DIFFERENTIATED COACHING OF HIGH SCHOOL MATHEMATICS TEACHERS: THE INSTRUCTIONAL TRAJECTORY OF PERFORMANCE TASKS

by

Deborah J. Crawford
A Dissertation Submitted to the Graduate Faculty of George Mason University In Partial Fulfillment of The Requirements for the Degree of Doctor of Philosophy Education

Committee:

[Signatures]

Chair

Program Director

Mark R. Ginsberg
Dean, College of Education and Human Development

Date: July 23, 2015

Summer Semester 2015
George Mason University Fairfax, VA
Differentiated Coaching of High School Mathematics Teachers: The Instructional Trajectory of Performance Tasks

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Deborah J. Crawford
Master of Education
George Mason University, 1995
Bachelor of Science
Indiana University of Pennsylvania, 1980

Chair: Margret Hjalmarson, Associate Professor
College of Education and Human Development

Summer Semester 2015
George Mason University
Fairfax, VA
This work is licensed under a creative commons attribution-noderivs 3.0 unported license.
Dedication

This study is dedicated to my seven children: Ashley, William, Matthew, Sarah, Hannah, Katherine, and Edward, and to my parents, Edward and Cornelia.
Acknowledgements

Deep thanks and gratitude go to my family and friends who have supported me during this journey of learning. Thanks especially to my mother for her ongoing encouragement, Phil for his unconditional support, and to Hannah and Sarah for babysitting the twins. A special thanks goes to Dr. Margret Hjalmarson for advising me along the way and in particular for pointing me in the direction of curricular design research. Thanks to Drs. Jennifer Suh and Erin Peters-Burton for providing guidance and ideas to focus my project.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
<tr>
<td>Chapter One. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1. Purpose Statement</td>
<td>1</td>
</tr>
<tr>
<td>2. Rationale</td>
<td>1</td>
</tr>
<tr>
<td>3. Importance of Study</td>
<td>5</td>
</tr>
<tr>
<td>4. Local Contribution of Study</td>
<td>8</td>
</tr>
<tr>
<td>5. Research Questions</td>
<td>11</td>
</tr>
<tr>
<td>6. Conceptual Framework</td>
<td>12</td>
</tr>
<tr>
<td>7. Categories</td>
<td>13</td>
</tr>
<tr>
<td>8. Definition of Terms</td>
<td>15</td>
</tr>
<tr>
<td>9. Dissertation Overview</td>
<td>19</td>
</tr>
<tr>
<td>Chapter Two. Review of the Literature</td>
<td>21</td>
</tr>
<tr>
<td>Literature Categories</td>
<td>22</td>
</tr>
<tr>
<td>Data Base Search Parameters</td>
<td>22</td>
</tr>
<tr>
<td>Introduction: Literature Review Foci</td>
<td>23</td>
</tr>
<tr>
<td>Constructivist Paradigm of Reform Teaching Practices</td>
<td>30</td>
</tr>
<tr>
<td>Mathematical Knowledge for Teaching (MKT)</td>
<td>34</td>
</tr>
<tr>
<td>Task Selection</td>
<td>40</td>
</tr>
<tr>
<td>Professional Development of Math Teachers</td>
<td>42</td>
</tr>
<tr>
<td>Chapter Three. Methodology</td>
<td>56</td>
</tr>
<tr>
<td>Research Questions</td>
<td>56</td>
</tr>
<tr>
<td>Research Design</td>
<td>57</td>
</tr>
<tr>
<td>Mixed Methods Design</td>
<td>63</td>
</tr>
<tr>
<td>Context and Participants</td>
<td>74</td>
</tr>
<tr>
<td>Data Collection</td>
<td>80</td>
</tr>
<tr>
<td>Summary</td>
<td>94</td>
</tr>
<tr>
<td>Chapter Four. Mixed Method Results</td>
<td>96</td>
</tr>
<tr>
<td>Teacher Profiles</td>
<td>96</td>
</tr>
<tr>
<td>IQA Observation Instrument Results</td>
<td>100</td>
</tr>
<tr>
<td>Academic Rigor</td>
<td>112</td>
</tr>
<tr>
<td>Accountable Talk</td>
<td>118</td>
</tr>
<tr>
<td>Performance Task Surveys</td>
<td>125</td>
</tr>
<tr>
<td>Survey Results</td>
<td>126</td>
</tr>
<tr>
<td>Summary</td>
<td>143</td>
</tr>
</tbody>
</table>
Chapter Five. Comparative Teacher Cases: Findings in Context ................................... 145
  Division Impetus for the Design of the Professional Learning Model .................. 145
  Teacher Selection .................................................................................................. 146
  Teacher Background ............................................................................................ 147
  Goal Setting ......................................................................................................... 149
  Case Studies: A Comparison of the IQA Dimensions ........................................ 150
  Will’s Experience as a New Teacher .................................................................... 159
  Ashley’s Experience as an Experienced Teacher ............................................... 184
  Sarah’s Experience as a Special Education Teacher ............................................ 202
  Interpretive Summary ......................................................................................... 222
  Salient Themes ..................................................................................................... 225
  Chapter Six. Conclusions ..................................................................................... 234
    Summary ............................................................................................................ 234
    Discussion of Findings: Research Question One .............................................. 237
    Discussion of Findings: Research Question Two .............................................. 248
    Implications for Teacher Practices and Professional Learning Models ............ 254
    Limitations of the Study .................................................................................... 257
    Implications for Research ................................................................................. 258
    Closing Ideas ...................................................................................................... 260
  Appendix A. Performance Task Surveys .............................................................. 262
  Appendix B. IRB Exemption Letter and Informed Consent Form ......................... 264
  Appendix C. Video Observation Feedback Survey .............................................. 267
  References .............................................................................................................. 274
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Table 1</em>. Data Collection and Data Analysis Procedures</td>
<td>90</td>
</tr>
<tr>
<td><em>Table 2</em>. Data Sources and Research Questions</td>
<td>91</td>
</tr>
<tr>
<td><em>Table 3</em>. Lesson Analyses Count</td>
<td>102</td>
</tr>
<tr>
<td><em>Table 4</em>. Task Potential</td>
<td>103</td>
</tr>
<tr>
<td><em>Table 5</em>. Task Implementation</td>
<td>104</td>
</tr>
<tr>
<td><em>Table 6</em>. Student Participation in Discussion of Task</td>
<td>104</td>
</tr>
<tr>
<td><em>Table 7</em>. Rigor of Teacher Questioning</td>
<td>104</td>
</tr>
<tr>
<td><em>Table 8</em>. Task Potential vs. Implementation by Teacher Group</td>
<td>105</td>
</tr>
<tr>
<td><em>Table 9</em>. Means for Each of the 10 IQA Rubric Categories for Academic Rigor and Accountable Talk Over the Four Lesson Cycles</td>
<td>107</td>
</tr>
<tr>
<td><em>Table 10</em>. Academic Rigor Dimensions by Teacher Group</td>
<td>108</td>
</tr>
<tr>
<td><em>Table 11</em>. Accountable Talk Dimensions by Teacher Group</td>
<td>109</td>
</tr>
<tr>
<td><em>Table 12</em>. Correlations: Instructional Quality Assessment Dimensions</td>
<td>111</td>
</tr>
<tr>
<td><em>Table 13</em>. Lesson Goal-Setting</td>
<td>131</td>
</tr>
<tr>
<td><em>Table 14</em>. Academic Rigor: Potential of Task</td>
<td>152</td>
</tr>
<tr>
<td><em>Table 15</em>. Task Potential Crosstabulation by Teacher</td>
<td>155</td>
</tr>
<tr>
<td><em>Table 16</em>. Task Implementation</td>
<td>157</td>
</tr>
<tr>
<td><em>Table 17</em>. Task Potential vs. Task Implementation</td>
<td>158</td>
</tr>
</tbody>
</table>
**List of Figures**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Data Collection Framework</td>
<td>86</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Means for Student and Teacher Linking Over Four Lesson Cycles</td>
<td>121</td>
</tr>
</tbody>
</table>
Abstract

DIFFERENTIATED COACHING OF HIGH SCHOOL MATHEMATICS TEACHERS: THE INSTRUCTIONAL TRAJECTORY OF PERFORMANCE TASKS

Deborah J. Crawford, Ph.D.

George Mason University, 2015

Dissertation Advisor: Dr. Margret Hjalmarson

The purpose of this study was to explore the impact of a video lesson analysis model on the capacity of secondary mathematics teachers to deliver performance tasks with higher cognitive demand by increasing student reasoning through classroom discourse. Of interest was how to personalize teacher professional development given the typical constraints of a public school division in terms of resources. This study provides one model of using formative assessment on 10 key dimensions of cognitive demand and classroom discourse through digital media to differentiate teacher professional development during mathematical tasks with self-assessment and peer feedback. This study used survey, observational ratings on the Instructional Quality Assessment (IQA) instrument, and a descriptive case study to give information on successes and challenges to add to the knowledge base of individualized, job-embedded professional development models for secondary mathematics teachers.
Findings indicated that teachers improved in all 10 dimensions when they set a specific goal to improve their classroom discussions based on a baseline video, self-assessed with criteria from the IQA rubrics, and considered coaching feedback from peers and a coach. Rigor of teacher questioning was the most significant dimension aligned with increased cognitive demand by students in mathematical task implementation. Teachers viewed self-assessment of their own lesson videos against the common Instructional Quality Assessment criteria as the most valuable component of the lesson study. Teachers improved in their capacity to deliver mathematics tasks with higher cognitive demand and explored how to personalize embedded professional development through coaching feedback.

This study offers one model of how divisions can reach these goals with the realities of limited resources such as coaching staff, funding, and time. Therefore, this model includes efforts to explore the use of current web tools to expand professional development through virtual collaboration.
Chapter One

Purpose Statement

The purpose of this study was to explore how to increase capacity of secondary mathematics teachers to deliver performance tasks with higher cognitive demand by increasing student reasoning through classroom discourse. Of interest was how to personalize teacher professional development given the typical constraints of a public school division in terms of resources. This study provides one model of using formative assessment on 10 dimensions of accountable talk and academic rigor through digital media to differentiate teacher professional development during the trajectory of mathematical task planning, implementation and self-assessment with peer feedback. This mixed-methods survey and case study gives information on successes and challenges to add to the knowledge base of individualized, job-embedded professional development models for secondary mathematics teachers.

Rationale

As a result of increased workplace demands for all students to acquire fundamental content knowledge of functions and statistics beyond introductory algebra by completion of secondary education, the Common Core State Standards (CCSS) (Common Core State Standards Initiative, n.d.) and other aligned state curricula now require similar knowledge by the end of high school. Additionally, schools are focusing
on the larger goal of improving transition readiness for postsecondary college and careers by developing critical thinking skills, use of data for decision making, and habits of mind. Together, students need to leave high school with application of higher mathematics content beyond introductory algebra and geometry and the ability to problem solve in an ever-changing workplace that requires critical and creative thinking, collaboration, and communication skills. Such process skills can be honed through mathematical content as students solve rich tasks.

The goal of teachers is to prepare all of their students for futures in college and careers amidst a fast-shifting environment. Teachers well know that students need experiences in thinking flexibly to solve complex problems, often as members of a team, and in using data analysis to make decisions. No teacher would debate that our students need to be problem solvers, communicators, or flexible thinkers in nonroutine situations. During K-12 education, which mathematics students have access to and how students learn that mathematics determines their preparation for these life numeracy and problem-solving skills as well as habits of mind such as persistence and flexibility.

During task selection, teachers decide which mathematics students are exposed to; a lower level-thinking task will never require higher level thinking. Choosing a rich instructional task is one indicator of quality teaching. However, the level of cognitive demand that a student experiences depends ultimately not only on task selection but also on the way the task is deployed. The implementation of the task during its trajectory determines the quality of student learning (Boston, 2012a).
Brophy (1999) said that practices that engage students in thoughtful discourse about powerful ideas make a difference in such learning. Students in classrooms where cognitively rich tasks are given routinely experience the greatest gains in student performance (Hiebert et al., 2003; Oettinger, 2011; Stein, Grover, & Henningsen, 1996). The National Council of Teachers of Mathematics (NCTM), through the 1989 and 2000 Process Standards, has outlined a vision and framework for mathematics classrooms to engage students in learning mathematics through problem solving, reasoning and proof, connections, communication and representation (NCTM, 1989, 2000). The Mathematical Practices describe how to enact the Process Standards in a classroom community by articulating specific roles and behaviors of teachers and students (NCTM, 2014).

**Teacher is key.** The teacher’s role is key in determining the quality of mathematics instruction that students experience. High quality instructional practices are associated with greater improvements in student performance (Darling-Hammond, 2000; Darling-Hammond & Youngs, 2002; Hattie, 2012; Monk & King, 1994). “Students’ understanding of mathematics, their ability to use it to solve problems, and their confidence in doing mathematics are all shaped by the teaching they encounter in school” (Darling-Hammond, 2000).

**Challenges.** In spite of the importance of these instructional decisions, many teachers are not sure how to teach mathematics through problem solving. Cultural teaching norms for teaching mathematics have focused on a limited number of methods; students in typical classrooms spend most of their instructional time “acquiring isolated skills through repeated practice” (Stigler & Hiebert, 1999, 2009). Previous secondary and
postsecondary courses typically were taught procedurally and preservice methods courses
often did not prepare teachers to teach conceptually to meet this goal. Inservice
professional development efforts, until the last decade, typically were conducted as
discrete workshops that were not sustained or job-embedded. Most high school teachers
have not had common planning time by course. Such limited time to collaborate on
planning for lessons requiring new skills and little, if any, feedback from busy
administrators unfamiliar with mathematics pedagogy have limited such attempts by
teachers to change. Many teachers have not been able to practice new methods risk-free
from resulting test scores, and reflect to refine new methods collaboratively with peers.
Nor have teachers typically been given opportunities for self-analysis against criteria with
feedback on their attempts to change instructional practices. With the emergence of
Professional Learning Communities (PLC) during the last 20 years, more school
divisions have given teachers time to collaborate on planning, instruction, and embedded
assessment. Principals have been doing walk-throughs; however, their feedback varies in
usefulness based on focus, coding, and specificity. Few recognize how process standards
or mathematical practices are enacted in students and teacher moves during the trajectory
of a task. These challenges have limited the input of opportunities for teachers to grow in
mathematical content knowledge for teaching (MKT) and research-informed methods
during traditional professional development models, even if embedded.

Often the lessons and assessments developed during valuable PLC collaboration
times have not often been different from the mathematics lessons developed in isolation.
Without new input or feedback aligned to a set of criteria, teacher teams tend to design lesson plans using the activities and tasks that they are most familiar with using.

Even where teachers have selected tasks requiring higher levels of thinking, the cognitive demand often declines during the trajectory of the task in the learning cycle. With the exception of Japan, higher-achieving countries did not use a greater percent of high-level tasks than in the United States in the Third International Mathematics and Science Study (TIMSS) 1999 study. All other referenced countries were, however, more successful in not reducing these tasks to procedural exercises (Smith & Stein, 2010).

