

*p< 0.1, **p<0.05.

Limitations of the Study

Before turning to a discussion of my findings, I acknowledge several limitations of this study. First, my assessment of teachers' enactment of CDTs, and, hence, change in their enactment of CDTs is based on two class periods of instruction for each year. I take the better of the two class periods to represent teachers' best shot at enacting CDTs rather than representing their typical enactment of CDTs. It is possible that more information about how their best shot relates to their typical instruction might shed additional light on their improvement in enacting CDTs.

Second, while the types of interactions I focus on in this study involved activities close to teachers' practice (e.g., lesson planning, advice pertaining to mathematics teaching), I generally do not know if the activities that teachers engaged in during interactions pertained to the enactment of CDTs. Therefore, it is possible that within certain types of interactions teachers did not actually work on the enactment of CDTs, which could explain why interactions with colleagues were not significantly related to change in the enactment of CDTs in some settings. It is methodologically challenging to collect information on the specific activities that take place within interactions across a large sample of teachers. Therefore, we need more small-scale studies of how work on the enactment of CDTs within interactions in different settings influences teachers' enactment of CDTs. Also, future research should investigate ways to collect information about the content of interactions for large samples of teachers.

Finally, two other limitations of this study might explain the general lack of significant findings pertaining to expertise. First, the data on expertise was limited to a sample of 3 to 6 teachers per school. In some schools this sample consisted of all of the math teachers, but in other schools it was as little as a third of the math teachers. Therefore, it is likely that I have underestimated colleagues' expertise within collaborative teacher meetings for some schools. Second, I used multiple imputation based on existing scores and other information about participants to create expertise scores for teachers or coaches with missing scores, which might have introduced additional error into the measures of expertise. Multiple imputation was particularly necessary because of my use of colleagues' expertise scores from the prior year, which meant there were even more missing values. While this is an accepted method for accounting for missing data, it has the potential to introduce extra error into the measures of expertise, which is especially problematic with a small sample. It is therefore possible that the lack precision of the measures of expertise might have contributed to a lack of significant findings pertaining to colleagues' expertise.

Discussion and Conclusion

The enactment of CDTs is critically important in supporting students' development of conceptual understanding in mathematics (Stein & Lane, 1996; Zohar & Dori, 2003). Further, there is evidence that teachers in U.S. classrooms rarely enact CDTs, and that they struggle when they attempt to do so (Hiebert et al., 2005; Stein et al., 1996). Therefore, we need to learn more about productive supports for mathematics teachers' learning to enact CDTs. Prior research on professional learning suggests that ongoing interactions with relatively accomplished colleagues, involving activities that are close to practice might support teachers' development (Bruner, 1996; Lave & Wenger, 1991). Work with a coach, collaborative teacher meetings, advice-seeking interactions, and professional development each have the potential to meet those criteria. In this study, I investigated whether change in teachers' enactment of CDTs over time was related to interaction with colleagues in those settings, with special attention to the available expertise.

Despite the fact that this analysis involved middle school mathematics teachers from four school districts that were attempting to support teachers to improve their practice, there was not widespread change in teachers' enactment of CDTs. This is further confirmation that changing teachers' practice is difficult and complicated, even when improvement is a district focus (Coburn, 2001; Coburn & Stein, 2006; D. K. Cohen, 2011; Matsumura, Sartoris, Bickel, & Garnier, 2009). Yet, over one-half of the teachers did improve in some aspects of the enactment of CDTs. Given the lack of widespread change, identifying types of interactions that were related to the improvements that some teachers made is particularly important. Unfortunately, in this analysis, interactions in most settings were not significantly related to improvement in teachers' enactment of CDTs, although there were a few significant relationships that I discuss below.

With regard to change in teachers' enactment of CDTs over three years, there was slightly more change in teachers' task selection than in their maintenance of the cognitive demand of high-level tasks. In particular, approximately 61% of teachers who initially posed a low level task eventually posed a high-level task, whereas 45% of teachers who initially decreased the cognitive demand of a high-level task they posed eventually maintained the cognitive demand. There is good reason to believe that selecting a high-level task is not as challenging as maintaining the cognitive demand of a high-level task because maintaining the cognitive demand of a high-level task goes beyond choosing from materials to managing interactions in the classroom (Stein et al., 1996). Further, findings from Paper 2 of this dissertation suggest that teachers' knowledge and beliefs are related to their task selection and maintenance of the cognitive demand in different ways, which implies that perhaps different supports are appropriate for task selection and for maintenance of the cognitive demand of high-level tasks.

The findings of this analysis shed some light on how interactions with colleagues might support teachers' enactment of CDTs. First, results from models of task selection and maintenance of the cognitive demand suggest that work with a coach was significantly and negatively related to teachers' development in enacting CDTs except in the case when a coach had very sophisticated mathematical knowledge for teaching. When a coach had very sophisticated mathematical knowledge for teaching (i.e., two standard deviations above the mean), working with a coach supported teachers' development in maintaining the cognitive demand.

One possible explanation for the negative relationship between working with a coach and the enactment of CDTs is a selection bias in coach-teacher matching: coaches tended to work more with teachers who were weaker instructionally. This interpretation is consistent with several other pieces of evidence: 1) the district designs for coaching in Districts A, C, and D expected coaches to work with the neediest teachers, and 2) the significantly higher mean number of interactions with a coach for teachers in the primary maintenance sample compared with teachers in the "All Maintain" group. Although this interpretation could explain why the teachers who did not improve worked more with a coach, it does not explain why those teachers still did not seem to improve.

A possible reason why the teachers did not seem to improve as a result of working with the coach is that, on average, the MKT and VHQMI of coaches in this sample were not more developed than the teachers they were expected to support. The fact that coaches were generally not relatively accomplished might have limited their ability to support teachers' development. The marginally significant result suggesting that teachers who worked with coaches whose MKT was well developed were more likely to maintain the cognitive demand of the high-level tasks gives some credence to this interpretation. In sum, these findings on coaching suggest that, in general, coaches worked with weaker teachers, and that only those teachers who worked with relatively accomplished coaches improved.