**Importance of Study**

Reform math classrooms are set up as learning communities where students discuss their thinking about solutions to problem situations, often in a real-world context, for the purpose of decision making. It only makes sense to align teacher’s learning targets and professional development models with this outcome. Providing teachers with opportunities for continuous learning while immersed in the job of teaching makes sense as a vehicle for change toward this goal. The importance of this study is twofold: Improve the capacity of math teachers to deliver mathematics tasks with higher cognitive demand and explore how to personalize embedded professional development through coaching feedback. Finally, the study is situated in the current dilemma of how divisions can reach these two goals with the realities of limited resources such as coaching staff, funding, and time. Therefore, this model includes efforts to explore the use of current web tools to expand professional development through virtual collaboration.
Previous studies have well documented the connection between teachers’ beliefs about the use of mathematical tasks in reform classrooms and task selection choices and resulting trajectories (Boston & Smith, 2009; Stein, Engle, Smith, & Hughes, 2008; Stein et al., 1996). The tasks that teachers select for their students to do, determine the access to particular mathematics for those students. Next, how the task unfolds in the learning environment determines whether and how students engage in reasoning and communication of their thinking to others. Studies in the last decade have identified teacher pedagogical moves that keep tasks at high thinking levels (Lappan, Smith, & Jones, 2012; Smith, Bill, & Hughes, 2008). Stein and Smith have studied for over two decades how to design professional development to prepare preservice teachers and transform inservice elementary and middle school teachers to teach mathematics through problem solving with rich tasks (Boston, 2012a; Smith et al., 2008; Smith, Hughes, Engle, & Stein, 2009; Stein et al., 1996). Smith and Stein’s 2011 five practices of orchestrating productive mathematical discussions move the teacher’s role to that of facilitator. A needed next step in current research will be to study exactly how teachers and students create the classroom environment where the practices can be deployed (Williams-Candek & Smith, 2015). This study examines in detail how teachers moved to a facilitator’s role by creating classroom learning communities for discourse. Key was that teachers were not told to become facilitators or how to set up this culture. This study fills a small step in understanding the process of realization by teachers of the need to adjust teaching behaviors so that student reasoning can be made visible through classroom discourse. Only then can social construction of understanding occur through
the conversations between students and between teachers and students. The task is the vehicle through which the mathematics understanding is negotiated.

Similar studies with high school teachers have been published less frequently and more recently than those with elementary and middle school participants (Boston, 2012b). Recent studies have explored how to provide embedded professional development with high school teachers and specifically to the importance to the learning of algebra (Boston, 2013; Boston & Smith, 2009; Smith & Stein, 2010). Jong argued for differentiation of professional development in her mixed-methods survey and case study of two elementary teachers having the same teacher preservice program but differing in teaching placements and background experiences (2009). Current studies provided data on instructional quality from three main sources: student achievement tests, value-added measures like student growth percentiles, and classroom observations (Boston, 2014). This study fills a gap in the literature on evaluating improvement in instructional quality through classroom observation in four ways: by studying inservice high school mathematics teachers rather than preservice, K-8, or college; in utilizing cost-effective coaching in a blended model using video, rather than real-time observations; by focusing on formative feedback including self- and peer assessment; and by treating the professional learning process as an iterative, rather than as a linear or intervention design.

This study was built on these previous studies to explore how embedded professional development on the use of performance tasks can be differentiated for individual secondary teachers using a blended approach of face-to-face and virtual coaching via video and other digital online media. Teachers collaborated with a coach...
and peers using commonly used free applications and web tools such as Google Apps for Education (GAFE). To incrementally improve their mathematics instruction, teachers in this study focused on student reasoning and classroom discourse occurring during classroom task implementation. Practices focused on using student thinking as formative data to guide instructional decisions. The same reasoning was applied to teacher learners; the researcher and teachers used formative data from observations together to plan next steps in learning. The professional development activities became the work of the teachers of this course, collaborating together. In doing so, the model in this study utilizes Stigler and Hiebert’s six principles for gradual improvement in teaching (1999, 2009).

**Local Contribution of Study**

This study was situated firmly in the professional learning of high school teachers of a mathematics course designed to increase equity of access to college and careers. A cohort of Algebra, Functions, and Data Analysis (AFDA) teachers collaborated for one school year to study their individual task implementation across 10 dimensions. Success for struggling AFDA students extended their learning beyond geometry to meet diploma requirements. Bruner reminded teachers in 1960 that schools waste valuable time by withholding authentic problems until all “prerequisite” skills are acquired. Gojak (2011) argued for the value of engaging students in problem solving without having specifically taught the related procedures by using strategies for learning through these rich problems. These teachers’ AFDA students were at the end of their high school math courses, typically without having had much exposure to problem solving. The goal of this professional development was to improve the capacity of teachers to facilitate student
reasoning about a few big ideas emerging from selected problem tasks by increasing the quality of and participation in classroom discussions. The model focused on two parts: academic rigor of the task as selected and implemented and the accountable talk by both students and teachers occurring during classroom discourse.

Expanding professional development collaboration statewide. A second importance for this study existed in the local context of AFDA as a relatively new option to meet the minimum graduation requirement. At the time of this study, no current organized professional development was available to help divisions provide individualized professional development to mathematics teachers who were trying to use performance tasks in a classroom community of struggling learners. An earlier Mathematics and Science Partnership grant had provided initial higher education in course content taught through problem-solving tasks to about 25 pioneer AFDA teachers four years prior. A second grant funded an Algebra and Functions course for algebra teachers, some of whom also taught AFDA. Out of this teacher pool, only 1 out of the 10 teachers in this study had participated. Across the state, the course had not met with success when taught traditionally with direct teaching and a textbook. In many cases, a lone teacher taught AFDA at a high school with no collaboration used in this division’s course design. In contrast, during the first three years of implementation, other divisions had viewed this AFDA team’s course design as an example of successful implementation. To help other implementing teachers fill this professional development gap, this team had created collaboration experiences in the course design process and shared digital resources for this new course. Previous team members had made a point to
get involved in statewide collaboration with other AFDA teachers. This team had shared their journey of creating an inquiry-based math course for struggling students by speaking at regional and state conferences, serving as invited speakers, inviting teachers from other divisions to collaborate, and by sharing the division AFDA curriculum website. Additionally, this division had welcomed site visits from other divisions who wished to see AFDA students doing mathematical tasks.

In the year of this study, the professional development needs of the teacher team had changed. With staffing changes and teacher mobility, 6 out of the 10 division AFDA teachers were new to the course. Only 2 had been part of the original cohort’s grant-funded courses and collaboration. This pioneering division now needed similar professional development in the pedagogy of implementing tasks in a reform classroom as the teachers they had been mentoring. Public school divisions routinely experience such loss of collective experience and high-functioning professional learning teams. With limited professional development funds that have included cuts in funding for teacher collaboration and coaching, districts have struggled with how to orient and coach teachers. Teaching an inquiry-based course, where students learn mathematics through problem solving, requires a complex skill set (Williams-Candek & Smith, 2015). This study attempted to add crucial information on how a cohort of teachers assigned to teach a mathematics course for struggling learners collaboratively studied their own task implementation through a cost-effective video lesson study using social media. Two overarching research questions guided the study.
**Research Questions**

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
   a. For what purposes did high school mathematics teachers select tasks for students?
   b. What did high school mathematics teachers want to change about task implementation in their classrooms?
   c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?
   d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?
   e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?
   a. How did teachers describe the change process in their own task implementation?
   b. Which coaching activities did teachers perceive as leading to growth?
   c. How did the learning process compare across three teacher groups: New, Special Education, and Experienced?
   d. What suggestions did teachers offer to improve coaching activities in future models?
Conceptual Framework

My conceptual framework situated this study and was shaped by post-Vygotskian activity theorists such as Davydov and Radzikhovskii (1995) to describe how humans come to know. As a result, I grounded my professional development in a social constructivist approach to learning mathematics, with a particular focus on modeling as central to this AFDA course (Boaler, 1999; Lesh & Doerr, 2003; Lesh, Lester & Hjalmason, 2003). Lave and Wendel’s 1991 social practice theory coined “a community of practice” and “the connectedness of knowing,” and NCTM’s vision framed my pedagogical content knowledge of how a classroom learning community operated (NCTM, 1989, 2000). The setting of a public school professional development model necessitated both pragmatic and integrated dialectic paradigms to make use of this research. Findings were applied internally for AFDA curricular and instructional program design refinement and externally to share as a model for others.

Researcher perspective. My constructivist approach to teaching and learning mathematics began with the publication of the National Council of Teachers of Mathematics’ Standards in 1989. I began studying the research behind the Standards later in my graduate studies. Dewey (1916), Shulman, Schoenfeld, Lester, Cobb, Ball, Boaler, Lannin, Ma, Smith, and many others have studied reform classrooms in depth and have informed the way I coach math teachers.

Paradigms. Jennifer Greene’s explanations of the different paradigms were helpful in seeing where I stood in both the pragmatist and dialectic, integrated camps. Maxwell’s second and third editions of Qualitative Research Design, An Integrated
Framework (2002, 2013) explained how the integration of various paradigms, differently in the context of each study, questioned the existence of distinct paradigms. In this study, I take a decidedly pragmatic stance in the interest of utility of this study for professional development of teachers and program design. The study examined activities through a social-constructivist lens while the interpretation of the data was through a dialectic paradigm. Thus, the three paradigms meld in the particular context of this study.

**Choice of mixed methods.** Literature on comparative case studies and mixed-methods research formed another category (Maxwell, 2013). My search topics included the following: reform mathematics, inquiry-based learning, pedagogical qualities of high-quality instruction, mathematical knowledge for teaching (MKT), mathematical tasks, cognitive demand, classroom discourse, and the professional development of mathematics teachers. Research models used were action research, curriculum design (Hjalmarson & Lesh 2008) and program evaluation.

**Categories**

To highlight pivotal researchers and studies that have formed my thinking on this topic of teacher professional development designed to increase the cognitive demand of classroom mathematics tasks, I have cited references from each related category in the literature review. Research was integrated throughout the sections in addition to the literature review chapter. The overarching categories include: the evolution of thinking about what constitutes teacher quality; mathematics knowledge for teaching (MKT); a description of standards-based, mathematics learning communities; beliefs of teachers about task selection; cognitive demand of tasks; classroom discourse including teacher
moves affecting the task trajectory; and the models of professional development of mathematics teachers associated with change in practices. This research began with a broad assumption that the teacher quality matters in the degree and depth of student learning (Darling-Hammond, 2000; Monk & King, 1994; Sanders & Horn, 1994; Sanders & Rivers, 1996). My doctoral studies and larger literature review have crystalized down to the specific knowledge needed to plan and study this professional development-coaching model. The model itself is of interest as much as the change in task implementation practices of the specific group of teachers.

**Math teacher professional development.** Stein and Smith from the University of Pittsburgh, along with later coauthors such as Boston and Hughes, have studied professional development of preservice teachers for two decades (Boston & Smith, 2009; Smith & Stein, 2011; Stein et al., 2008; Stein et al., 1996; Stein, Smith, Henningsen, & Silver, 2000, 2009). Their seminal collective works have framed my thinking about job-embedded professional development of teachers who are learning to teach mathematics through rich tasks. In studying reform classrooms, these researchers have focused on the importance of the cognitive demand of performance tasks in mathematics classrooms. I have been using their collective works beginning with Mathematical Tasks Framework: a representation of how mathematical tasks unfold during instruction (Stein et al., 1996). Their resulting Task Analysis Guide has been used universally as a key tool for coding descriptors for the cognitive depth of mathematics tasks to rank task potential. Stein et al.’s (2000, 2009) casebook on the professional development of mathematics teachers has been part of my library as a math supervisor for years. I have used Smith and Stein’s
2011 book and related articles on orchestrating productive mathematical communication in classrooms with my teachers as professional development on how to structure and use data from student communication during tasks. Thus, Smith’s collective works have been pivotal in my literature review on the professional development of teachers using tasks in reform mathematics classrooms (Boston & Smith, 2009).

**Professional development as design.** Lesh, Lester, and Hjalmarson (2003) have framed how to use the engineering design of curriculum and professional development of teachers to design such cognitively rich programs. I have used their framework to design the professional development of the AFDA course design. Loucks-Horseley, Love, Stiles, Mundry, and Hewson (2009) edited a book on professional development of math and science teachers from which I have incorporated the key components of job-embedded, long-term sustainability and continuous growth into the model.

**Definition of Terms**

The following terms were used in this study in the specific context of a mathematics classroom: *performance task, cognitive demand, Process Standards, Virginia Process Goals, Mathematical Practices, problem solving, reasoning, communication, connections, representation, classroom discourse, precision, strategic use of tools,* and “doing” mathematics. Ten dimensions were rated by teachers during classroom video analyses: *participation, student linking, teacher connections/linking, teacher press, student responses, classroom discussion, task potential, task implementation, questioning,* and *mathematical residue.* Assessment terms used in this study included *formative assessment, summative assessment, common assessment, high-*
stakes testing, and rubric. Another program term for student learning used was College and Career Readiness. Professional development terminology included job-embedded, sustainable, and Professional Learning Community (PLC).

Central to this study is the term performance task, which is used synonymously in this study with math task or rich task. The following definition is taken from a typical public school division publication, because it is one that teachers commonly are given:

Performance tasks are formative assessments designed to provide the classroom teacher with information about their students’ thinking and learning which should be used as feedback to modify instructional [moves] and learning activities in the classroom. The performance task is designed to be the lesson, which is built around problems and explorations. The problems are used as a medium for learning mathematics instead of an application at the end of learning. (Little-Kaumo & Fritz, 2001)

Rich tasks can be synonymous with performance tasks implying that they integrate multiple mathematics concepts and skills in a multistep context, forming natural connections that can be intra- or interdisciplinary. Boston and Smith (2009), whose Instructional Quality Assessment instrument was used to collect data on task selection and implementation in this study, defined tasks as, “A mathematical task is a single complex problem or a set of problems that focuses students’ attention on a specific mathematical idea” (p. 136).

Real-world tasks can be both authentic and rich, offering connections through the given scenario, often leading to some decision-making action. All terms refer to tasks
selected by teachers to be used instructionally by students as the vehicle for constructing understanding of mathematical content.

Tasks offer differing levels of thinking; the term *cognitive demand* is used to mean the level of complexity of student thinking required by the task. Lower level thinking in tasks was described by Stein, Smith, and Silver (1999) as requiring memorization or procedures without connections. Higher level tasks offer opportunities for students to engage in procedures with connections or “doing” mathematics (Stein et al., 1999). Stein and Smith define “*doing mathematics*” as requiring cognitive effort and complex, nonalgorithmic thinking. Students explore to understand the mathematics, self-regulate when stuck, and then access and synthesize their knowledge to persist in solving problems (Boston & Smith, 2009; Smith, 2000).

*Process Standards* refer to the National Council of Teachers of Mathematics classic set of five universal expectations for how students learn mathematics (NCTM, 1989, 2000). The process standards were first published in 1989 to include *problem-solving*, *reasoning*, *connections* and *communication*. These process standards were revised in 2000 to include *representation*. The term *Process Goals* refers to the Virginia Department of Education term for the same set of process standards as the five from NCTM. The CCSSI’s Mathematical Practices (n.d.) adds two more: *precision* and *strategic use of tools*. *Classroom discourse* refers to the conversations student-to-student and student-to-teacher in the classroom community that serve to make student thinking visible. Learners communicate their mathematical thinking using five representations:
verbal, concrete, numerical, graphical/pictorial, and symbolic/algebraic through virtual or face-to-face interactions.

The NCTM Process Standards describe a vision for classroom learning communities where students construct understanding of mathematics through problem solving. The term *problem solving* in this study will be defined as a process whereby students are given frequent experiences in their math program to think deeply about, communicate their thinking and reason to solve complex problems. These problems are often real world and connected in many ways to personal learning. In problem solving, students develop transferrable strategies that can be refined over time to be used with other problem situations. Students emerge as thinkers who can solve problems in the workplace with a team and communicate these solutions to others for decision making.

Preparing students to be able to think and continuously learn is at the heart of *College and Career Readiness (C&CR)*, a movement to better prepare students in K-12 for postsecondary transitions. C&CR projects have created mathematics, English, and work-readiness proficiencies for students to be deemed college ready. The course in this study is part of the C&CR pathway for students who will be college and career bound but not yet ready skill-wise.

The study uses the term *formative assessment* as the process by which teachers and students gather information to provide feedback for use in making next instructional decisions to move learning forward (Bloom, Hasting, & Madaus, 1971; Hattie, 2012; Lesh, Hoover, Hole, Kelly, & Post, 2000; Wiliam, 2011). *Summative assessment* is an evaluation of the learning at a given point in time. To contrast, *formative assessment* is
ongoing while *summative* is typically an event. *Common assessments* here are defined as performance measures given by a division to all students in particular courses. *High-stakes testing* refers to state or other testing that is tied to graduation requirements. *Rubrics* are scoring tools used to assess student and teacher understanding in the study classrooms.

*Professional Learning Community (PLC)* in this study refers to a collaborative learning effort by a team of teachers who are working on particular common goals. Teacher teams involved in continuous learning about their work while on the job are receiving *job-embedded professional development* (PD). *Sustainable* professional development refers to learning opportunities that are ongoing where teachers can go through the iterative design cycle of creating and testing prototypes of new learning. In this study, teachers were able to reflect, adjust, and implement new strategies with greater skill over time. In *personalized learning*, planning decisions by the coach and teachers are based on individual performance feedback. In a *blended or hybrid* learning model, teachers grow through both face-to-face and virtual collaboration. The following Google Applications for Education (GAFE) were used by teachers and coaches to collaborate in this video lesson analysis model: Google Sites, Docs, Sheets, Forms, You Tube, Google+ and division email.

**Dissertation Overview**

This dissertation study is part of a larger five-year program implementation design study of the AFDA course in a K-12 public school district. This study connects the purpose and research questions from the parent AFDA study to the research questions in
this study. This dissertation is composed of six chapters: introduction, literature review, methodology, data analysis, comparative case study, and discussion. Original data are available embedded in the paper or in appendices.
Chapter Two

The literature surrounding the research questions was situated in the evolution of the research on the importance of a teacher’s ability to keep classroom mathematical tasks at a high level of cognitive demand. This literature review has informed me of the current knowledge in the area of provision of targeted, content-specific professional development to improve teacher practice in mathematical task selection, implementation, and reflection to continuously improve teacher practice. The topic of professional development of mathematics teachers related to improvement in the use of student reasoning through classroom discourse during task deployment has been studied in depth since the mid-1990s. My overarching idea was that improvements in teaching and learning by teachers occur in a similar iterative design model as used in the curriculum design of this AFDA course. My premise is that professional development should be differentiated by the teachers themselves through collaborative feedback between and among teachers, peer teachers and the coach. This approach mirrors a constructivist paradigm in the learning of any student. Furthermore, this study design fills a niche in providing such coaching in a cost-effective manner through a blended approach of face-to-face coaching, video, and virtual learning.

The study gives needed information on how teachers set instructional improvement goals, select and implement tasks and grow from self-assessing real
performance against set criteria with feedback followed by reflection and research of strategies to try next. The detailed descriptions of the actual coding of student and teacher moves during classroom discourse are helpful to learn how teachers and students act and change in the classroom community. The change part of the story gives a glimpse of how teachers create a classroom mathematics community where one had not previously existed. Williams-Candek and Smith (2015) have indicated that this piece is missing in current research.

**Literature Categories**

The overarching categories include: the evolution of thinking about what constitutes teacher quality, necessary content and math knowledge for teaching (MKT), task selection criteria, beliefs of teachers about task selection related to cognitive demand, student reasoning and classroom discourse in a standards-based mathematics learning community, teacher moves affecting the maintenance of the cognitive demand in the task trajectory, and the models of professional development of mathematics teachers associated with change in practices. Each category exists within the context of a constructivist paradigm of student learning mathematics through problem solving.

**Data Base Search Parameters**

I organized my review of the literature by searching databases with keyword searches, all relevant mathematics education journals, international studies primarily from Australia, Canada, England, and Japan, mathematics education, international and design research handbooks, key international studies including TIMSS and PISA,
conceptual framework philosophy, pedagogy for high-quality instruction, and mathematical knowledge for teaching literature.

I conducted keyword searches using mathematical tasks or cognitive demand or reform mathematics; classroom discourse; teacher quality and student achievement in mathematics, pedagogy for quality teaching (general); mathematical tasks and cognitive demand; professional development of teachers and math knowledge for teaching and pedagogy; and discourse processes, mathematics education, and qualitative research.

To keep updated on new studies published during the data collection, analysis and writing phases of this study, I subscribed to digital device alerts for studies related to rich mathematical tasks, cognitive demand and classroom discourse. I read articles of interest and added them to End Note.

Introduction: Literature Review Foci

This research begins with a broad assumption that teacher moves in the classroom matter in the degree and depth of student learning (Darling-Hammond, 2000; Hattie, 2012; Monk, 1994; Rowe, 2003, Sanders & Rivers, 1996). In his meta-analyses of over 50,000 empirical research studies, John Hattie found that from among the six groups of factors influencing successful learning in schools-student, home, school, teacher, curricula, and teaching-the factor ‘teacher’ seems to have the strongest effect (Hattie, 2012). Hattie’s argument is that “when teaching and learning are visible, there is a greater likelihood of students reaching higher levels of achievement” (2012, p. 21). This literature review begins with this premise, first exploring the search over time for characteristics of teacher quality and focusing on the necessary pedagogical content
knowledge for teaching mathematics (MKT) (Hill, Rowan, & Ball, 2005). Within MKT, the study targets successful moves that mathematics teachers make during task selection and implementation to keep the cognitive demand high. This model examines the literature on how teachers as learners can increase the visibility of student reasoning through classroom discourse by using formative feedback during classroom tasks. A review of the professional development literature regarding the specific knowledge and factors to be considered when educators plan personalized, collaborative professional learning with formative feedback to effect teacher growth closes this chapter.

**Factors in student achievement.** In a landmark 1966 study of about 600,000 students, 60,000 teachers, and 4,000 schools, James Coleman asserted that student background defined outcomes, mattering more than the experience of schooling. Included in the schooling characteristics unrelated to student achievement was teacher quality. Over 40 years of study of the effect of various teacher characteristics ensued. Studies in the 1980s and 1990s focused on certification and teacher content knowledge, which led to increased certification requirements by states. In 1994, data drawn from the Longitudinal Study of American Youth found that subject area preparation had a positive effect on student learning gains. However, effects were found to diminish with time and vary across types of students (Monk, 1994). Sanders created a stir with his 1994 and 1996 value-added growth studies on the effect of teacher quality on mathematics student achievement across grades three, four, and five (Sanders & Horn, 1994; Sanders & Rivers, 1996). These studies used a large state database to track student performance growth year-to-year rather than at a single collection point in time. The results suggested
that the teacher mattered the most in student outcomes. Sanders’ studies focused the research trajectory on teacher quality to include the teacher’s qualities beyond certification and content knowledge.

**Role of teacher certification.** In a flurry of teacher quality and student achievement studies over the next ten years, there was still controversy with varying findings from large-scale data sets and empirically-based studies. Scheerens and Bosker argued that schools mattered only 20%, teacher and classroom quality 20%, while differences in background among students accounted for 60% of variance in student performance (Bosker, 1997). Background, however, could include effects of previous schooling and effective teachers. The National Center for Statistics reported in 1999 that teacher verbal ability, general ability, and content knowledge matter while certification and professional development activities did not. The recommendation was to loosen the certification requirements to more easily let career switchers or those without education preservice credentials teach, if they had the content knowledge (Statistics, 1999). Legislation soon followed in the reauthorization of the Elementary and Secondary Education Act which mandated highly-qualified teachers in every classroom (Whitehurst, 2002).

Stanford’s Darling-Hammond spoke out emphatically that certification requirements created a stronger teacher pool, with her research on an analysis of empirically-based studies that related to what mattered in teacher quality. Her stance was that certification requirements requiring educational coursework in content as well as pedagogy, including preservice experiences such as student teaching were related to
academically stronger new teachers as evidenced by teacher exam scores and improved success and retention (Darling-Hammond, 2002; Darling-Hammond, 2000). Other researchers around this time concurred that certification in mathematics mattered (Goldhaber, 2000; Rivkin, Hanushek, & Kain, 2001). Both content knowledge and the pedagogy from education courses were found to be necessary in secondary mathematics. Both NAEP and Rowan found that fourth grade student performance data indicated that teacher experience mattered; however, there was no relationship between student achievement in math and a teacher’s major, master’s degree, or certification in math (NAEP, 2000; Rowan, Correnti, & Miller, 2002). Recently, John Hattie’s synthesis of over 900 meta-analyses ranked teacher subject matter knowledge as “136 out of 150 influences on student achievement with an effect size of 0.09” (Hattie, 2012, p. 268).

Other teacher qualities. Since the 1960s, researchers have studied which teacher qualities influence student learning. Most studies have been qualitative, based on case studies, comparison studies of new and experienced teachers, and international comparisons of teachers’ math content knowledge (Hill & Ball, 2009; Hill et al., 2005). Few link mathematical knowledge to student achievement (Hill et al., 2005; Rowan & Miller, 1997). Cognitive ability, in numerous studies of largely verbal skills and meta-analyses, was often cited as first in importance, followed by targeted content-specific training, and experience (Greenwald, 1996; Harris, 2011; Rowan, 2002; Whitehurst, 2002). Master’s degrees, entrance exam scores, certification, and general workshops had lower effect sizes on student achievement (Harris & Sass, 2011; Whitehurst, 2002). In elementary grades, math courses and area of certification are not associated with better
results; in only a handful of studies does this matter in secondary (Hill & Ball, 2009; Hill et al., 2005). Researchers have concurred that teacher qualities such as experience and content-specific pedagogy do matter (Harris, 2011; Hill et al., 2005).

Furthermore, like Rivkin, Rowan urged researchers using large-scale data sets to examine carefully research methodologies, which had led to varying causal relationships involving teacher effects on student performance (Rivkin, Hanushek, & Kain, 2001, 2005; Rowan, 2002). These wildly varying results over 30 years, have, in the past decade, focused into general consensus that teacher quality matters, because of the primary role of the teacher in creating the learning experience for students. Experienced, math-certified teachers with higher cognitive ability who have had targeted, content-specific professional development were thought to fare better in creating effective learning experiences for students.

**Domains of mathematical content knowledge for teaching.** Yet cognitive ability and subject-specific content knowledge are not the variables associated with the highest gains in student achievement. Shulman first identified the areas of knowledge needed by teachers: content knowledge, pedagogical content knowledge for teaching, and curriculum knowledge (Shulman, 1986; Wilson, Shulman, & Richert, 1987). Ball et al. created a model of MKT that identified domains within the two overarching area of subject matter knowledge and pedagogical content knowledge (Ball, Thames, & Phelps, 2008). Under subject matter knowledge are common content knowledge such as knowing an algorithm to get a correct answer, specialized content knowledge such as being able to model integer operations with a charge model, and horizontal mathematical knowledge of
the larger vertical mathematical curriculum (Hill & Ball, 2009). The umbrella of
pedagogical content knowledge is formed from three domains: knowledge of content and
students, knowledge of content and teaching, and knowledge of the curriculum. These
areas blend together as MKT in the classroom in a teacher’s work.

A teacher’s knowledge of mathematical content knowledge for teaching (MKT)
generates effect sizes almost double that of general cognitive ability (Hill et al., 2005).
Two recent studies have used the CKT-M measure of mathematical content knowledge
for teaching to assess teacher MKT and then to compare it to with student achievement.
Hill et al. gave this measure to over 300 elementary teachers of grades one through three.
The researchers were interested to find that as early as first grade, MKT was associated
with student performance. Economists gave the same CKT-M test to over 400 teachers
along with tests for general cognitive ability and personality. MKT had an effect size of
almost double that of cognitive ability or personality on student achievement (Rockoff,
2008). Although teachers do need to understand the mathematics they are teaching, they
must additionally be able to understand how students construct knowledge of
mathematics, know what students are thinking and see from a student vantage point why
they are giving a particular response (Ball & Foranzi, 2010; Walshaw & Anthony, 2008).
It seems that the domain of having enough content-specific knowledge for the level of
math course is necessary, but being able to produce student learning matters more (Hill et
al., 2005).

In 2003, Ken Rowe from the Australian Council of Educational Research and
John Hattie (2012), found that although student background and characteristics mattered,
they were much less influential on student outcomes elementary through secondary as
compared to the quality of the teaching and learning experience (Hattie, 2012; Rivkin,
2005; Rowe, 2003; Whitehurst, 2002). These studies form the basic tenant of this study:
that teachers and classroom learning experiences are more important to student learning
than any other factor, including student background. Research recently has supported the
idea that if basic content knowledge appropriate to the level of mathematics being taught
and general pedagogy are in place from certification requirements, mathematics teachers
will benefit most from additional content-specific MKT to improve the student learning
experience (Hill et al., 2005). The NCTM Process Standards describe this experience in
terms of the student look-fors of problem solving, reasoning and proof, communication,

This study seeks to explore a model for improving a teacher’s ability to create a
classroom where students are reasoning at high levels and communicating their thinking
to other students as well as the teacher. Such classroom discourse makes student
mathematical thinking visible to teachers and other students in order to provide formative
assessment and feedback opportunities to move learning to the next step in
understanding. Teacher moves to create a mathematical learning community are at the
heart of reform mathematics, based on the constructivist notion that students construct
and adjust understanding of mathematics through interaction with experiences in a social
setting. Teaching is complex; creating and deploying tasks in such a classroom requires a
very complex skill set that must be learned (Monk, 1994; Stein et al., 1996; Walshaw &
Anthony, 2008; Williams-Canchek & Smith, 2015).
The goal of professional development in this study is to coach teachers through feedback and self-reflection, in order to improve their ability to keep tasks at a high thinking level (Boston, 2013; Smith & Stein, 2010). Just as with students, teachers enter this professional development on a continuum of background experiences and beliefs; therefore, differentiation of the coaching is inherent in the model. In summary, teacher quality studies point to the teacher and the learning experience as paramount in student achievement; when targeted, professional development is aligned with the mathematics standards-based reform practices, student achievement is likely to improve (Stein et al., 1996).

**Constructivist Paradigm of Reform Teaching Practices**

**Teacher beliefs about problem solving.** Teaching through problem solving is often discussed as a key attribute of reform pedagogy. Defining problem solving is not easy. One oft-cited definition is that a problem is a situation where the student does not know the answer. Charles and Lester’s 1984 study of problem solving offers an early reform argument for inquiry-based mathematics. Schoenfeld’s 1983 study uses plane geometry as a context helps teachers understand what shapes peoples’ behavior as they solve problems. His qualitative study details how to analyze problem-solving performance on three levels: access to cognitive resources stored in long-term memory, executive decision-making, and belief systems. A study of 105 secondary math teachers sought to correlate teachers’ beliefs about pedagogy to their ability to correctly rank algebraic tasks by difficulty level (Nathan & Koedinger, 2000). Teachers who practiced reform pedagogy were more able to rank tasks by difficulty and better assess students’
mathematical development. The symbol-precedence view seems to mediate teachers’ judgments regarding students’ mathematical development (Nathan & Koedinger, 2000). A 2009 survey of 1222 K-12 math/science teachers designed to measure beliefs about inquiry found elementary teachers to believe in and use inquiry more than secondary teachers. Math teachers used inquiry more than science teachers (Marshall et al., 2009).

Mathematical modeling is central to the functions approach used in many of the AFDA tasks in this study. This literature review confirmed a vision of key features of the discussions that would occur during an following data collection in such tasks. Justi, et al in 2002 said that modeling is not widely practiced despite the diversity of available models. In order to help students understand conceptually, she advocates that teachers understand models in general, know how and when to use models, have the ability to create good teaching models, have the skills to conduct modeling activities in class, and understand how students construct their own mental models and how to deal with those expressed models in class (Justi et al., 2002). Obviously, the development and acceptance of mathematical models is complex and not easily accomplished, yet national standards call for students to connect math and science to real world phenomena learned through authentic activities (Carrejo & Marshall, 2007). The authors show how in statistics, students always start with a real world context through data collection and then connect to a formal, abstract mathematical model. Students can then ask if the model makes sense in the given context. Modeling then can be described as “constructing an idealized, abstract model and comparing it to a real system” (Giere, 1999, p. 50).
Modeling certainly has become easier with the use of technologies, particularly the graphing calculator and data collection equipment such as CBLs, CBRs, Lab Pro and various probeware. Use of these technologies allows students to collect real world data, and organize, display, analyze and interpret it in real time. After 30 years of debate about the use of calculators in math classrooms, most of the debate now is over how the tools are to be used. Ellington completed a 2003 meta-analysis of 54 calculator studies and found that operational and problem-solving skills improved when calculators were an integral part of teaching and learning. Many studies show the value of calculator-based labs such as the investigation of distance vs. time plots using a CBR to help even struggling students justify understanding of functions (Stylianou, Smith, & Kaput, 2005; Yerushalmy, 2006). Students investigate physical representations of slope as a rate of change in math and as velocity in science and the y-intercept as the initial condition, or starting position. The way teachers in this study used calculators as a problem-solving tool or for just computational purposes aligned with the cognitive demand of the tasks in which they were used. In these investigations, students and teachers become partners in developing mathematical ideas and solving math problems (NCTM, 1989). This partnership in a community of learners (Boaler, 1999), requires communication of mathematical thinking by all parties, or discourse.