Another key finding from this study is that teachers' advice-seeking interactions were positively related to the selection of high-level tasks. In other words, teachers who reported seeking advice more often were more likely to select high-level tasks. Further, it appears to be the frequency of interactions rather than the number of people with whom teachers interact that is important. One possible interpretation is that the teachers who seek advice more often are actually learning to select high-level tasks through their advice-seeking interactions. Alternatively, it is also possible that the teachers who seek advice about math instruction more frequently have a stronger desire to improve than those who seek advice less often. Future research should investigate this result further by taking account of teachers' propensities toward change while examining the effect of advice-seeking interactions on their enactment of CDTs.

Even if it is the case that teachers learn to enact CDTs from their advice-seeking interactions, the policy implications are challenging. Teachers' advice-seeking networks are emergent and cannot be mandated (Smylie & Evans, 2006; James P Spillane, Reiser, & Gomez, 2006). However, there are some indications that district and school policies can influence these advice-seeking interactions. Most of the research investigating the formation of teacher networks has focused on the characteristics of individuals (e.g., teachers and colleagues), but some recent studies have begun to investigate how district policy and other aspects of school context might influence teacher networks (Coburn, Choi, & Mata, 2010; Coburn & Russell, 2008). In particular, there is evidence that district policy that creates structures for interaction with a particular focus (e.g., mathematics curriculum implementation) can influence teacher networks

(Coburn et al., 2010). Further, aspects of the school context can have an impact on: (1) the formation of teacher networks (Gibbons, Garrison, & Cobb, 2011), and (2) the influence of district policies on the formation of teacher networks (Atteberry & Bryk, 2010). More research is needed on how and under what conditions particular support structures might influence teachers' advice-seeking behavior.

Overall, there is little indication from this sample that teachers' learning through interactions was moderated by their colleagues' mathematical knowledge for teaching, vision of high-quality mathematics instruction, and sophistication of their instructional practices. Further, colleagues' expertise was generally not significantly related to change in teachers' enactment of CDTs. These findings contradict a number of prior studies suggesting the importance of colleagues' expertise within interactions (e.g., Frank et al., 2004; Gibbons, 2012; Horn & Little, 2010; Penuel et al., 2009). There are several possible explanations for the contradictory findings. First, as I described above, the sampling of teachers within schools and the use of multiple imputation may have contributed to a lack of significant findings pertaining to expertise. Second, it might be that colleagues' expertise was generally not developed enough to support teachers in improving their instructional practices. The marginally significant and positive moderating relationship of coach MKT on work with a coach gives some credence to this interpretation. Third, it is possible that these forms of colleagues' expertise are not as relevant to the enactment of CDTs as other forms of expertise (e.g., knowledge of the curriculum materials). Lastly, as described in the limitations above, perhaps teachers did not actually work on the enactment of CDTs during interactions. If that was the case, we would not necessarily expect there to be changes in teachers' enactment of CDTs, regardless of the available expertise. In future research, it will be important to consider specific forms of expertise, in the context of specific activities in which teachers engage with others during interactions, as they relate to the development of associated practices.

The findings of several prior studies indicate that time allocated for collaboration is not sufficient for supporting teachers' development of specific practices (Horn & Little, 2010; Kruse & Louis, 1995). The results of this study confirm those findings for collaborative teacher meetings and extend the findings to work with a coach and formal professional development. There is evidence that, in general, the mere occurrence of interactions—whether they are with a coach, in collaborative teacher meetings, or within formal professional development—is not sufficient to support teachers' development in enacting CDTs. In future research, it will be important to consider the actual activities during interactions when investigating how they support teachers' development. However, the results from this study suggest that the occurrence of advice-seeking interactions both within and outside of schools might support teachers' improvement in enacting CDTs. Future research should investigate advice-seeking interactions to understand how they might support teachers' improvement of classroom practice. As schools and districts work to support teachers' development of inquiry-oriented instructional practices, it will be important to create structures that promote interactions with colleagues (e.g., mathematics coaches, collaborative teacher meetings) and be intentional about activities that will foster the desired development.

APPENDIX A

VISION OF HIGH QUALITY MATHEMATICS INSTRUCTION (VHQMI)

Interview Questions:

“If you were asked to observe a teacher's math classroom for one or more lessons, what would you look for to decide whether the mathematics instruction is high quality?”

> “Why do you think it is important to use/do _____ in a math classroom? Is there anything else you would look for? If so, what? Why?”

For each of these three topics the participants did not identify spontaneously, we prompted by asking, respectively,

1) What are some of the things that the teacher should actually be doing in the classroom for instruction to be of high quality?

2) What type of tasks do you think the teacher should be using for instruction to be of high quality?

3) Can you please describe what classroom discussion would look and sound like if instruction was of high quality?

VHQMI Rubric Categories include:

Role of the Teacher

Classroom Discourse (including: Patterns and Structure of Classroom Talk, Nature of Talk, Student Questions, Teacher Questions, Student Explanations)

Mathematical Tasks

Nature of Classroom Activity

Abbreviated versions of the VHQMI Rubrics are provided on the following pages. See Munter (Under review) for a more thorough explanation of the measure.