**Student communication of reasoning via classroom discourse.** Teachers establish classroom learning communities that can enact reform practices of learning mathematics through problem solving (Hufferd-Ackles, Fuson, & Sherin, 2004). If students are engaged in worthwhile tasks, and are asked to explain their reasoning, then
the task itself gives them something to talk about (Stein et al., 2000, 2009). Smith and Stein (2011) describe the practices of orchestrating such classroom discourse that leads to students understanding the big ideas of the mathematics inherent in the task. During planning, teachers need to anticipate student solution strategies and misconceptions and plan essential questions to ask based on student understandings. During task implementation, teachers need to monitor student thinking made visible in oral discourse and in written work. For the summary discussion following the task, teachers need to select and sequence student solutions to be shared to summarize the big mathematical ideas that were the learning target. Facilitate this discourse and scaffold, monitor, and facilitate the conversations. Students are often unclear and need to elaborate or focus their thinking (Kieran & Dreyfus, 1998). The importance of discourse is more than just communication in order to be understood. It goes back to the sociocultural construction of understanding in a social setting (Cobb, 1999). Kids in a classroom test their individual constructions against others’ thinking; they then reflect and adjust their own thinking based on the evidence presented in the social experience.

In summary, the pedagogy needed to teach in a reform manner is more complex than is required to teach in a traditional manner. Students who participate in a traditional classroom are viewed as understanding when they can successfully follow procedures; students in an inquiry-based classroom must show they understand by creating, manipulating and explaining mathematical models, as well as justifying when asked (Ma, 1999,). Phil Duran, representing the Common Core (CCSS) committee, said in the 2013 Virginia Council of Mathematics Supervisors’ meeting that in the United States, teachers
ask themselves while planning a lesson, how they can teach students to answer this particular type of problem correctly. In contrast, Japanese teachers ask themselves how students can understand the mathematics from this particular problem.

**Mathematical Knowledge for Teaching (MKT)**

Ball and Hill describe MKT as the teacher “knowledge necessary to carry out the work of teaching mathematics” (Hill et al., 2005). Hill (2010) defines MKT in the context of high quality math instruction as: use of mathematical explanations and representations, interpretations of student responses and the ability to avoid math errors and imprecision.

The real mathematical thinking that goes on in a classroom, in fact, depends heavily on how deeply the teacher understands mathematics (Ma, 1999). If a teacher’s knowledge is limited to procedures, then it makes sense that the teacher would struggle if asked to establish a classroom of inquiry-based instruction (Ma, 1999). She explains why such teachers cling so much to textbooks; the textbooks provide a consistent source of content knowledge that can be studied and retold to students. The problem, of course, is that students do not learn from telling but must construct their own understanding. In her example of the staircase problem where students investigate slope, Sherin (2002), shows how teaching can become learning for the teacher. Math content emerges as a collection of practices of the classroom community (Cobb, 1999). He says the choice of activities and practices reflects content choices. That means that content and pedagogy are intertwined in school mathematics. Ball and Foranzi stated simply that, “Student learning of mathematics depends fundamentally on what happens inside the classroom as teachers and learners interact over the curriculum” (Ball & Foranzi, 2010, p. 17)
Hill and Ball have studied which specialized mathematical knowledge is most crucial for a teacher. They identified three areas of knowledge: interpreting and analyzing student work, providing mathematical explanations, and connecting pictorial to symbolic representations (Hill & Ball, 2009). These researchers focused on mathematical explanations and representations in a summer school laboratory while other similar programs focused on other aspects of MKT. Students gained more when teachers focused efforts on classroom discourse and representations (Hill & Ball, 2009; Hill et al., 2005). As a result, the teacher professional development in this study will focus on these practices.

Under the domain of understanding student work, teacher tasks are to exhibit the following skills: provides mathematical explanations that are understandable to a child, analyze student errors and know where the child is in his or her thinking, understand unconventional student solutions, select the next step in terms of examples, and assessing representations in textbooks or other curriculum resources (Hill & Ball, 2009). Teachers in this study will code for many of these behaviors.

A search for meaning. One example of teacher MKT from algebraic content occurs in the teaching of the concept of functions. Reform classrooms emphasize a functions-based approach algebra rather than an equations approach. In a functions-approach, students start with a real world physical context, go to the graph right away to interpret, and write an equation to fit the data. A transformational approach to graphing is used to understand functions by comparing them to parent functions using dynamic representations on the calculator or computer. In an equations approach, students start
with an abstract equation and then go to the table and graph it; they use it in a contextual application only if there is time. Thus, the different approaches have different practice in terms of symbolic rules, ordered pairs, social data, use of physical situations, and controlling images (Mesa, 2004). Functions can then be represented as a physical field or as an abstract concept of function. The functions-approach asks students to develop an abstract equation to fit, or model, a physical data set. The physical field helps students understand the abstract concept. Paz and Leron (2009) describe a functions approach as being, “…the slippery road from actions on objects to functions and variables.” Erickson (2006) also describes the search for meaning in the parameters, such as a physical representation of slope as a rate of change and the problems students have coordinating the data and the models. He quips that curve fitting and math modeling are neither complicated nor new; it seems strange that mathematics teachers do not use this “science” way to understand functions (Erickson, 2006, p. 23).

Likewise, Gerber and Reineke (2005) explain how databases can be used to understand statistics in a real world context. AFDA has a community project that students do with a team to analyze a data set from a business, organization, or environmental science setting to answer a question or form an inference. The data base project outlined in the Gerber and Reineke example, uses local sources of data for student statistics projects.

In short, students can connect functions to real phenomena as they represent that concrete situation as a table, graph, equation/expression and verbally in words (VandeWalle & Lovin, 2005). Fluency with functions means being able to move
comfortably between and among the different representations, and understanding the connections between them. For example, students forming connections would not just locate the max/min/vertex in the table but would additionally connect the meaning of these critical attributes in terms of the physical situation. Fluency is needed to get to the real value using a function for decision-making back in the context of the situation. Again, students and teachers utilize complex behaviors and deep content understandings to connect and synthesize interpretation of data by both students and teachers.

Sherin’s idea of teachers modifying their content knowledge during instruction comes to play here. Students often explain or discover an idea that reframes the more procedural/static view of the concept for the teacher. The community of practice would have to be in place for teachers to act as learners rather than dispensers of knowledge.

**Understanding students.** A teacher’s practice is an evolution of relationships (Lampert, 2004, as cited in Sowder, 2007). There are teacher-to-student interactions, student-to-student, teacher-to-teacher, teacher-to-administrator, and many types of classroom group relationships. The teacher orchestrates the content, the representations, and the people in the room (Sowder, 2007). Together, as a community of learners, they negotiate meaning. At the time of this study, teachers were setting specific, measurable, attainable, realistic and timely (SMART) goals, collecting and analyzing data on student results as part of their evaluation process. However, in this study, the design was nonevaluative and collaborative with the same SMART focus on reflection and adjustment of practice leading to growth.
Teachers need to have a deep understanding of the content they teach but also of the children they serve. Professional development (PD) should be designed to strengthen the teacher-student learning connection. Darling-Hammond & Sykes (1999) offer the relationship elements of ideal professional development efforts. Learning theory should guide professional development in terms of how learning opportunities affect eventual learning. Professional development should be embedded in the context of student curriculum. In addition to understanding content deeply, teachers must understand how students learn the curriculum and develop varied ways of presenting it to students. Through varied assessments, teachers must find multiple ways to examine evidence of that learning. With regard to change, teachers need opportunities through professional development to fully understand new implementations. Finally, the teacher evaluation process would be tied to student learning (Darling-Hammond, 2008).

Unfortunately many educational reforms or new implementations bypass classrooms with economically disadvantaged or ELL backgrounds (Spillane, 2001, quoted by Hufferd-Ackles et al., 2004). Some researchers have found evidence that some children, particularly low SES, are not able to access the classroom discourse (Boaler, 1999; Lubienski, 2002). Likewise, our district’s AFDA class has an enrollment of a proportionally higher percentage of economically disadvantaged and students with disabilities. However, in AFDA we purposefully chose to design lessons where students would learn through problem solving in performance tasks and modeling to fit hands-on data collection and analysis scenarios. The National Math Advisory Panel Report (2009) also emphasizes relationships as a basis for motivation especially with students who need
intervention. If opportunity to learn is measured as opportunities to engage, whether those chances are equitably distributed among students depends on the teacher. Teachers decide how participation is negotiated in a classroom (Franke, Kazemi, & Battey, 2007).

Students react differently to inquiry-based lessons after years of direct teaching with little expectation for discourse. Stein, Smith, and Silver (1999), point out that shaping student-learning opportunities is often hard for students used to many years of traditional instruction. It may take time to try out new materials, modeling of expectations and lots of wait time over many investigations for students to understand how to function in a reform class. The same is true of open-ended assessments after years of plug-and-chug multiple choice.

Cobb, Yackel, and Wood (1993) remind teachers of the importance of the classroom community since students construct meaning individually but also socially through interactions. If the classroom is not set up for equity in discourse with students expected to reason and justify their thinking, this social construction of meaning will be reduced. Many “lower level” classrooms are planned with a traditional curriculum in mind thanks to a theory that low socio-economic status (SES) or English Language Learners (ELL) kids are not as able to learn in a reform classroom versus one offering direct instruction with less inquiry or discovery (Lubienski, 2002). Students miss the chance to construct knowledge in this type of setting, continuing to be inexperienced in the ways of problem solving. This finding is inconsistent with other research that found that all students, including low SES, benefit from engaging deeply in mathematics (Boaler, 1998, as cited in Sowder, 2007). Situated cognition demands a rich rather than a
reduced classroom environment. Reduced equity of opportunity can perpetuate and increase any gaps, since students are not getting as much practice with problem solving and communicating math thinking through discussion as students in a reform or advanced curriculum. The AFDA curriculum offers the best practices of inquiry through problem-based learning. Students work in teams to learn together and discuss findings. They often are asked to make decisions about data, justifying their reasoning. The best evidence from the first year has been the rich and varied samples of student work of real world tasks and performance assessments. Teachers and students alike have said that they have been asked think more deeply than in a traditional classroom.

**Task Selection**

Boston and Wolf (2006) identified the most important first step in supporting student reasoning at a higher cognitive demand in any classroom is to select tasks that provide opportunity to learn mathematics through problem solving. Cohen and Ball emphasized the need for teachers to become serious learners of practice rather than learners of strategies and activities (Cohen & Ball, 2001). In collaborative content PLC meetings prior to this study, the teacher teams in this study focused on task selection and use of performance tasks for formative assessment purposes. The team had matured in the task selection process from solely a content focus to a combined content/pedagogical one. However, they had not yet studied their implementation through lesson study or other peer observations. The individualized professional development studied in this treatment, will focus on developing the necessary mathematical knowledge for teaching (MKT), coined by Deborah Ball, to deploy classroom tasks at a high level of cognitive demand.
Peg Smith’s 1999 book, *Professional Development of Math Teachers, Task Analysis Guide* rubric for evaluating tasks for cognitive demand, and her 2011 *Five Practices for Orchestrating Productive Mathematics Discussions* book with Mary Kay Stein (Smith, 2000; Smith & Stein, 2011; Smith, Hughes, Engle, & Stein, 2009) have been tools I have used in my work with teachers to evaluate tasks for cognitive demand during planning and to implement tasks. Multiple evolving studies on professional development of teachers around use of math tasks have been used. Smith describes changes in cognitive demand before, during, and after professional development in her study of 18 practicing teachers (Boston & Smith, 2009). The *Task Analysis Guide* (Stein & Smith, 2003, 2009) to measure cognitive demand of tasks has been used by AFDA teachers for planning prior to this study. Tasks were coded by teachers and the researcher fusing the Task Potential rubric from the *Instructional Quality Assessment* (IQA) instrument of the following descriptors: memorization, procedures without connections, procedures with connections, and “doing” mathematics (Boston, 2008, 2009, 2012; Stein, 2008). This instrument mirrors Stein and Smith’s 1999 *Task Analysis Guide* in the Task Potential rubric.

**Teacher moves in task trajectory.** This section outlines key teacher moves that appear repeatedly in studies of reform classrooms where cognitive demand remains high as students work with tasks. Of interest are the moves that I will study with teachers under the goals of student reasoning and communication about mathematical thinking. I will use Boston’s Instructional Quality Assessment Rubrics IQA rubrics to assess implementation of tasks in the domains selected by teachers as instructional growth goals.
Problem-posing is explored in each teaching episode with description of the types, cognitive demand, distribution and patterns of questions asked by the teacher. Chapin, O’Connor, O’Connor and Anderson (2009) along with Hughes, Smith, Boston and Hogel (2008) have published elementary and middle school cases of lesson cycles from launch of the task through the implementation and discussion to analysis of student work. This study adds to this research thread by analyzing high school cases of the problem posing and classroom discourse occurring during task implementation.

**Professional Development of Math Teachers**

My professional development question was how to embed individualized professional development right at the insertion point of new learning. In previous exit interviews, AFDA teachers thought they knew what worked or did not work in the classroom, but transcripts showed these conversations were limited to just that dichotomy. Details and elaboration were not present nor the analysis of student work. Teachers also said mathematical tasks were good or bad in terms of content alignment; comments were not based on criteria such as Smith’s 2009 rubric for evaluating the cognitive depth of such tasks. In year three, AFDA teachers had learned to choose tasks based on not only the mathematical content and the process skills, but also the cognitive depth of student thinking about the problems. For example, at one team meeting, teachers began to talk about choosing tasks that would make student understanding of quadratic functions visible. The next step was for teachers to build capacity in implementing these
tasks so that students engaged in problem solving and reasoning while communicating and connecting their ideas by representing thinking in different ways. The professional development plan in this study seeks to help teachers analyze their own moves to keep tasks at a higher level and know where students are in their thinking as made visible on the tasks. Teachers will self-assess their own moves during task selection and implementation against 10 dimensions of Academic Rigor and Accountable Talk during classroom discussions with the IQA rubrics (Boston, 2012).

**Literature on stages of teacher learning.** In their QUASAR project, Stein and Brown, 1997, defined learning as a transformation in the ways teachers participated in the collaborative activities and in the community that was being formed. The educational problem with AFDA was clear: how to design a course to teach the big ideas of functions and statistics to the lower 30% of the population of students needing a math course beyond geometry. Equity would demand that this course be taught with the most current ideas in course design, rather than a remedial skills approach. The second problem was that many teachers themselves had come to AFDA without teaching skills concurrent with reform mathematics, including inquiry-based instruction. In year one, four out of the eight teachers had created a sense of shared purpose and had begun to acquire the new skills, largely due to the university courses they had taken together. However, three special education teachers felt largely disenfranchised due to not having taken the university course with the others. Other factors included: little to no common planning which was needed to create the sense of belonging that the other teachers experienced and a perceived lack of content understanding. Additionally, at one site, the one special
educator position was, in reality, a revolving door of four people. I had responded by instituting six collaboration days during years one through three and four days in years four and five for this division AFDA content team to collaborate on unit planning to include PBL and community projects. By Year Five, turnover had created a new crop of inexperienced AFDA teachers. Without funding for six substitute days or the free graduate cohort classes that the original team had enjoyed, I designed this video lessons study to increase the cognitive demand during instruction. The model focuses on classroom discourse and connections between representations during tasks as per Ball’s research of which practices impact student growth most (Ball & Rowan, 2009). I designed a coaching model to continue the embedded, sustained collaboration between team members virtually across three high schools and to move past unit planning to improve instructional delivery of tasks. Borasi and Fonzi (2002) reminded us that teachers need both vision and commitment to school reform efforts with an attitude of inquiry about their own practice. Without the commitment to that vision, the new team of AFDA teachers in this study would not have been able to put in the time and hard work of examination of practice that this video lesson study required. I was prepared for this reality, so I used the two summer PD days prior to the lesson study for this purpose. A key motivator was teacher analysis of a personal baseline video with peer feedback and self-assessment to set an improvement goal for the study.

Edelson, in his learning-for-use model (2001), offers a design framework with three steps: motivate, construct, and refine. For years one and two, I was at the motivational level with the new teachers coming in to teach AFDA for the first time and
for the special educators who have not yet felt a shared sense of purpose. By year three, four teachers were operating at the construction level and providing coaching to the others who were at stage one. This study is situated during year five where I was back at the motivational level with six out of ten teachers who were new to the team. One was a long-term substitute. Three other teachers were operating at the construction level in their own classrooms. Another teacher was, at this time, serving as a leader to other teachers and sharing with other divisions. Looking ahead, three were at the cusp of becoming experienced teachers that I wanted to develop into leaders of this course. With such a range of experience and readiness, differentiation of professional learning was needed and is offered by this model. Following this study, two teachers presented at the state math conference, all of the AFDA team teachers shared resources with a large division who toured their program, and three teachers became invited speakers to a literacy conference. I was asked to speak at conferences and meetings about the video lesson study as well. In these ways, all of the members of the AFDA team, including the researcher, received differentiated professional development based on our learning continuum. The premise for Edelson’s model for learning originated from Whitehead, 1929, who said that shallow memorization leads to inert knowledge that cannot be recalled later to be of use in applications (Whitehead, 1929, cited by Edelson, 2001). Edelson, then, combines content and process in a teaching and learning model that I feel is useful for both student and adult learning. Edelson’s model begins with activities designed to create demand and elicit curiosity, similar to cognitive dissonance. This problematic gap or limitation in
understanding sets the purpose for needing the new learning that lies ahead. In this video lessons study, this event was the baseline video with goal setting. The second phase is construction where the learners receive directly or indirectly communication from others that allows them to build new knowledge. Activities should provide direct experiences to observe this new learning in action. This aligns with the AFDA teachers trying out new lesson strategies in their video taped classes. Thirdly, teachers apply the new knowledge in ways that are meaningful and reflect on the new knowledge and experiences in order to re-index their knowledge (Edelson, 2001). Specifically, the four lesson cycles allow teachers to try out new strategies, reflect with a rubric against their goals and receive feedback by several peers and a coach. Interestingly, in a problem-based course like AFDA, this model aligns with scientific inquiry, which consists of three phases: exploration, invention, and discovery (Lawson, 1995).

Stein and Brown (1997) said that learning could be viewed as a transformation of participation. They identified four stages: performance assisted by others, performance assisted by self, performance independent and automated, and performance reviewed and revised. In these ways, coaching by mentors comes to play in the video lesson study model used in this study.

Loucks-Horsley and Stiegelbauer in 1991 agreed that mandating change would not work. In their adapted framework of stages of concern that adopters of innovations go through during implementation, they recommended developing teacher leaders to serve as change agents. Stein and Brown in their 1998 project found that they needed a different approach with their second school, because that particular school community
had not yet agreed to the reform ideas. In this school, researchers relied on master teachers to lead the way using them to mentor new-to-reform teachers. This research team studied the program as the unit of analysis rather than individual teachers. I, too, used the AFDA program as the unit of analysis in a previous program implementation parent study. In this study, I studied individual teachers as the unit of analysis in one analysis and the teacher groups in another through the comparative case study: new, special education and experienced. I layered both analyses in my findings.

Action research is an ideal way for teachers to learn more about teaching and learning of mathematics in application to their own practice (Borasi & Fonzi, 2002). Stigler and Heibert (2009) and Cochran-Smith and Lytle (1999), also ask teachers to study their own practice. However, they view research as the third, most sophisticated stage in knowledge acquisition in the area of teaching. Knowledge for practice would come first (Cochran-Smith & Lytle, 1999). Teachers in this stage would typically take coursework or participate in formal PD programs, or follow a reform curriculum adopted by the district. One example would be the Developing Mathematical Ideas professional development materials that use case studies to examine student work, and reflect on videos of students being interviewed in order to understand the development of the big ideas of K-7 mathematics (Schifter, Bastable, & Russell, 2000). The Big Ideas program also asks participants to examine the practice of others (Sowder, 2007). In both, the venue is a formal professional development event lasting at least a year. All of our original AFDA teachers had completed two lesson studies as a division math professional
development prior to the video lesson study. Only two of the nine teachers in this study had been on the original team to have experienced lesson study.

The second phase of knowledge-in-practice would require more skill. Examples could include Professional Learning Communities (PLC), communities of practice, content teams by course or a lesson study group. Teachers in this stage would meet in small content teams to plan, reflect and adjust as well as observe each other. By participating in this coaching model, the new AFDA teachers were operating in this first phase. For example. Schoenfeld’s Assessment Project model: Teaching for Robust Understanding of Mathematics offers teachers a rubric to use to assess practice across five dimensions (Schoenfeld, 2014). In this study, I selected Boston’s Instructional Quality Assessment Instrument that gives feedback on the domains of Academic Rigor and Accountable Talk across ten dimensions (Boston, 2012). I selected this instrument, because it focused on keeping the cognitive demand of student thinking high during task implementation through student reasoning during classroom discourse. Another benefit was that the IQA tool was familiar to our teachers, having its roots in the Task Analysis Guide framework developed by Stein and other researchers (Stein et al., 2009).

Teachers would examine their own practice by using student work to promote collective inquiry (Kazemi & Franke, 2004). I would place the experienced AFDA teachers in this part of phase two. By year five, AFDA teachers had begun to use student work from tasks and projects to reflect and adjust in an iterative cycle. Examples were their functions, statistics and probability showcase projects. Secondly, they had been
discussing better measurement tools for such workplace skills such as teamwork, communication, and problem solving.

The third phase would require even more as teachers collaborate with researchers to connect their teaching to larger social, cultural or political issues. In this stage teachers might take on action research or become part of a larger study. AFDA teachers at this stage serve as mentors or coaches to teachers coming into their PLC from other divisions. Initiating a group of teachers to a community of practice or expecting action research in the first year of implementation probably would be beyond their zone of proximal development (Vygotsky, 1978, cited by Davydov, 1995). The experienced AFDA teachers had been operating at this level by presenting at conferences and participating in the previous AFDA goals and ensuing action research such as the one on how best to teach quadratic functions. Each year, AFDA teachers set a new goal based on data as a team, collecting data to measure growth.

**Reasons for ineffective professional development.** Thus, mandating change to teachers who either have not yet bought into reform or are at stage one in the Stein and Brown model, will not work, because they either need to feel cognitive dissonance that current practices are not working and/or need to be coached by others. Following this adage with my group, I would need to use the experienced teachers at stages two and three to coach the teachers who are at stages one and two. As a facilitator, I would need to figure out how that would look in different sites and contexts. As Sowder (2007) succinctly says, success equals teacher change. Twenty years ago, Darling-Hammond warned us that change is a process, or growth over time. Consequently, I would expect
that full implementation of our AFDA goals would easily take five years. This video lesson study is part of that iterative professional learning cycle.

Sowder, in 2007, describes the goal of any professional development, including preservice training, to prepare teachers to be continuous learners of teaching practice. Until the university course designed for AFDA teachers, only one out of eight of my teachers had taken any significant PD on pedagogy in math education in their careers, which ranged from 2 to 24 years. Interestingly, the youngest teacher with a recent master’s in teaching (MAT) was the most successful at implementation and developed most of the common assessments and tasks. The general education teachers who served as defacto leaders at each school were the ones with master’s degrees, even if in other areas. The teachers who had the most trouble adjusting were the ones without recent course work in any area and who did not yet have advanced degrees. Motivating teachers to continue learning goes back to the first step of creating the cognitive dissonance in order to create the need for the new learning. Add in the mobility factor year to year, and the need for a differentiated professional development response is guaranteed. Clearly, differentiation of teacher professional development is as needed as differentiation to meet diverse needs of students in the classroom. Providing that level of support is a challenge. Limited professional development funding, only six hours of in-service time with teachers outside of this grant, and responsibility for multiple sites and projects would hinder such individualized professional development. The use of peer coaching is the one strategy I have relied on this year; limitations have been lack of common planning at one site and productive common planning at the other two. Most importantly, teachers view
that mentoring as extra time they do not have to give for free. In her dissertation, Ma
cautions that American teachers do not have time to plan carefully; she goes further and
says that what we expect of our teachers is “unreasonable” (Ma, 1999, p. 276).

Hargreaves (1995) concurs; teachers will reject professional development if the
requirements are imposed or overwhelming. Teachers will not participate if they are not
involved in development, information is not directly related to their daily work, or is
composed of one-shot workshops. Affective factors of fear of failure or of being used as
an example also play a role in rejection. Hargreaves summarizes this well by saying it is
not that the professional development contains bad knowledge; it is a matter of it not
fitting into a context of particular teacher identities, moral purposes and an existing
culture (Hargreaves, 1995). This video lesson study model seeks to provide professional
development that is personalized and sustained over time. The digital online tools make it
reasonable to access with flexible times and choices of partners and devices.

**Collective math learning.** Cobb is more hopeful in his 1999 description of
teacher professional development in emergent terms. He says that the sequence would
begin with collective development of an idea such as inquiry-based instruction, and the
use of analysis of classroom practice to support that development. Implementation is seen
as an idea-driven application. He describes researchers supporting a group of
collaborating teachers as they experiment with an idea by analyzing it, adapting it to their
context, followed by testing and refining it. In this model, teacher participants helped to
create a better model. Three examples were the creation of a viewing form, expansion of
collaboration to include Google + and the addition of nonmath observers from other
content areas and outside of the AFDA team to give feedback. Ball and Cohen (1996) agree in saying that research-based instructional sequences support both teacher and student learning. Incidentally, Cobb says that collective mathematical learning does not deny individual initiative and creativity (Cobb, 1999).

**Pedagogical learning.** In 1989, NCTM described a vision for mathematics to be taught through problem solving with students reasoning and justifying their understanding and communicating their thinking through classroom discourse. Add in the emergence of technologies, which have been aptly described by Kaput (1992) as a “newly active volcano; the mathematical mountain changing in front of our eyes” (1992, p. 515). NCTM (1989) gave graphing calculators credit for the emergence of a new classroom dynamic where students and teachers partnered in developing mathematical ideas and solving problems. Today most teachers additionally use multiple devices such as Smart Boards, tablets, iPods, data collection equipment, computer applications such as Google Apps for Education and graphing calculator applications. It is an arduous task for teachers to keep up with the radical shift in how mathematics is taught in an inquiry-based classroom in addition to learning how to use its new tools for problem solving. Professional development must help traditional teachers transform their classroom pedagogy on both counts. In this study, new, special education and experienced teachers learned to facilitate student teams to solve problems through investigation, using technologies to simulate, form predictions, collect and analyze data, calculate, and present findings leading to decision making. Collaborative team content meetings were held for two days in the summer prior to this study and at the end of the year to allow
new and experienced teachers to listen and learn from each other. Teacher participants gave specific feedback based on instructional goals for facilitating classroom discourse and keeping the cognitive demand high during task implementation.

**Modeling.** A modeling approach to the teaching and learning of mathematics shifts the focus of the lesson from finding a particular solution to one problem to creating a system of relationships that can be generalized and reused in other problems (Doerr & English, 2003). The importance of quantitative reasoning in the K-12 mathematical experience has increased during this past decade due to the need for an informed citizenry and higher workplace demands for ability to interpret and communicate implications from data (outputs) from a computer or other technological tool (Doerr & English, 2003). Their study asked middle school students to develop models to evaluate, select, and rank data for the purpose of decision-making. The real-life context for data analysis and higher level thinking involved in ranking and decision-making are central to 21st Century skills. A second attribute of these tasks was that they contained a powerful mathematical big idea, which students could generalize to other problem settings (Doerr & English, 2003). I included this study not just because it was about statistics, which is a large focus of the AFDA curriculum, but because it was a great example of the reform mathematics idea of having students look past a particular case to generalize. The authors described models in the following way:

Models are systems of elements, operations, relationships, and rules that can be used to describe, explain, or predict the behavior of some other familiar system. The
modeling process begins with the elicitation stage, which confronts students with the need to develop such a model (Doerr, 1997; Lesh & Doerr, 2003).

Lannin, Barker, and Townsend (2006), list factors of algebraic generalization strategies: input value, structure of the problem task, prior strategies used, visual imagery of the task, and social interactions with teachers and other students. This aligns with Cobb’s argument that learning is a process that is an integration of individual construction of meaning with the social constructivist model of sociocultural conceptual development (Cobb, 1995). Cobb et al. said even earlier in 1992 that what is important is not whether students are constructing, but the quality of those constructions. That links to the question about balance between procedural vs. conceptual understanding. Ma (1999) says that students need to know both how and why. In Adding it Up, the NRC defined understanding as having five components: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. Again, Edelson quotes Whitehead from 1929 when he talks about learning for use. Like many others, he argues for integration of content and process in an inquiry-based approach (Edelson, 2001). The investigation provides a context, engaging the student. Cobb gives a statistics example of the relationship between a students’ reasoning in problem solving and the practices he is engaged in doing (Cobb, 1999). When the student comes to a limit in his own knowledge, in the investigation, he becomes motivated to acquire new skills or knowledge (Edelson, 2001, Schank, 1999). Thus, the learner must initiate learning!

Teachers as learners. Putnam and Borko (2000) took the research on situated learning for students and applied it to teachers who were learning new ways of teaching.
They noted that less attention had been paid to teachers learning new methods than students and that the process was essentially the same. They used the three themes of situating the new learning in a particular physical and social context, recognizing that learning was social in nature, and that learning was distributed across the teacher, other persons in the school and the tools used (Putnam & Borko, 2000). In their study, teachers learned new roles in creating the learning experiences for their students. The researchers used videotapes of classroom lessons and teacher journals to reflect and student notebooks of work samples of problems to embed the new learning in the teachers’ practice. Borko, Jacobs, Eiteljorg, and Pittman in 2008 showed a model for using videotape to analyze student discussion as I am doing in this study. Recently, Borko, Jacobs, Seago, and Mangram in 2014 facilitated teacher professional development by discussing videos of classroom lessons. My video lesson study design has similar components of context, collaboration and distributed learning, with the collaboration and feedback delivered virtually through convenient digital and online media now available across time and space.
Chapter Three

This chapter describes the methodological components of the interactive design of this mixed-method study. The discussion begins with a restatement of the goals and questions that frame the research design. The following sections are in the chapter: a mixed-method orientation, qualitative and quantitative research design elements, participants and settings, data collection procedures, data analysis, validity, and limitations of the design.

The purpose of this study was to explore one model of personalized coaching of high school mathematics teachers through digital video lesson analysis. Specifically, teachers targeted how to keep the cognitive demand high during the trajectory of a performance task. The target outcome was to increase opportunities for students to engage in reasoning and communication of mathematical thinking through the classroom discourse occurring during tasks. The following research questions were designed using a mixed-method interactive design (Maxwell, 2002, 2013).

Research Questions

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
   a. For what purposes did high school mathematics teachers select tasks for students?
b. What did high school mathematics teachers want to change about task implementation in their classrooms?

c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?

d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?

e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?

a. How did teachers describe the change process in their own task implementation?

b. Which coaching activities did teachers perceive as leading to growth?

c. How did the learning process compare across three teacher groups: new, special education, and experienced?

d. What suggestions did teachers offer to improve coaching activities in future models?

Research Design

The design of this study was mixed methods using three comparative case studies. I collected data from multiple sources in order to answer each research question: pre/post teacher beliefs survey, pre/post teacher video coaching reflection, lesson videos, the selected mathematical tasks, and observation Instructional Quality Assessment (IQA)
rubrics (Boston, 2012). I integrated all evidence when interpreting data in order to better understand task implementation and to increase trustworthiness of the findings. The design offered flexibility to explore interesting questions emerging from data from the different sources. A process approach made more sense in this comparative case study for examining classroom discourse occurring during task implementation in the context of individual teachers’ classrooms (Maxwell, 2013). Thus, the open-ended survey questions, video analysis of observation data, and teacher reflections were analyzed using qualitative methods.