Level	Description	Potential ways of characterizing teacher's role
4) Teacher as 'more knowledgeable other'	Describes the role of the teacher as *proactively* supporting students' learning through co-participation. Stresses the importance of designing learning environments that support problematizing mathematical ideas, giving students mathematical authority, holding students accountable to others and to shared disciplinary norms, and providing students with relevant resources (Engle & Conant, 2002).	*Influencing classroom discourse*: Suggests that the teacher should purposefully intervene in classroom discussions to elicit & scaffold students' ideas, create a shared context, and maintain continuity over time (Staples, 2007).
		Attribution of mathematical authority: Suggests that the teacher should support students in sharing in authority (Lampert, 1990), problematizing content (Hiebert et al., 1996), working toward a shared goal (Hiebert et al, 1997), and ensuring that the responsibility for determining the validity of ideas resides with the classroom community (Simon, 1994).
		Conception of typical activity structure: Promotes a 'launch-explore-summarize' lesson (Lappan et al., 1998), in which a) the teacher poses a problem and ensures that all students understand the context and expectations (Jackson et al., in press), b) students develop strategies and solutions (typically in collaboration with each other), and c) through reflection and sharing, the teacher and students work together to explicate the mathematical concepts underlying the lesson's problem (Stigler & Hiebert, 1999).
3) Teacher as 'facilitator'	Focuses on the forms of "reform instruction" without a strong conception of the accompanying functions that underlie those forms: either (a) views the teacher's role as *passive*, as students discover new mathematical insights as the result of collaborative problem solving (e.g. "romantic constructivism"), or (b) describes a transitional view that incorporates both teacher demonstration or introduction (e.g., at the beginning of the lesson) and 'turning it over' to the students (who then make the remaining 'discoveries'). Description likely stresses 'rules' for structuring lessons, discussion, etc. or describes posing problems and asking students to describe their strategies but does not detail a proactive role in supporting students in engaging in genuine mathematical inquiry (Kazemi & Stipek, 2001).	*Influencing classroom discourse*: Describes the teacher facilitating student-to-student talk, but primarily in terms of students taking turns sharing their solutions; Hesitates to 'tell' too much for fear of interrupting the 'discovery' process (Lobato et al, 2005).
		Attribution of mathematical authority: Supports a 'no-tell policy': Stresses that students should figure things out for themselves and play a role in 'teaching.' Suggests that if students are pursuing an unfruitful path of inquiry or an inaccurate line of reasoning, the teacher should pose a question to help them find their mistake, but the reason for doing so focuses more on not telling than helping students develop mathematical authority. Is open to students developing their own mathematical problems, but these inquiries are not candidates for paths of classroom mathematical investigation.
		Conception of typical activity structure: Promotes a 'launch-explore-summarize' lesson (Lappan et al., 1998), in which a) the teacher poses a problem and possibly completes the first step or two with the class or demonstrates how to solve similar problems, b) students work (likely in groups) to complete the task(s), and c) students take turns sharing their solutions and strategies and/or the teacher clarifies the primary mathematical concept of the day (i.e., how they 'should have' solve the task).

Figure A1. Abbreviated VHQMI Rubric: Role of the Teacher

2) Teacher as 'monitor'	Describes the teacher as the primary source of knowledge, but stresses the importance of providing time for students to work together, to try on their own and make sense of what the teacher has demonstrated, to (first) explain things to each other, and then get help from the teacher.	*Influencing classroom discourse*: Suggests the teacher should promote student-student discussion in group work.
		Attribution of mathematical authority: Suggests a view of teacher as an "adjudicator of correctness" (Hiebert et al, 1997). Students may participate in 'teaching' but only as mediators of the teacher's instruction, adding clarification, etc. If students are pursuing an unfruitful path of inquiry or an inaccurate line of reasoning, the teacher stops them and sets them on a 'better' path.
		Conception of typical activity structure: Promotes a two phase, 'acquisition and application' lesson (Stigler & Hiebert, 1999), in which a) the teacher demonstrates or leads a discussion on how to solve a type of problem, and then b) students are expected to work together (or "teach each other") to use what has just been demonstrated to solve similar problems, while the teacher circulates throughout the classroom, providing assistance when needed.
1) Teacher as 'deliverer of knowledge'	Describes the teacher as the primary source of knowledge, focusing primarily on mathematical correctness and thoroughness of explanations (i.e., showing all steps). Description suggests that students are *welcome* to ask questions, but that there is no expectation that the teacher will facilitate student collaboration or discussion.	*Influencing classroom discourse*: Focuses exclusively on T→S discourse. Considers quality of teacher's explanations in terms of clarity and mathematical correctness.
		Attribution of mathematical authority: Suggests that the responsibility for determining the validity of ideas resides with the teacher or is ascribed to the textbook (Simon, 1994). (This includes insistence that teachers be mathematically knowledgeable and correct.)
		Conception of typical activity structure: Promotes efficiently structured lessons (in terms of coverage) in which the teacher directly teaches how to solve problems. Periods might include time for practice while teacher checks students' work and answers questions, but this is likely quiet & individually-based with no opportunity for whole-class discussion. Description suggests no qualms with exclusive lecture format.
0) Teacher as 'motivator'	Suggests that the teacher must first and foremost be sufficiently captivating to attract and hold students' attention.	

Figure A1, continued.