In seeing the connections between qualitative and quantitative research as an integrated system, I used results of the qualitative data from the teacher lesson reflections, paragraph text survey items, observations, and selected math tasks to better understand the quantitative data surrounding the surveys and Instructional Quality Assessment (IQA) observational rubrics. Triangulation of evidence was used to answer research questions with multi-layered responses. However, the most important reason for using a mixed-methods design was to obtain richer data in order to study this model across more contextual dimensions. Thus, mixed methods were used with a dialectic paradigm in the study design as well as in the interpretative process with validity checks. Finally, a comparative case study of three representative teachers painted a picture of task implementation in high school mathematics classrooms during the video lesson study experience. Three teachers fell into each of three groups: experienced, new to AFDA, and special education.
The goal of the quantitative sources was to find any associations between task selection and implementation practices, and any of the 10 dimensions of the domains of academic rigor and accountable talk moves, rather than explaining how or why. Numerical analysis was more appropriate for the survey questions answered with multiple choice, scale, or a check list, the ranking of mathematical tasks, and observation rubrics. Frequency data were matched with physical causation (Maxwell & Loomis, 2003) in the descriptive data to explain what the numbers meant in the context. The qualitative data sources were also used to describe how teachers responded to the video lesson analysis format and their perceptions of which practices were valuable in terms of professional growth. I looked for alignment between the survey data and the qualitative observations. I had planned to create emerging questions from any discrepancies between data sources and answer those by going back to the original classroom video footage. However, I did not find any discrepancies. I selected five emerging themes of interest from the multiple data sources to study across three groups of interest: new, experienced, and special education teachers. The following themes were explored in the comparative case study: task selection vs. task implementation, student engagement: participation patterns in classroom discourse, teacher questioning, teacher vs. student moves, and engagement in the coaching process.

Situating the methodological choices. Design Research was used as the primary methodological framework in creating the AFDA course curriculum with a mixed-methods design to study task selection and task implementation in this study (Hjalmarson, 2008; Hjalmarson & Lesh, 2014). Just as with the overarching course
design, the evolution of the video lesson analysis study as a professional development model followed the same design principles. An iterative cycle of the following steps was used: problem identification, creation of a conceptual framework, curricular design, identification of systems of use, creation of a prototype, trials, testing and intervention, implementation of models and generalization of the course design to other settings (Hjalmarson & Lesh, 2008; 2014; Middleton, Gorard, Taylor, & Bannan-Ritland, 2008). The results of this study will be used as evidence to continue to adjust the video lesson study prototype as a model of embedded professional learning.

This study sought to provide information about how teachers study their own mathematical task selection and implementation by analysis of video of their lessons with peer coaching feedback. The academic rigor of the task itself was analyzed with IQA rubric criteria across then dimensions. The research questions studied how the accountable talk generated by teacher and student interactions moved a task along its trajectory. Either the task sustained a high cognitive demand, or it was reduced to a lower-level task (TIMMS, 1999). Of particular interest was the quality of the discourse between students and teachers as well as between students.

Study teachers each set a personal goal from the academic rigor and accountable talk domains after viewing and self-assessing a video of a task they had recently taught. During this goal-setting process, teachers collaborated face-to-face in pairs or in triads to give rubric feedback to each other and to discuss and agree on goals. Once a goal was identified by teachers from a baseline video, the question was how to support teachers along their individual trajectories to improve capacity to keep a high level of student
thinking as the task unfolded. These teachers collaborated during four lesson cycles for two purposes: to meet a personal instructional goal and to help peer teachers reach their goals by providing coaching feedback on a common set of criteria. Teachers also knew they were part of the design process of creating and improving this model of professional development by providing survey and feedback during coaching conversations. Study findings have been used to adjust local professional development practice of the video lesson study model for the following year. Locally, the findings from this study were used as a part of program monitoring of the division math goal to implement performance tasks.

**Researcher’s identity.** As a researcher and math supervisor, I was interested in capturing the thoughts and practices of a group of teachers from one school division who are implementing a task, problem, and project-based course in order to understand how to improve task implementation as evidenced by student reasoning in classroom discussions. The goal was to embed effective teacher professional development into teachers’ peer collaboration through real-time formative feedback by nonevaluative observers such as fellow teachers and instructional coaches. In my role as math supervisor, this study dovetailed with the larger AFDA course implementation design. With multiple sources of data used to measure professional development impact on students and teachers in terms of task selection and implementation, this study added information on how to design embedded professional development as part of ongoing course implementation. I used each of these case studies to reconstruct the story of the professional development effort through both teacher and researcher perspectives. To do
this, I focused on teacher and student behaviors during task implementation. I selected themes from questions of interest emerging from the analysis of the study’s five data sources. I then looked deeply at these questions by coding the classroom videos to describe the experiences of three representative teachers in a comparative case study.

**Qualitative design orientation.** The philosophical view of qualitative methods that I chose for this study was one that included a variety of methods and forms of interpretation that were naturalistic, historical, emerging, critical, shifting, and changing (Lincoln & Canella, 2004). This view mirrored the educational theory of social constructivism that underlies the agenda for the AFDA curriculum and the instructional pedagogy used in classroom discourse during student mathematical tasks. While the quantitative methods and analyses were prestructured to answer the research questions, the interpretive analysis from the qualitative sources was inductive with no predetermined categories (Maxwell, 2013). I used the qualitative part of the study’s design to understand the process of how the phenomena of task implementation evolved in the classrooms of nine high school math teachers who were intentionally studying their own task implementation.

Since AFDA is delivered differently from traditional math courses by teaching through the inquiry provoked by instructional tasks, it was crucial to understand where individual teachers on the AFDA team were operating in the selection of and implementation of performance tasks. I also wanted to know how teachers were responding to this paradigm shift in pedagogical expectations through the evidence provided by this model. Lastly, I wanted to determine if a digital video lesson analysis as
a professional learning model was both effective and realistic for teachers in a public high school setting.

**Mixed-Methods Design**

The design of this study was mixed methods using a comparative case study. I collected prestructured data from multiple sources in order to find what each one showed me about each research question. I integrated all evidence when interpreting data in a more inductive approach in order to better understand the phenomenon of task implementation in a secondary mathematics classroom. The design built in flexibility to explore relevant emerging data including discrepant data from different sources through the lens of three representative teachers by checking back to alignment with original video footage and compare to IQA ratings, open-ended reflection survey items and coaching feedback.

**Multiple data sources.** Hufferd-Ackles et al. (2004) collected data for their study in three phases: observations of four teachers each by three different observers, a case study of one teacher, and a comparison study to four other teachers. She then triangulated the data to increase trustworthiness of the findings. In a similar fashion, I had 10 study teachers and 19 peer observers, in addition to the researcher, watch and code videos over four lesson cycles for a total of 114 observations. Each teacher in the video lesson study completed a self-assessment and a range of four to six peer observations on each lesson. These peer observers viewed the lesson videos from private YouTube links on the lesson study site. They used an electronic viewing form to tally and record question types asked by students and teachers along with question examples and suggestions. Observers used
the data from the viewing form to complete the 10 IQA rubrics as part of an electronic feedback lesson survey using a Google forms link embedded on a Google site. These observers, the researcher, and the teacher in the video, all completed and sent a survey containing the rubrics, along with reflections and descriptive feedback, to teachers for feedback as part of each lesson cycle.

Boaler’s 1999 case study offered a clear picture of the multiple sources of data we could have chosen to collect in this case study of teacher professional development. In her studies of a cohort of students in two different United Kingdom schools, Boaler used seven teacher interviews, 80 student interviews, about 100 one-hour lesson observations, 300 questionnaires with both open and closed questions, 305 contextualized free-response assessments, 104 architectural projects and tests, 61 long term learning tests, 188 flat design activity and tests, and 290 student answers from a national examination. Her extensive data collection is of interest, because it studied a similarly-sized student and teacher population as my AFDA cohort and investigated the influence of the classroom as a community on learning. Her study was useful in that it was a comparison study between schools having different ideas about classroom culture which is similar to the comparative case study between teachers from three schools in this study, representing three different roles: (1) the experienced teachers having over five years of experience; (2) the new teachers in the first year of teaching this course; and (3) the special education teachers coteaching in an inclusion or self-contained math classroom. Boaler’s study helped me see what a comparative case study over three years might look like.
A process approach made more sense in a comparative case study for examining the context of math classrooms and individual teachers (Hammersley, 2000; Maxwell, 2013). I compared the experience of the study across three groups: experienced AFDA teachers, new teachers and special educators. Thus, the beliefs surveys, video analyses of a common teacher, teacher lesson reflections and observation data were analyzed using qualitative methods. Numerical analysis was more appropriate for the survey questions answered with multiple choice, scale, or a check list. These frequency data were matched with physical causation (Maxwell, & Loomis, 2003) in the descriptive data to explain what the numbers meant in the context. Events of interest in the survey data and the IQA rubrics, created emerging study questions, which were used to delve further into the classroom video footage to explain and connect to other sources during the case studies.

In seeing the connections between qualitative and quantitative research as an integrated system (Maxwell, 2013), I used results of the qualitative data (the beliefs survey, video classroom footage, and teacher and peer observer written reflections) together with the quantitative data (the survey frequencies, comparison of means, and correlations between the ten task implementation dimensions from the IQA observation rubrics) to answer the research questions. Triangulation of evidence was used to answer research questions; both connections and discrepancies between data sources were used to form emerging study questions. Thus, mixed methods were used with a dialectic paradigm in the study design but also in the interpretative process and validity checks.

The unit of analysis in this study was the teacher. I wanted to know how the cognitive demand in task selection was related to the choices of teacher moves and of
questions posed by teachers and students during implementation. Research Question 1, parts d. and e. evaluated the changes in the academic rigor occurring during classroom discourse over the course of the lesson cycles. Teachers clarified what was working and not working in the model to help them build capacity to keep tasks at a high thinking level. Multiple data sources helped to evaluate impact of professional development activities on both teacher beliefs about performance tasks and on the actual student reasoning occurring during classroom discourse. Emerging data and resulting questions provided insight into validity threats and study limitations. The results were used to direct future adjustments to the video lesson analysis model in the division and to inform related studies.

Quantitative methods. The quantitative data sources consisted of selected questions on the pre- and postsurvey questions, the ranking of the selected tasks in the lessons for cognitive demand, and the IQA observation rubric scaled responses across 10 dimensions.

Surveys. The presurvey was administered during the late summer preceding the school year of the study, and the postsurvey at the end of the school year in June (Appendix A). These were the first and last activities in the study. The surveys examined teachers’ current beliefs and practices regarding selection and implementation of mathematical tasks in their classrooms. The prestudy survey asked 11 questions including the following: task definition, characteristics of rich tasks, task purposes, frequency of use, purposes of class discussions, question types, student participation, student reasoning, and how students typically showed work. Lastly, the survey asked teachers to
reflect on what they would like to change about the way they use performance tasks and about their classroom discussions. Of the survey questions, two were ordinal and four items were nominal and could be reported as frequencies. The remaining five were open-ended and reported in qualitative results.

I gave this same survey to all mathematics teachers during the preschool days prior to the start of the school year as part of division-wide implementation of performance tasks. Study participants would have taken the survey regardless of their participation status. I chose to write my own survey based on division needs in terms of teacher readiness for implementation of performance tasks, which were being implemented in the year of this study. The purpose of the postsurvey was to formatively check where teachers were with task selection and implementation, noting changes in the frequency and use of tasks in order to plan for next steps in division implementation the following year.

The postsurvey contained the same 11 questions as the presurvey as well as three additional questions. One of these questions was nominal with a checklist that could be reported with frequencies: Which of the professional development activities/tools (in the checklist) were useful to you to improve something in your teaching?

**Task analysis guide.** Division teachers were familiar with Stein and Smith’s Task Analysis Guide (TAG) (1998, 2011) from previous division professional development. Participant teachers had experience using the TAG to evaluate cognitive demand and to modify unit performance tasks. Therefore, it was simple to ask them to evaluate the potential of each task in the video lesson study using the Instructional Quality
Assessment (IQA) rubrics for Academic Rigor on a continuous scale of increasing quality from one to four. Melissa Boston (2012b) built on Stein and Smith’s original TAG to write the instrument (Boston, 2012b; Boston & Smith, 2009; Smith, Hughes, Engle & Stein, 2009; Stein et al., 1996). The IQA rubric on Task Potential used in this study has the same levels of cognitive demand and similar language in the descriptors as the Task Analysis Guide for lower and higher level thinking. At their first meeting, teachers used the familiar TAG and the new Academic Potential rubric from the IQA tool to code the tasks used in their baseline videos as well as practiced ranking the task used by their partner after viewing the lesson video. Partners compared rankings and had 100% agreement on the baseline. Categorical data on academic potential of task rankings (1-4) were considered as discrete categories of performance when reported out in frequencies. The accountable talk rubrics were written as a continuous scale of increasing quality from 0 to 4. A score of zero meant that no classroom discussion occurred. I reported the data both ways.

**Instructional quality assessment rubrics (IQA).** The classroom observation tool, written by Melissa Boston and used with written permission, was the instrument used by all teachers and coaches to provide feedback on the lesson videos. The instrument scores two overarching areas: Academic Rigor and Accountable Talk. The dimensions under Academic Rigor included: (1) Task Potential; (2) Task Implementation; (3) Student Discussion Following the Task; (4) Questioning; (5) Mathematical Residue. The dimensions under Accountable Talk included: (1) Participation; (2) Teacher’s Linking; (3) Students’ Linking; (4) Asking (Teacher Press); (5) Providing (Student Responses).
All IQA mathematics lesson observation rubrics and checklists were reported as frequencies over the four lesson cycles. I ran a comparison of means by lesson to observe the change in ratings over time, followed by correlations of IQA variables in order to look for associations.

**Qualitative methods.** The qualitative data sources consisted of: (1) open-ended items on the pre- and postperformance task beliefs surveys; (2) the pre- and poststudy written coaching analysis of a common online video math lesson taught by an Algebra I teacher; (3) classroom video footage; (4) Instructional Quality Assessment instrument data across 10 dimensions; (5) ranking of classroom tasks for cognitive demand using the Academic Potential rubric from the IQA instrument (Boston, 2012).

**Survey.** Following an orientation to the study, the first activity for teachers was to take a beliefs survey on the selection and use of mathematical performance tasks. All division mathematics teachers had taken the same survey during the preschool week, because our division math goal was to implement performance tasks in every unit of every mathematics course in Grade Six through Algebra II. The division purpose was to increase student problem-solving and reasoning through classroom discourse during the tasks. Thus, this study worked in tandem with the concurrent professional development initiatives existing in the division.

Five out of the 11 items on the teacher prestudy beliefs survey on performance tasks were open-ended and coded using qualitative methods. Data from the following questions were coded using open axial coding and then categorized into themes: (1) How would you define a math task? (2) What makes a task “rich”? (3) For what purpose do
you use class discussions? (4) What is the number one thing you would like to change in
the way you use performance tasks? (5) What is the one thing that you would like to
change about your classroom discussions? The poststudy survey asked the same
questions as the presurvey along with three more. Two of these “poststudy-only”
questions were open-ended: (1) What was the most useful part of this professional
development? (2) Please suggest possible improvements for other teachers studying their
classroom tasks and discussions. I compared the themes from these seven questions to
those generated from the same teachers from the video feedback written reflections of the
lesson of the anonymous “Barbara” teacher. As a second approach, I then used a matrix
to locate connections between and among the teachers from these two data sources. I
selected a connecting strategy to locate relationships between the data sources rather than
fracture the data into discrete categories and recombine into themes (Maxwell, 2013). I
wanted to see emerging themes from both inductive and deductive approaches.

**Video peer feedback.** During the initial summer professional learning and during
the final face-to-face meeting, the AFDA teachers viewed a common video of an Algebra
I teacher teaching a lesson on quadratic functions from a mathematics professional
development site. Following the video, each teacher wrote a peer coaching reflection and
uploaded it to a Google Docs folder or submitted it via email. The purpose of this video
reflection was to collect baseline and poststudy data on teacher beliefs about what math
classrooms should look like in terms of task selection, implementation and within that,
classroom discussions. These data would be compared and coded against self-reported
pre- and postsurvey data. Gauging teachers’ beliefs about effective instruction is complex
(Jacobs & Mirata, 2002). The idea was that teachers would better show their true beliefs through application in a context than by answering a general pedagogical question. The context arises from the task of providing written feedback on another teacher’s specific video lesson. Self-reporting in a survey would hold more risk of bias with teachers giving what they believe to be the correct response. By triangulating these two data sources for the same purpose, I felt I could get a more accurate measure of teacher beliefs about task implementation and, specifically, how a discussion should look.

Determining teacher beliefs is tricky, so I did not want to rely solely on self-reported survey data. Jacobs and Morita (2002) compared Japanese and American beliefs about classroom instruction by having teachers evaluate a video lesson. Study findings indicated that participants’ coding of a common video in a context was more reliable than a self-survey of classroom practice. I adjusted the Jacobs and Morita study procedure as a model to design the second data source for getting at the beliefs on the use of performance tasks in this study. I used data from the pre- and postbeliefs survey and a video coaching analysis to describe teachers’ beliefs about task selection, implementation and classroom discourse as well as teacher and student lesson moves.

Teachers watched an Algebra I teacher named Barbara teach a quadratics lesson and wrote a peer coaching reflection for the teacher. This assignment was done in the summer before the lesson cycles began to give teachers experience with giving peer coaching feedback on a content topic from their curriculum (quadratics) before the actual data collection of their own lessons began. The exercise was also designed to lessen the anxiety about giving and receiving feedback, since the teacher was an unknown person
from a professional development video. I then correlated results to the survey to check for discrepancies and alignment. I utilized the same procedure with the postsurvey and postvideo reflection of Barbara. Using multiple data sources gave me a more reliable picture of each teacher’s beliefs about task implementation, highlighting teacher moves and student discussion before and after the professional learning. Secondly, I used these sources to add to the self-reported data on what teachers wanted to change with regard to using tasks in their classrooms.

**Baseline video.** At the initial two-day summer professional development, teachers analyzed a personal baseline video taken at the end of the previous school year in order to set goals. Before taping, the only direction given to teachers was mathematics to record on a day when a task was being used instructionally and to include in the footage of the classroom discussion that followed the task. In the summer collaboration, I explained the IQA rubrics to the teachers and had each teacher practice applying them to their own video and a partner’s lesson. The teachers compared the scoring of the rubrics and talked about any differences. As a team, teachers then came to agreement on what the rubric levels meant. Teachers completed a self-reflection on their baseline video and set a personal goal for the study from one of the 10 dimensions in the IQA instrument. In this way, the IQA rubrics were used as a baseline and as a training tool for cohort teachers and peer observers on the use of the IQA instrument to achieve a measure of inter-rater reliability. Each peer observer was also trained on the instrument for the same purpose.

**Lesson cycle videos.** Classroom video footage was taken for each of the 36 lessons and shared via a private YouTube link on a Google site within the division’s
Google domain, so that it was secure for teachers and students. I emphasized to teachers that this project was situated in a safe, risk-free environment for self-reflection and peer coaching. Key to this safety was the creation of a cohort of teachers who would observe one another. The site itself was shared only with cohort teachers and any peer observers they invited to participate. I made a point of saying that their evaluating administrators would not be participating. Each teacher analyzed his or her own lesson using the IQA rubrics. Additionally, four to six peer teachers served to provide coaching feedback to each teacher for each lesson. Teachers were responsible to seek out additional peer observers in addition to the teachers in the cohort.

Directions. The observers tallied question types and sample question examples along with suggestions on a video viewing form, which was created by one of the teacher participants. Next teachers used these data to code the IQA rubrics and submitted data to the researcher via a Google form. Finally, observers emailed a copy of the survey form to the teacher of the lesson. The IQA observation rubrics were the key instrument used to code the lesson observational data from peer and self-observers. The classroom lesson footage was used as the initial data source from which the rubrics were coded. As a primary source, these videos were used to check emerging questions of interest or discrepancies that arose during the data analysis of the other data sources.

Access and permissions. I submitted a form for a request for data collection to the Director of Program Evaluation and Assessment of the school division and was approved before beginning data collection. I submitted and received International Review Board approval through George Mason University (Appendix B). The participating
teachers signed a consent form (Appendix B) to participate in the study. As some of the professional development activities were a part of the existing school division’s professional learning, the teachers received license renewal points for any completed portion, regardless of participation in the study. Following the completion of the study, teachers each received a certificate and license renewal points for participation.

**Context and Participants**

**Context.** This section describes the rich context of this study, including division, summer professional activities, classroom setting/resources, and duration.

**Division.** The Algebra Functions and Data Analysis (AFDA) course design process took place in a rural-suburban public school district of about 13,000 students. During the year of this study, about 312 students were enrolled in 14 sections of AFDA from three base high schools along with additional students from long-term suspension and other alternative programs.

AFDA teachers had been collaborating as a content learning team four to six times per year over the past five years of new course implementation. The local school district had funded these collaboration days. These original team members had been functioning as a community of practice with autonomy. For example, the team had created a Google website and had a volunteer teacher who sent out minutes and meeting dates. With the recent influx of new AFDA teachers, a new structural team identity had not yet been formed. Outside of the research questions in this study, I wanted to examine the process of how this new teacher team learned to work together through this lesson study to implement tasks and provide feedback to each other through digital media.
**Summer professional activities.** During the summer prior to the study, division mathematics teachers met with their peers by course team to continue professional development on modifying, selecting and implementing performance tasks. This video lesson study was part of the division math professional development plan and tied to the division math goal for students to become problem-solvers and reason through the classroom discourse occurring during rich tasks. Teacher representatives collaborated that summer to write tasks for every unit for seven courses ranging from Grade Six to Algebra II. The AFDA teachers met for two days as part of this larger project.

During this summer collaboration and in preschool meetings, teachers of all mathematics courses from Grade Six through Algebra II practiced the components of the task cycle related to keeping cognitive demand high for students. Examples of lesson planning activities included the ranking of tasks for cognitive demand and selection of tasks based on alignment with content standards, process goals and cognitive depth. Teachers also discussed how to modify tasks to increase the cognitive complexity. During the preschool inservice, all division teachers participated in a task implementation cycle to orient them to the process of using the performance tasks selected by teacher representatives during the summer. Led by the teacher leaders, participants solved a selected task themselves in two different ways and shared out with peers different solution paths and strategies in the group. Teachers worked in groups to check for vertical alignment with the state curriculum framework and identified the thinking levels on Bloom’s Taxonomy, Webb’s Depth of Knowledge, and the cognitive demand on the TAG. Teachers predicted student misconceptions and planned next questions to ask to
move learning forward based on the formative feedback. They recorded on-the-spot student interventions for typical misconceptions. Thus, the AFDA teachers had already been a part of this professional learning on mathematical task design and selection during this particular summer. The video lesson analysis study offered these teachers embedded coaching on the next step of task implementation.

**Classroom setting/resources.** Division math classrooms were each equipped with an electronic whiteboard and a laptop/projector presentation station equipped with math software including Smart View for graphing calculator display. All AFDA students had access to a TI-84 and TI-Nspire graphing calculator either from a class set issued to the teacher or from a personal one issued to or owned by the student. Teachers and students had access to Lab Pro on the network, Lab Quest and CBL-2 units and probeware for data collection purposes. A dedicated math lab and class set of laptops was available at each school. The alternative program had laptops for each student. All buildings offered wireless connections and allowed students to use personal devices. Study teachers had Google accounts within the protected division domain to allow for easy use of private YouTube, Google+, Google Forms and Sites to upload and view videos and to submit digital observation rubric forms for feedback.

The students did not have a print copy or digital license for a textbook as the course had been written as a task, problem and project-based course. Teachers borrowed lab equipment from the mathematics and science workrooms when needed and used manipulatives such as Lego blocks, color tiles, dice and spinners gathered from various free or low-cost sources.
**Duration.** This mixed-methods study collected course design and implementation teacher and classroom data over one summer and during the following school year. Activities spanned two summer professional development days in July, a preschool meeting in August, virtual collaboration over four lesson cycles during the school year and a final face-to-face debriefing the following June at the end of the school year. Thus this professional development activity fit into the division’s existing structure for professional development and issuance of license renewal points by June 30 during a given school year. This school division had been collecting AFDA course design and program implementation data for five years, including the school year of this study.

This study was situated within the overarching professional development design for course implementation of AFDA within the division. Due to the small population of AFDA teachers, purposeful selection was used to invite all existing AFDA teachers at the time of the study to participate. The teachers would have had the same professional development regardless of participation in the study. One AFDA teacher had an assignment change, and he stayed in the study. Another teacher completed two lesson cycles before leaving mid-year for a different job. He was not included in the study. Purposeful selection was used for the case studies to achieve representation of all three different teacher groups: new, special education and experienced (Maxwell, 2013).

**Participants.** This section details all the 10 teacher participants, as well as contextual information on their students and the researcher.

**Teachers.** Ten AFDA teachers began as participants in this study. One teacher left in January to take a new job. Since 7 out of the 14 sections were cotaught, the original 10
teachers formed 6 teaching teams. Six out of the 10 teachers were new in their AFDA assignments, being in years one or two; this resulted in the need for this professional learning. Three were special education teachers and seven were general education math teachers. The three special education teachers were endorsed in K-12 special education. None had a mathematics endorsement. All general education math teachers were endorsed in 6-12 secondary mathematics. Six out of the seven general education teachers were mathematics majors. Five out of 10 study teachers were endorsed in special education; 2 general education teachers were dual-endorsed in both mathematics and special education. At the beginning of the study, six teachers were male and four female. The mean age was 34.2. Five, or half of the teachers, had a master’s degree, and one was in the dissertation phase of a doctoral program. The mean number of years of completed teaching experience was about nine.

For data analysis purposes, the remaining teacher team fell into three groups: (1) Three experienced teachers of at least four years of teaching AFDA; (2) Three teachers in their first two years of teaching AFDA; (3) Three special education teachers assigned to coteach AFDA. These groups formed the three categories for the case studies. One representative teacher from each group was analyzed in depth.

From the experienced group, one original member of the first year teaching cohort of the AFDA course design process remained. However, this teacher’s schedule was switched in July after the study began due to a late resignation in the department. The teacher chose to remain in the study and was videoed in Algebra II and Calculus. Two teachers who began teaching in the second year of AFDA were also considered veterans.
A fourth experienced teacher began the study, completed two cycles, but then left the school to take a different position. This teacher’s data are not included in these results.

Five of the teachers had little to no experience in teaching an inquiry-based course with a problem-based approach rather than a textbook. Thus, the team as a professional learning community (PLC) was operating as a fairly new team. Three veteran leaders who had led the development of the course for the first three years had recently taken other assignments, leaving only two experienced AFDA teachers to serve as mentors. One high school had no veteran AFDA leaders and was operating with a long-term substitute for most of the study. Such mobility is typical in a public school division and a reason for the necessity of ongoing job-embedded professional development on the pedagogical math knowledge for teaching.

**Students.** About 300 AFDA students from year five of the course implementation were enrolled in the classrooms in this study. This cohort contained a disproportionate number of students with disabilities (S w/ D). Economically disadvantaged (ED) students came predominately from rural and migrant settings. The cohort was 82% white. The division percentage of Students with Disabilities was about 12% and Economically Disadvantaged (ED) about 30%.

**Researcher.** As a participant/researcher, I worked as the district mathematics supervisor in the division being studied. My position allowed me to serve as a coach without any evaluative role. I feel I have a strong working relationship with all participants. Students were used to seeing me in the classrooms in a coaching capacity. I have been the division leader of the AFDA course design process for the past five years. I
have been leading design/action research on this project since its inception. In order to mitigate participant-researcher bias, multiple observers coded each lesson video. Observers came from the nine teachers in the study and ten other peer observers to provide corroboration on all IQA observation rubric data. Each of the 36 lessons had three to six observers completing rubrics for feedback.

**Participant selection.** All current AFDA teachers were part of this study including both general education and special education team teachers. The sample was both a purposeful and an inclusive sample in that every AFDA teacher took part and purposeful in that that all three teacher groups were represented in the case study. All division teachers were implementing rich tasks this year, so professional development efforts on task selection and implementation had already been planned for teachers in all mathematics courses, including AFDA. The topic of this study was a part of the division mathematics goal to increase critical thinking through use of rich performance tasks used formatively during instruction along with data emerging from research questions from previous years of AFDA course implementation.

**Data Collection**

**Procedure.** The procedure for data collection was divided into several parts as explained in the following subsections.

**Prestudy data.** Teacher participants met face-to-face with the researcher to understand the overall purpose and components of the study as well as to complete consent forms (Appendix C). Teachers then filled out an online survey on their beliefs about mathematical performance tasks (Appendix A). The survey was designed to ask
questions about task selection practices, beliefs about cognitive depth, instructional uses of tasks, reasoning and classroom discourse. Model feedback included self-identified improvements needed related to task implementation. All study teachers then viewed an identical teaching video and analyzed it by commenting on the various components of the lesson. These video analysis data were collected digitally at any convenient time from any location. The researcher later triangulated the self-reporting of beliefs about performance tasks in the survey with the coaching reflection of a lesson video of a classroom task viewed from a professional development site. Teachers set personal goals from the areas of classroom discourse and student reasoning described in the rubric domains. The researcher and teachers completed four lesson cycles during the course of the school year.

**Prelesson cycle.** For each lesson, I collected the particular task being observed and situated it within the curricular units. The teacher, peer observers, and I evaluated the task on the Academic Rigor: Potential of the Task rubric (Boston, 2012b), which was adapted from the Task Analysis Guide (Boston & Smith, 1998, 2011).

**Lesson cycle protocol.** During the lesson, observers took observation notes, collecting data on a teacher-created video lesson viewing form. Peers and coaches tallied questions by one of six types: probing, mathematical meaning, generating discussion, procedural/factual, other mathematical, and nonmathematical (Boaler, 1999). The lesson was videoed with a choice of device: cameras, flip-cameras, tablets, and phones were used. Teachers at one school placed the camera on a bookcase and taped the whole
lesson, uploading the pertinent discussion piece. Some teachers had peers video their lessons while others had the researcher observe and record.

Following the lesson, the teacher or researcher uploaded the video to a private YouTube channel and shared the link via email or Google+. The teacher then viewed the video and tallied question types and examples on a viewing sheet created by one study member. The videoed teacher then completed a self-assessment of the lesson using the online IQA rubrics on a Google Form. The observers then completed the same IQA rubrics on an identical Google form. Some questions were only for the peer observers to provide feedback to the teacher while other questions were just for the teacher of the lesson. The observers sent a copy of their form directly to the teacher for feedback. All rubric response data were automatically collected in a Google sheet, a spreadsheet accessed by the researcher.

**Professional resources.** Each teacher and I discussed choices of research-based resources related to that teacher’s goal. Coaching communication took place via email, Google+ and face-to-face. The other peer observers offered coaching suggestions for the teacher to consider implementing in class over time and before the next video. In the summer PD, I had shared a bank of about 200 professional resources located on the division math site on the topics related to this lesson study including classroom discourse, task implementation and cognitive demand. After the baseline and first videos, I pointed the teacher to specific assignments. For example, after the baseline video, all participants were given the same article: *Never Say Anything a Kid Can Say* (Reinhart, 2000) to provide a vision for incremental improvement of classroom discourse. Reinhart
encourages teachers to select one or two changes to practice and refine before tackling others as a strategy for continuous improvement. After Lesson 1, teachers were asked to choose a resource to study and reflect on strategies that they would like to try. Teachers focused on an improvement goal in the video lesson study and practiced these few selected strategies each cycle, measuring growth by the IQA rubric feedback.