Level	Patterns/structure of Classroom Talk	Nature of Classroom Talk	Student Questions	Teacher Questions	Student Explanation
	Description	Description	Description	Description	Description
4	Promotes whole-class conversations, including student-to-student **talk that is student-initiated**, not dependent on the teacher (Hufferd-Ackles, Fuson, & Sherin, 2004); Promotes developing & supporting a "mathematical discourse community" (Lampert, 1990),	Suggests that classroom talk should be conceptually oriented—including articulating/refining conjectures and arguments for explaining mathematical phenomena—for the purpose of supporting students in 'doing mathematics' and/or spawning new investigations.	Values student questions that drive instruction, leading to new mathematical investigations, questions characteristic of 'doing mathematics' (e.g., generalization).	Describes the role of teacher questions that are conceptually oriented ('why' questions) in driving investigations, helping students explain their problem-solving strategies, and/or helping the teacher understand students' thinking (Borko, 2004)	Student explanations include both explanation and justification (Kazemi & Stipek, 2001) with little prompting from the teacher (Hufferd-Ackles, Fuson, & Sherin, 2004)
3	Promotes whole-class conversations (about ideas, not just whole-class lecture or task set-up), but description places the **teacher at the center of talk**, likely doing most of the prompting and pressing, or calling upon students/groups to take turns presenting their strategies.	Insists that the content of classroom talk be about mathematics (e.g., asking questions, providing explanations), but description of such talk either (a) characterizes talk that is of a calculational orientation; or (b) fails to specify expectations for the nature/quality of the questions, explanations, etc.	Values student questions in the math classroom, but description suggests that procedurally-oriented questions are adequate; possibly considers the occurrence of student questions primarily among groups of students (and not during whole-class instruction).	Either (a) stresses the importance of asking conceptually-oriented questions (and details such questions with more than 'catch-phrases' such as or 'higher-order') but does not elaborate on the function of such questions in progressing classroom discourse or understanding student thinking, or (b) suggests that the teacher's questions can serve such functions but describes questions of a calculational orientation ('how' questions)—which would not actually achieve the intended function.	Description suggests an emphasis on student explanations of strategies that have primarily a calculational (rather than conceptual) orientation (Thompson et al, 1994; Kazemi & Stipek, 2001) or are not characterized
2	Values student-student discourse but describes it exclusively in the context of **small group/partner work** (if there's mention of whole-class discussion, it's characterized only as an option, not a vital element)	Insists that the content of students' classroom talk (with each other) be about mathematics, but provides no description of content (i.e., does not specify things such as questions and explanations).	Emphasizes the *presence* of student questions in the math classroom; may consider students' questions as differentiable in quality, but provides no specific criteria	Names the quality of teacher questions as an important criterion, but either (a) provides no criteria for differentiating in quality, (b) uses only 'catch-phrases' (e.g., 'higher-order', 'extension') to describe the quality of questions, or (c) examples include probing for steps taken or questioning to determine whether (but not how) a student understands ('what/how' questions, but not 'why' questions).	
1	Describes traditional **lecturing and/or IRE** (Mehan, 1979), or IRF (Sinclair & Coulthard, 1975) dialogue patterns. (Note that this can occur in a 'whole-class' setting, but is not considered a genuine whole-class discussion.)		Does not value student questions, or suggests that students should be welcome to ask questions, but that the presence of student questions is not inherently a good aspect of classroom discourse.	Names the presence or quantity of teacher's questions as an important criterion, or describes a scenario where students offer one-word or short-phrase answers to questions the teacher asks as (s)he demonstrates, or suggests that the role of teacher's questions is to keep students on task.	

Figure A2. Abbreviated VHQMI Rubric: Classroom Discourse

Level
4) Emphasizes tasks that have the potential to engage students in "doing mathematics" (Stein, Grover, & Henningsen, 1996; Smith & Stein, 1998), allowing for "insights into the structure of mathematics" & "strategies or methods for solving problems" (Hiebert et al, 1997).
3) Emphasizes tasks that have the potential to engage students in complex thinking, including tasks that that allow multiple solution paths or provide opportunities for students to create meaning for mathematical concepts, procedures, and/or relationships. "Application" is characterized in terms of problem-solving. However, tasks described lack complexity, do not press for generalizations, do not emphasize making connections between strategies or representations, or require little explanation (Boston & Wolf, 2006). Instead, they emphasize connections to "the real world, or "prior knowledge." Reasons for multiple strategies are not tied to rich discussion or making connections between ideas.
2) Promotes 'reform'-oriented aspects of tasks without specifying the nature of tasks beyond broad characterizations (e.g., "hands-on," "real world connections," "higher order"), and without elaborating on their function in terms of providing opportunities for "doing mathematics" (Stein, Grover, & Henningsen, 1996; Smith & Stein, 1998). "Application" is characterized in terms of "real world" *context* and/or students being active.
1) Emphasizes tasks that provide students with opportunity to practice a procedure before then applying it conceptually to a problem (Hiebert et al, 1997)
0) (a) Does not view tasks as inherently higher- or lower-quality; or (b) Does not view tasks as a manipulable feature of classroom instruction

Figure A3. Abbreviated VHQMI Rubric: Mathematical Tasks

Level	Description
2) Specifies WHAT Ss should be doing using typical reform language, without describing the nature of classroom activity in content-specific ways--focuses primarily on the organization/structure of the activity (form view).	Describes what students should be doing without mention of the content of their interactions (i.e., describes a 'non-traditional' classroom, full of activity, but does not specify how the activity is specific to mathematics). If reasons WHY particular forms of activity are important are provided they are not in terms of supporting students' participation in doing mathematics.
1) Stresses the importance of students being engaged and "on-task", either taking for granted the *quality* of classroom activity (i.e., students should be doing whatever the teacher asked), or specifying traditional classroom activities as what should take place.	(a) Stresses THAT students should be engaged and participating in classroom activities (i.e., on-task, paying attention), without specifying WHAT those activities should be; OR, (b) Describes nature of classroom activity as traditional classroom activity.

Figure A4. Abbreviated VHQMI Rubric: Nature of Classroom Activity

APPENDIX B

ORIGINAL AND MULTIPLY IMPUTED EXPERTISE DATA FOR PAPER 3

Table B1

Comparing Expertise in Original and Imputed Sample

	Original Sample					Imputed Sample				
	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N
MKT	-0.06	0.81	-2.01	2.01	584	-0.18	0.95	-3.84	3.56	1808
VQHMI	2.44	0.63	0	4	816	2.42	0.72	-0.95	5.15	1808
High Potential	0.70	0.46	0	1	499	0.68	0.47	0	1	1808
Maintain	0.45	0.50	0	1	347	0.42	0.49	0	1	1229

REFERENCES

Aguirre, J., & Speer, N. M. (1999). Examining the Relationship Between Beliefs and Goals in Teacher Practice. *The Journal of Mathematical Behavior, 18*(3), 327-356.

Allison, P. D. (2005). *Causal inference with panel data.* Paper presented at the Amer. Sociol. Association Annual Meeting.

Anderson, J., White, P., & Sullivan, P. (2005). Using a schematic model to represent influences on, and relationships between, teachers' problem-solving beliefs and practices. *Mathematics Education Research Journal, 17*(2), 9-38. doi: 10.1007/bf03217414

Atteberry, A., & Bryk, A. S. (2010). Centrality, connection, and commitment: The role of social networks in a school-based literacy initiative. In A. J. Daly (Ed.), *Social Network Theory and Educational Change* (pp. 51-76). Cambridge, Massachusetts: Harvard Education Press.

Ball, D. L. (1991). Research on teaching mathematics: Making subject matter knowledge part of the equation. East Lansing, MI: Michigan State University.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession* (pp. 3-32). San Francisco: Jossey-Bass.

Barraugh, A. L. (2011). *Situated mathematics instruction: The interaction of teacher knowledge, beliefs, and policy interpretation.* (Doctor of Education), University of California, San Diego, San Diego, CA.

Barrett Paterson, V. U. (2009). *Teachers' beliefs and practices in relation to reform oriented mathematics teaching.* (Doctor of Philosophy), Temple University.

Barton, A. R. (2005). *The impact of a specialized mathematics professional development program on instructional teacher behaviors and student success.* (Doctor of Education), Texas A & M University- Kingsville.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., . . . Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal, 47*(1), 133-180.

Beswick, K. (2005). The beliefs/practice connection in broadly defined contexts. *Mathematics Education Research Journal, 17*(2), 39-68. doi: 10.1007/bf03217415

Bidwell, C. E., & Yasumoto, J. Y. (1999). The collegial focus: Teaching fields, collegial relationships, and instructional practice in american high schools. *Sociology of Education, 72*(4), 234-256.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education, 24*(2), 417-436. doi: http://dx.doi.org/10.1016/j.tate.2006.11.012

Boston, M., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education, 40*(2), 119-156.

Boston, M., & Wolf, M. K. (2006). Assessing academic rigor in mathematics instruction: The development of the instructional qualiy assessment toolkit. CSE Technical Report 672. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Brophy, J. (1986). Teaching and Learning Mathematics: Where Research Should Be Going. *Journal for Research in Mathematics Education, 17*(5), 323-346.

Bruce, C. D., & Ross, J. A. (2008). A Model for Increasing Reform Implementation and Teacher Efficacy: Teacher Peer Coaching in Grades 3 and 6 Mathematics. *Canadian Journal of Education / Revue canadienne de l'éducation, 31*(2), 346-370.

Bruner, J. (1996). *The Culture of Education*. Cambridge, Massachusetts: Harvard University Press.

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppesco, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal, 26*(4), 499-531.

Charalambous, C. Y. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal, 110*(3), 247-278.

Choppin, J. M. (2011). The impact of professional noticing on teachers' adaptations of challenging tasks. *Mathematical Thinking and Learning, 13*, 175-197.

Clarke, D. (2008). The mathematics teacher as curriculum maker. In P. Sullivan & T. Wood (Eds.), *Knowledge and Beliefs in Mathematics Teaching and Teaching Development* (Vol. 1, pp. 133-151). Rotterdam, The Netherlands: Sense Publishers.

Cobb, P., & Jackson, K. J. (2011). Towards an empricially grounded theory of action for improving the quality of mathematics teaching at scale. *Mathematics Teacher Education and Development, 13*(1), 6-33.

Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating Teachers' Instructional Practices in the Institutional Setting of the School and District. *Educational Researcher, 32*(6), 13-24.

Cobb, P., & Smith, T. M. (2008). District development as a means of improving mathematics teaching and learning at scale. In K. Krainer & T. Wood (Eds.), *Participants in Mathematics Teacher Education: Individuals, Teams, Communities, and Networks* (Vol. 3, pp. 231-254). Rotterdam, The Netherlands: Sense Publishers.

Coburn, C. E. (2001). Collective Sensemaking about Reading: How Teachers Mediate Reading Policy in Their Professional Communities. *Educational Evaluation and Policy Analysis, 23*(2), 145-170.

Coburn, C. E. (2005). Shaping Teacher Sensemaking: School Leaders and the Enactment of Reading Policy. *Educational Policy, 19*(3), 476-509.

Coburn, C. E., Choi, L., & Mata, W. (2010). "I would go to her because her mind is math": Network formation in the context of a district-based mathematics reform. In A. J. Daly (Ed.), *Social Network Theory and Educational Change* (pp. 33-50). Cambridge, Massachusetts: Harvard Education Press.

Coburn, C. E., & Russell, J. L. (2008). District Policy and Teacher's Social Networks. *Educational Evaluation and Policy Analysis, 30*(3), 203-235.

Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. In M. I. Honig (Ed.), *New Directions in Education Policy Implementation: Confronting Complexity* (pp. 25-46). Albany, NY: State University of New York Press.

Cohen, D. K. (1990). A Revolution in One Classroom: The Case of Mrs. Oublier. *Educational Evaluation and Policy Analysis, 12*(3), 311-329.

Cohen, D. K. (2011). *Teaching and its predicaments*: Harvard University Press.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Collopy, R. (2003). Curriculum materials as a professional development tool: How a mathematics textbook affected two teachers' learning. *The Elementary School Journal, 103*(3), 287-311.

Confrey, J. (1986). A critique of teacher effectiveness research in mathematics education. *Journal for Research in Mathematics Education, 17*(5), 347-360.

Cooney, T. J. (1985). A Beginning Teacher's View of Problem Solving. *Journal for Research in Mathematics Education, 16*(5), 324-336.

Cross, D. (2009). Alignment, cohesion, and change: Examining mathematics teachers' belief structures and their influence on instructional practices. *Journal of Mathematics Teacher Education, 12*(5), 325-346. doi: 10.1007/s10857-009-9120-5

Darling-Hammond, L. (2007). The flat earth and education: How America's commitment to equity will determine our future. *Educational Researcher, 36*(6), 318-334.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction: Results from a Three-Year Longitudinal Study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Diaz, D. G. P. (2004). *The role of standards-based curricula in teacher learning and instructional reform.* (Doctor of Philosophy Doctoral Disseration), Clemson University.

Doyle, W. (1988). Work in mathematics classes: The context of students' thinking during instruction. *Educational Psychologist, 23*(2), 167-180.

Elliott, R., Kazemi, E., Lesseig, K., Mumme, J., Carroll, C., & Kelley-Petersen, M. (2009). Conceptualizing the work of leading mathematical tasks in professional development. *Journal of Teacher Education, 60*(4), 364-379.

Elmore, R. F., Peterson, P. L., & McCarthey, S. J. (1996). *Restructuring the Classroom: Teaching, Learning, and School Organization*. San Francisco, CA: Jossey-Bass Inc.

Escudero, I., & Sánchez, V. (2007). How do domains of knowledge integrate into mathematics teachers' practice? *The Journal of Mathematical Behavior, 26*(4), 312-327.

Evertson, C. M., Emmer, E. T., & Brophy, J. (1980). Predictors of effective teaching in junior high mathematics classrooms. *Journal for Research in Mathematics Education, 11*(3), 167-178.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A Longitudinal Study of Learning to Use Children's Thinking in Mathematics Instruction. *Journal for Research in Mathematics Education, 27*(4), 403-434.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and it's impact. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 148-164). New York: MacMillan.

Frank, K. A., Kim, C., & Belman, D. (2010). Utility theory, social networks, and teacher decision making. In A. J. Daly (Ed.), *Social Network Theory and Educational Change*. Cambridge: Harvard University Press.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. *Sociology of Education, 77*, 148-171.

Franke, M. L., Carpenter, T. P., Fennema, E., Ansell, E., & Behrend, J. (1998). Understanding teachers' self-sustaining, generative change in the context of professional development. *Teaching and Teacher Education, 14*(1), 67-80.

Franke, M. L., Carpenter, T. P., Levi, L., & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal, 38*(3), 653-689.

Franke, M. L., & Kazemi, E. (2001). Learning to Teach Mathematics: Focus on Student Thinking. *Theory into Practice, 40*(2).

Freeman, D. J., & Porter, A. C. (1989). Do Textbooks Dictate the Content of Mathematics Instruction in Elementary Schools? *American Educational Research Journal, 26*(3), 403-421.

Fullan, M. (2000). The return of large-scale reform. *Journal of Educational Change, 1*(1), 5-27.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results from a National Sample of Teachers. *American Educational Research Journal, 38*(4), 915-945.

Gibbons, L. (2012). *Examining mathematics coaches in supporting teachers to develop ambitious instructional practices: A three-paper dissertation investigating context, knowledge, and practice.* (PhD), Vanderbilt University.

Gibbons, L., Garrison, A., & Cobb, P. (2011). *Teacher networks and the role of the mathematics coach: How institutional factors influence coach centrality*. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA.

Good, T. L., & Brophy, J. E. (1987). *Looking in classrooms*. New York: Harper & Row.

Goos, M., Dole, S., & Makar, K. (2007). Designing professional development to support teachers' learning in complex environments. *Mathematics Teacher Education and Development, 8*, 23-47.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *The Teachers College Record, 111*(9), 2055-2100.

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45*(1), 184-205.

Gwet, K. L. (2010). *Handbook of Inter-Rater Reliability* (Second ed.). Gaithersburg, MD: Advanced Analytics, LLC.

Hartmann, D. P., Barrios, B. A., & Wood, D. D. (2004). Principles of behavioral observation. In M. Hersen (Ed.), *Comprehensive Handbook of Psychological Assessment* (Vol. 3, pp. 108-137). Hoboken, N. J.: John Wiley & Sons.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28,* 524-549.

Herbel-Eisenmann, B. A., Lubienski, S. T., & Id-Deen, L. (2006). Reconsidering the Study of Mathematics Instructional Practices: The Importance of Curricular Context in Understanding Local and Global Teacher Change. *Journal of Mathematics Teacher Education, 9*(4), 313-345. doi: 10.1007/s10857-006-9012-x

Herbst, P. G. (2003). Using Novel Tasks in Teaching Mathematics: Three Tensions Affecting the Work of the Teacher. *American Educational Research Journal, 40*(1), 197-238.

Hiebert, J., Carpenter, T. P., Fennema, E., Fuson, K. C., Wearne, D., Murray, H., . . . Human, P. (1997). *Making sense: Teaching and learning mathematics with understanding*. Portsmouth, NH: Heinemann.

Hiebert, J., Gallimore, R., Garnier, H. E., Givvin, K. B., Hollingsworth, H., Jacobs, J., . . . Stigler, J. W. (2003). Teaching mathematics in seven countries: Results from the TIMSS 1999 video study. Washington, DC: U.S. Department of Education.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Charlotte, NC: Information Age Publishing.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., . . . Gallimore, R. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis, 27*(2), 111.

Hill, H. C., Ball, D. L., Blunk, M. L., Goffney, I. M., & Rowan, B. (2007). Validating the ecological assumption: The relationship of measure scores to classroom teaching and student learning. *Measurement: Interdisciplinary Research and Perspectives, 5*(2), 107-118.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372-400.

Hill, H. C., Blunk, M. L., Charalambos, C. Y., Lewis, J. M., Phelps, G., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511.

Hill, H. C., Schilling, S. G., & Ball, D. L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *The Elementary School Journal, 105*(1), 11-30.

Hill, H. C., Sleep, L., Lewis, J. M., & Ball, D. L. (2007). Assessing teachers' mathematical knowledge: What knowledge matters and what evidence counts? In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (Vol. 1, pp. 111-156). Charlotte, NC: Information Age Publishing.

Horn, I. S. (2007). Fast Kids, Slow Kids, Lazy Kids: Framing the Mismatch Problem in Mathematics Teachers´ Conversations. *Journal of the Learning Sciences, 16*(1), 37.

Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal, 47*(1), 181-217.

Jamar, I., & Pitts, V. R. (2005). High expectations: A "how" of achieving equitable mathematics classrooms. *Negro Educational Review, 56*(2/3), 127-134.

Kaufman, J. H., & Stein, M. K. (2010). Teacher learning opportunities in a shifting policy environment for instruction. *Educational Policy, 24*(4), 563-601.

Kazemi, E., Franke, M., & Lampert, M. (2009). *Developing pedagogies in teacher education to support novice teachers' ability to enact ambitious teaching.* Paper presented at the 32nd Annual conference of the Mathematics Education Research Group of Australasia.

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education, 7*, 203-235.

Kruse, S., & Louis, K. S. (1995). Developing professional communities in new and restructuring urban districts. In K. S. Louis & S. Kruse (Eds.), *Professionalism and community: Perspectives on reforming urban schools* (pp. 187-207). Thousand Oaks, CA: Corwin Press, Inc.

Lambdin, D., & Preston, R. (1995). Caricatures in innovation: Teacher adaptation to an investigation-oriented middle school mathematics curriculum. *Journal of Teacher Education, 46*(2), 130-140.

Lampert, M. (2001). *Teaching Problems and the Problems of Teaching*. New Haven: Yale University Press.

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers' and teacher educators' learning in and for practice. *Elementary School Journal, 109*(5), 491-509.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*: Cambridge University Press.

Lloyd, G. M. (1999). Two teachers' conceptions of a reform-oriented curriculum: Implications for mathematics teacher development. *Journal of Mathematics Teacher Education, 2*(3), 227-252.

Lloyd, G. M., & Behm, S. (2002). *The impact of prospective elementary teachers' experiences with reform-oriented mathematics curriculum materials*. Paper presented at the Annual Meeting of the Americal Educational Research Association, New Orleans.

Lloyd, G. M., & Wilson, M. (1998). Supporting Innovation: The Impact of a Teacher's Conceptions of Functions on His Implementation of a Reform Curriculum. *Journal for Research in Mathematics Education, 29*(3), 248-274.

Manouchehri, A. (2004). Implementing mathematics reform in urban schools: A study of the effect of teachers' motivation style. *Urban Education, 39*(5), 472-508.

Manouchehri, A., & Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *The Journal of Educational Research, 92*(1), 27-41.

Manouchehri, A., & Goodman, T. (2000). Implementing Mathematics Reform: The Challenge within. *Educational Studies in Mathematics, 42*(1), 1-34.

Matsumura, L. C., Sartoris, M., Bickel, D. D., & Garnier, H. (2009). Leadership for literacy coaching: The principal's role in launching a new coaching program. *Educational Administration Quarterly, 45*(5), 655-693.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. B. (2006). Measuring reading comprehension and mathematics instruction in urban middle schools: a pilot study of the instructional quality assessment: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

McGinnis, J. R., Parker, C., & Graeber, A. O. (2004). A cultural perspective of the induction of five reform-minded beginning mathematics and science teachers. *Journal of Research in Science Teaching, 41*(7), 720-747. doi: 10.1002/tea.20022

Munter, C. (Under review). Developing visions of high-quality mathematics instruction. *Journal for Research in Mathematics Education*.

National Council of Teachers of Mathematics [NCTM]. (1989). Curriculum and Evaluation Standards. Reston, VA: Author.

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and Standards for School Mathematics*. Reston, VA: Author.

National Governors Association for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics*

Neuberger, J. A. (2010). *Fostering effective mathematics teaching: Professional coaching and teachers' instructional practices and beliefs.* (Doctor of Education), Rutgers, The State University of New Jersey.

Opdenakker, M.-C., & Van Damme, J. (2006). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and Teacher Education, 22*(1), 1-21. doi: 10.1016/j.tate.2005.07.008

Owston, R. D., Sinclair, M., & Wideman, H. (2008). Blended learning for professional development: An evaluation of a program for middle school mathematics and science teachers. *Teachers College Record, 110*(5), 1033-1064.

Penuel, W. R., Riel, M., Krause, A. E., & Frank, K. A. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *Teachers College Record, 111*(1), 124-163.

Peterson, P. L., Carpenter, T. P., & Fennema, E. (1989). Teachers' knowledge of students' knowledge in mathematics problem solving: Correlational and case analyses. *Journal of Educational Psychology, 81*(4), 558-569.

Peterson, P. L., McCarthey, S. J., & Elmore, R. F. (1996). Learning from school restructuring. *American Educational Research Journal, 33*(1), 119-153.

Philipp, R. A. (2007). Mathematic teachers' beliefs and affect. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teachign and Learning* (pp. 257-318). Charlotte, NC: Information Age Publishing.

Philipp, R. A., Flores, A., Sowder, J. T., & Schappelle, B. P. (1994). Conceptions and practices of extraordinary mathematics teachers. *Journal of Mathematical Behavior, 13*, 155-180.

Prawat, R. S., & Jennings, N. (1997). Students as Context in Mathematics Reform: The Story of Two Upper-Elementary Teachers. *The Elementary School Journal, 97*(3), 251-270.

Putnam, R. T., & Borko, H. (1997). Teacher learning: Implications of new views of cognition. In B. Biddle, T. L. Good & I. F. Goodson (Eds.), *International handbook of teachers and teaching* (Vol. 2, pp. 1223-1296). Dordrecht/Boston/London: Kluwer Academic Publishers.

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*(2), 213-228.

Quinn, D. M. (2002). The impact of principal leadership behaviors on instructional practice and student engagement. *Journal of Educational Administration, 40*(5), 447-467.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics, 128*(2), 301-323.

Raymond, A. M. (1997). Inconsistency between a Beginning Elementary School Teacher's Mathematics Beliefs and Teaching Practice. *Journal for Research in Mathematics Education, 28*(5), 550-576.

Remillard, J. T. (1999). Curriculum Materials in Mathematics Education Reform: A Framework for Examining Teachers' Curriculum Development. *Curriculum Inquiry, 29*(3), 315-342.

Remillard, J. T. (2000). Can Curriculum Materials Support Teachers' Learning? Two Fourth-Grade Teachers' Use of a New Mathematics Text. *The Elementary School Journal, 100*(4), 331-350.

Remillard, J. T. (2005). Examining Key Concepts in Research on Teachers' Use of Mathematics Curricula. *Review of Educational Research, 75*(2), 211-246.

Remillard, J. T., & Bryans, M. B. (2004). Teachers' Orientations toward Mathematics Curriculum Materials: Implications for Teacher Learning. *Journal for Research in Mathematics Education, 35*(5), 352-388.

Remillard, J. T., & Kaye, P. (2002). Supporting Teachers' Professional Learning by Navigating Openings in the Curriculum. *Journal of Mathematics Teacher Education, 5*(1), 7-34. doi: 10.1023/a:1013862918442

Robinson, V. M. J., Lloyd, C. A., & Rowe, K. J. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44*(5), 635-674.

Rousseau, C. K., & Powell, A. (2005). Understanding the significance of context: A framework to examine equity and reform in secondary mathematics. *The High School Journal, 88*(4), 19-31.

Schneider, R. M., & Krajcik, J. A. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education, 13*(3), 221-245.

Schoenfeld, A. H. (2011). *How We Think: A Theory of Goal-Oriented Decision Making and its Educational Applications*. New York: Routledge.

Sherin, M. G. (2002). When teaching becomes learning. *Cognition and Instruction, 20*(2), 119 - 150.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-21.

Skemp, R. R. (1978). Relational Understanding and Instrumental Understanding. *The Arithmetic Teacher, 26*(3), 9-15. doi: 10.2307/41187667

Skott, J. (2001). The emerging practices of a novice teacher: The roles of his school mathematics images. *Journal of Mathematics Teacher Education, 4*(1), 3-28.

Smith, M. S. (2000). Balancing Old and New: An Experienced Middle School Teacher's Learning in the Context of Mathematics Instructional Reform. *The Elementary School Journal, 100*(4), 351-375.

Smylie, M. A., & Evans, A. E. (2006). Social capital and the problem of implementation *New directions in education policy: Confronting complexity* (pp. 187-208).

Son, J.-W. (2008). *Elementary teachers' mathematics textbook use in terms of cognitive demands and influential factors: A mixed method study*. Doctoral Dissertation. Curriculum, Teaching, and Educational Policy. Michigan State University. East Lansing.

Sowder, J. T., Philipp, R. A., Armstrong, B. E., & Schappelle, B. P. (1998). *Middle-grade teachers' mathematical knowledge and its relationship to instruction: A research monograph*. Albany, NY: State University of New York Press.

Spillane, J. P., Diamond, J. B., Burch, P., Hallett, T., Jita, L., & Zoltners, J. (2002). Managing in the middle: School leaders and the enactment of accountability policy. *Educational Policy, 16*(5), 731-762.

Spillane, J. P., Reiser, B. J., & Gomez, L. M. (2006). Policy implementation and cognition *New directions in educational policy implementation* (pp. 47-64).

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating Productive Mathematical Discussions: Five Practices for Helping Teachers Move Beyond Show and Tell. *Mathematical Thinking and Learning, 10*(4), 313 - 340.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488.

Stein, M. K., & Kaufman, J. H. (2010). Selecting and supporting use of mathematics curricula at scale. *American Educational Research Journal, 47*(3), 663-693.

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50-80.

Stein, M. K., & Nelson, B. S. (2003). Leadership content knowledge. *Educational Evaluation and Policy Analysis, 25*(4), 423-448.

Stein, M. K., Remillard, J., & Smith, M. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 319-370). Charlotte, NC: Information Age Publishing.

Stein, M. K., Smith, M. S., & Silver, E. A. (1999). The development of professional developers: Learning to assist teachers in new settings in new ways. *Harvard Educational Review, 69*(3), 237-269.

Stipek, D., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17*, 213-226.

Stodolsky, S. S., & Grossman, P. (2000). Changing students, changing teachers. *Teachers College Record, 102*(1), 125-172.

Sullivan, P., & Leder, G. C. (1992). Students' Influence on Novice Australian Teachers' Thoughts and Actions regarding Mathematics Teaching: Two Case Studies. *The Elementary School Journal, 92*(5), 621-642.

Sun, M., Garrison, A., Larson, C., & Frank, K. A. (Under review). Exploring colleagues' professional influence on teachers' learning. *Teachers College Record*.

Superfine, A. C. (2009). The "problem" of experience in mathematics teaching. *School Science and Mathematics, 109*(1), 7-19.

Swafford, J. O., Jones, G. A., Thornton, C. A., Stump, S. L., & Miller, D. R. (1999). The impact on instructional practice of a teacher change model. *Journal of Research and Development in Education, 32*(2), 69-82.

Sztajn, P. (2003). Adapting reform ideas in different mathematics classrooms: Beliefs beyond mathematics. *Journal of Mathematics Teacher Education, 6*(1), 53-75.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15*(2), 105-127.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-147). New York: MacMillan.

Thompson, C. L., & Zeuli, J., S. (1999). The frame and the tapestry: Standards-based reform and professional development. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 341-375). San Francisco: Jossey-Bass.

Trenkamp, K. K. (2007). *An investigation of the factors that influence reform implementation practices of middle school mathematics teachers.* (Doctor of Philosophy), University of Maryland, College Park, College Park, Maryland.

Turner, J. C., Warzon, K. B., & Christensen, A. (2011). Motivating mathematics learning: Changes in teachers' practices and beliefs during a nine-month collaboration. *American Educational Research Journal, 48*(3), 718-762.

Walker, E. N. (2007). Rethinking professional development for elementary mathematics teachers. *Teacher Education Quarterly, 34*(3), 113-134.

Wang, J., & Paine, L. W. (2003). Learning to teach with mandated curriculum and public examination of teaching as contexts. *Teaching and Teacher Education, 19*(1), 75-94. doi: 10.1016/s0742-051x(02)00087-2

Wilson, S. M., & Berne, J. (1999). Teacher Learning and the Acquisition of Professional Knowledge: An Examination of Research on Contemporary Professional Development. *Review of Research in Education, 24*, 173-209.

Wood, T., Cobb, P., & Yackel, E. (1991). Change in teaching mathematics: A case study. *American Educational Research Journal, 28*(3), 587-616.

Yun-peng, M., Chi-chung, L., & Ngai-ying, W. (2006). Chinese primary school mathematics teachers working in a centralized curriculum system: a case study of two primary schools in North-East China. *Compare: A Journal of Comparative and International Education, 36*(2), 197-212.

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low-achieving students and higher order thinking. *Teaching and Teacher Education, 17*(3), 469-485.

Zohar, A., & Dori, Y. J. (2003). Higher Order Thinking Skills and Low-Achieving Students: Are They Mutually Exclusive? *The Journal of the Learning Sciences*, *12*(2), 145-181.