*Coaching feedback.* In future lessons, the coaching feedback was customized based on what the teacher needed from the IQA rubric and survey results. Peer teachers often gave specific strategies such as “connect the table to the graph sooner; what patterns do you see in the table?” Other coaching feedback was more general, “…give more wait time or individual think time to have students process and solve individually before moving to the pair/share part of group task exploration. “Notice how fast you jumped from ‘show me’ to ‘telling’ the big mathematical idea.” There were related videos by Jo Boaler, Lucy West, Peg Smith, Dan Meyer and others in the resources that shared questioning strategies for increasing opportunities for all students to access and think deeply about a question or task before the large-group discussion. One suggestion was to “focus more on the ‘Why?’ rather than the ‘What?’ in the content of the lesson.”

Thus, this community of learners coached each other using a common set of criteria but also offered the formative feedback specific to the actual context of the lesson.

In each lesson cycle, participant teachers observed a peer teacher in the study and used the common IQA tool to provide feedback. This interaction was meant for the teacher to practice looking for the same Accountable Talk moves in another teacher’s lesson. Each teacher self-assessed on personal goals from the 10 dimensions of Academic
Rigor and Accountable Talk domains from the online Google form. Additionally, the observed teacher received dimensions feedback and coaching suggestions from three to six peers in a safe coaching venue.

*Video lesson study analysis as division math professional development.* Summer participants also talked about task implementation in activities such as coding a video of an Algebra I lesson for teacher moves linked to student communication and reasoning. Teachers viewed a baseline video of a lesson they had taught, self-assessing with the 10 IQA rubrics and collaboratively coming to consensus on the IQA rubric scores for that lesson. Partner teachers then coached each other by suggesting alternate instructional strategies. Lastly, teachers reflected and wrote next steps for their goal. Teachers collaborated with the researcher and peer teachers for a total of four lesson cycles during the new school year following the same identified lesson cycle protocol.

*Poststudy data collection.* After the final lesson cycle, teachers individually filled out the poststudy survey on beliefs related to mathematical performance tasks, focusing on cognitive demand of the task trajectory and evidence of student reasoning and communication. The postsurvey contained all of the prestudy survey items and additional questions on self-growth. Questions asked which video lesson study activities were viewed as contributing to that growth and identified areas still needing support. The survey also asked teachers to evaluate reasonableness of the professional development as a whole in terms of time, submission of evidence digitally, feedback from the researcher and peers, and suggested changes for next uses (Appendix A). The researcher wrote a
memo on the same four areas adding an additional area examining reasonableness of staffing, stipends and resource costs to a division.

Teachers analyzed the same teaching video as in the prestudy reflection. Teachers each wrote a coaching reflection of strengths and feedback they would give that teacher. This type of peer feedback was familiar as it had been part of the study activities. Each teacher had provided peer feedback on at least four lessons. The researcher compared coaching statements evident from the video analysis to the self-reported belief statements in the postsurvey.

The final professional development event was a face-to-face half-day team meeting where teachers shared their feelings about the study and then completed a postsurvey and the video reflection. AFDA teachers were already meeting to collaborate on another project that day as part of their division collaboration. Thus, this video lesson study was embedded into standard team collaboration.
Communication Mode
- Digital
- Face-to-face
- Blended

Data Collection Framework

Prestudy Data (Preschool)
- Belief survey
- Video Analysis
- Goal setting

Four Teaching Cycles (School Year)

Poststudy Data (End of School Year)
- Belief survey
- Video coaching reflection
- Teacher Team Final Meeting

Researcher-Generated Data
- Task Analysis
- Video Viewing Form
- Lesson Video
- Goal-specific IQA rubric
- Coaching intervention notes
- Observation protocol

Teacher Generated Data
- Task Selection
- Task Analysis
- Self assessment: IQA rubric
- Peer observations
- Peer assessment: IQA rubric
- Task Survey

Planning and Task Selection
- Task selected (T)
- Task analysis evaluation (T/R)
- Set personal MKT goal (T)

Teaching the Lesson
- Video Viewing Form (T/R)
- IQA Rubrics (R)

Teaching Cycle
- Feedback/Reflection
  - Video analysis (T/R)
  - Goal-specific IQA rubric feedback (P/R)
  - Self-assessment: IQA rubric (T)
  - Researcher feedback on goal (T/R)
  - Plan next lesson cycle steps in goal (T)
  - PD assignment (T/R)
  - Peer Video Observation (P)
  - Peer-assessment: IQA rubric (P)
  - Feedback to peer (P/T)
  - Choice of research assignment

Figure 1. Data collection framework.
**Instrumentation.** This section describes the various types of instruments used in this study and their validity.

**Surveys.** The purpose of the pre- and postsurveys was to collect data on the evolving beliefs of high school teachers on the definition, purpose, selection, and implementation of rich mathematical tasks in their classrooms. This survey was selected for the study because it had been created by researcher with the division curriculum study team and was given to all 6-12 division mathematics teachers during the preschool period coinciding with the beginning of the video lesson study analysis. This study was aligned in purpose as a component of the division professional learning goal of implementing rich performance tasks for formative feedback in all 6-12 math courses. The overarching division purpose was to pilot this professional learning model in order to refine it before offering it to teachers in the future. Therefore, using the same measurement tool made sense.

**IQA observation rubrics.** The backbone of the data collection in this study came from the data collected by Melissa Boston’s 2012 Instructional Quality Assessment rubrics. These rubrics were originally created for use in teacher and program evaluation, but Boston and others saw the value in using them for teacher professional development beginning in about 2004 (Crosson, Boston, Levison, Matsumura, Resnick, & Wolf, 2004). Built on the extensive work of Stein and Smith who have worked with preservice teachers on the selection and implementation of rich mathematical tasks for two decades, the IQA instrument was well-grounded in current theory of how children construct understanding of mathematics (Boston & Smith, 2009). Key were opportunities for
students to make their mathematical reasoning visible through classroom discourse during rich tasks (Kazemi & Franke, 2014; NCTM, 2000; Stein & Smith, 2011). Teachers self-assessed during video observations on moves that “connected students’ approaches and the underlying mathematics” as outlined by Stein and Smith (2011) during “productive” classroom discussions.

Research on IQA rubric development. Teachers have long needed tools to help them self-assess their classroom interactions against specific criteria in order to reflect and adjust to improve (Elmore, 2002). There is a large amount of accrued research knowledge about what high quality instruction looks like in mathematics. The IQA rubrics offer a practical tool that can help teachers to self-assess and give peer feedback on task implementation against research-based criteria. Secondly, the tool can be used by principals and other instructional leaders to identify teacher leaders who can lead change efforts in mathematics and to identify next steps for growth in teachers in the school (Crosson et al., 2004). I selected this tool for all of these purposes.

Usually, evaluation tools that have been designed for large-scale summative assessment purposes have not had the ability to give the detailed, context-dependent feedback needed in formative assessment (Crosson et al., 2004; Junker et al. (2006). Even though the IQA rubrics were originally designed to be “external assessments designed to meet stringent requirements of reliability and validity”, the technical qualities of the IQA rubrics along with the specific descriptors were ideal for use in formative feedback (Crosson et al., 2004). The tool also has been tested for inter-rater reliability over the past ten years and has been refined to be hard to misapply (Boston, 2012).
Research suggests that the IQA rubrics are reliable in as few as two observations of the same teacher (Matsumura, Garnier, Slater & Boston, 2008).

**Data analysis.** I collected different types of data from each of six data sources in order to answer the research questions. Numerical data from surveys and tasks were collected for analysis. Descriptive data were used from the same data sources of open-ended survey questions, coaching reflections, and classroom videos. Unsorted data were collected and coded using axial and constant comparative analysis. I found themes and triangulated data and used connecting strategies to describe classroom mathematical task implementation in terms of academic rigor and accountable talk moves (Boston, 2012). Data were fit into an existing framework for course design (Lesh & Hjalmarson, 2007).
Table 1

Data Collection and Data Analysis Procedures

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Description</th>
<th>Data Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Surveys</td>
<td>Preperformance task survey ( (n = 9) )</td>
<td>Multiple coding methods, Frequencies, Descriptive Statistics, Cross-tabulations</td>
</tr>
<tr>
<td></td>
<td>Poststudy performance task survey ( (n = 9) )</td>
<td>Comparison of means, Correlations, Cognitive demand of tasks, Student reasoning and classroom communication, Value of coaching activities to practice</td>
</tr>
<tr>
<td>Teacher Goals</td>
<td>Prestudy Goal Setting ( (n = 9) )</td>
<td>Multiple coding methods, Categories: Coaching choices, Teacher choices, Value of PD activities reasoning &amp; communication</td>
</tr>
<tr>
<td></td>
<td>Poststudy Goal Analysis ( (n = 9) )</td>
<td></td>
</tr>
<tr>
<td>Teacher Coaching Video “Barbara” Reflections</td>
<td>Pre/Post Classroom Observation Reflections ( (n = 18) )</td>
<td>Multiple coding methods, Compared to the Pre/Post beliefs survey</td>
</tr>
<tr>
<td>Math Performance Tasks</td>
<td>Four tasks per teacher ( (n = 36) )</td>
<td>Coded with Academic Potential IQA rubric</td>
</tr>
<tr>
<td></td>
<td>IQA Task Potential Rubric ( (n = 36) )</td>
<td></td>
</tr>
<tr>
<td>Classroom Videos</td>
<td>4 classroom videos per teacher team ( (n = 36) )</td>
<td>Multiple coding methods, IQA Instrument-Focused on cognitive demand of task, student discourse and reasoning</td>
</tr>
<tr>
<td>Instructional Quality Assessment Rubrics</td>
<td>( n = 114 )</td>
<td>Frequencies, Descriptive Statistics, Cross-tabulations, Comparison of means, Selected Correlations</td>
</tr>
</tbody>
</table>
Table 2

Data Sources and Research Questions

<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Description</th>
<th>Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Surveys</td>
<td>Preperformance task surveys by teacher</td>
<td>For what purposes did high school mathematics teachers select tasks for students?</td>
</tr>
<tr>
<td></td>
<td>( n = 9 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Poststudy performance task survey</td>
<td>What did high school mathematics teachers want to change about task implementation in their classrooms?</td>
</tr>
<tr>
<td></td>
<td>( n = 9 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>What changes did teachers report in their own task implementation?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Which coaching activities did teachers perceive as leading to growth?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How did the learning process compare across three teacher groups: new, special education, and experienced?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What suggestions did teachers offer to improve coaching activities in future models?</td>
</tr>
<tr>
<td>Teacher Goals</td>
<td>Prestudy Goal Setting ( (n = 9) )</td>
<td>What did high school mathematics teachers want to change about task implementation in their classrooms?</td>
</tr>
<tr>
<td></td>
<td>Poststudy Goal Reflection ( (n = 9) )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lesson Goals ( (n = 36) )</td>
<td></td>
</tr>
<tr>
<td>Teacher Coaching</td>
<td>Pre/Post Classroom Observation Reflections ( n = 9 ) pre; 9 post</td>
<td>For what purposes did high school mathematics teachers select tasks for students?</td>
</tr>
<tr>
<td>Video Reflections of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Barbara”</td>
<td></td>
<td>What did high school mathematics teachers want to change about task implementation in their classrooms?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(continued)</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Data Sources</th>
<th>Description</th>
<th>Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematical Performance Tasks</td>
<td>Four tasks per teacher $(n = 36)$</td>
<td>How did the potential of the task relate to the cognitive demand experienced by students during implementation?</td>
</tr>
<tr>
<td></td>
<td>IQA Task Potential Rubric $(n = 114)$</td>
<td>How did the task selection compare across three teacher groups: new, special education, and experienced?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom Videos</td>
<td>Pre- and poststudy beliefs video analysis $(n = 18)$</td>
<td>How did the potential of the task relate to the cognitive demand experienced by students during implementation?</td>
</tr>
<tr>
<td></td>
<td>4 classroom observations per teacher $(n = 36)$</td>
<td>In what ways did academic rigor change over the course of four lesson cycles for teachers and students?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How did the learning process compare across three teacher groups: New, Special Education, and Veteran?</td>
</tr>
<tr>
<td>IQA Rubrics</td>
<td>Classroom lesson IQA Observation rubrics $(n = 114)$</td>
<td>How did the potential of the task relate to the cognitive demand experienced by students during implementation?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In what ways did academic rigor change over the course of four lesson cycles for teachers and students?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>How did the learning process compare across three teacher groups: new, special education, and experienced?</td>
</tr>
</tbody>
</table>
Validity. I have designed multiple measures of validity to increase authenticity, so that I can increase trustworthiness of the implications (Denzin & Lincoln, 2008). This is particularly important due to the pragmatic paradigm in which this study operates, however dialectic in approach to method. I needed to analyze the data from the stance of potential bias and design flaws in order to question findings, because I have been using the findings of this study to adjust this professional development model in my own division. I have been transparent about limitations in order to mitigate them with additional study addendums and to inform recommendations for future studies. This is important for other researchers and practitioners who wanted to apply this process to future studies or professional learning applications.

I used five validity checks in this study of the use of a video lesson analysis by teachers in an embedded professional development model. These checks included: integrating multiple data sources as a form of triangulation of data as well as of methods, comparison of methods to accepted models, instruments and definitions from literature, observations and coding of data by three to six observers, actively searching for discrepancies in data and negative cases (Maxwell, 2013), and member checking.

Results have been shared with the AFDA teacher team and the division Curriculum Study Team to plan adjustments in delivery of future iterations of use of lesson videos in professional learning. Member checking from AFDA teachers about the study and results were collected in a panel discussion following the data collection and used as emerging data to enhance the discussion section and add rigor to the analysis.
**Design strengths.** I viewed three strengths of this study design: triangulation of multiple data sources, integration of all data in order to answer each research question, and the authentic setting of this study within the larger context of a longitudinal program design and evaluation study. I aligned my design methods and analyses to these strengths to get their full effect.

**Limitations of the methods.** The major design flaws were my researcher and participant bias and the relatively short duration of a study of one school year. Had I collected task implementation data over two or more school years, I would have better been able to see the trajectory of learning by teachers over time. A longitudinal data set would have allowed me to measure teacher maintenance of change over time due to the treatment of the professional development. Having data collection points during the following school year would have been optimal to check for long-term transfer to practice. However, some of the teachers had changed course schedules the following year, and the timeline for the study did not make this possible.

**Summary**

This chapter discussed the plan and rationale for data collection and analysis around the research questions in this study. A data collection framework was presented to show how the blended learning professional learning model would work as an iterative design over four lesson cycles. An explanation for the use of a mixed-methods design with data sources and analysis protocols were explained. I described how the participants would participate in this embedded professional learning activity over the course of one school year in their school settings. I presented my rationale for selecting the IQA
instrument for teachers and coaches to use to self-assess and give feedback on lesson observations via video. I shared possible validity threats and my plan for mitigating their effects. Lastly, a discussion of interpretation, limitations and suggested next research steps was presented.

Chapters 4 and 5 report the findings and inferences of this study. Chapter 4 presents the quantitative and qualitative data from the survey results and the IQA rubric results by teacher, teacher groups (new, experienced, special-education) and for the study across 10 dimensions. Chapter 5 presents a comparative case study of the video lesson study coaching process as experienced by three representative teachers from the three teacher groups in the study. The final chapter integrates all findings in a summary by research question.
Chapter Four

The results from the five data sources in the study are reported in this chapter along with an interpretive summary by source to situate findings within the literature about the coaching of teachers through video lesson analysis. The purpose of the professional learning was for teachers and students to improve capacity to keep the academic rigor of mathematical tasks high during task implementation through the accountable talk moves occurring during classroom discourse.

Teacher Profiles

The first source collected was a prestudy survey of the nine teacher participants regarding their current beliefs about the selection, purpose, and use of rich mathematical tasks in their classrooms before taking part in the video lesson analysis professional learning. To complement self-reported pretreatment beliefs to check for corroboration and to better describe beliefs about task implementation, I also collected a video lesson analysis from each teacher of a classroom lesson video of an Algebra I teacher named “Barbara” teaching a quadratic functions lesson offered on a professional development site. This was a written peer-coaching reflection of Barbara’s task selection and implementation. I used open-axial coding on the written reflections of the video analysis and then organized the codes into categories and found emerging themes across teachers. I compared these major themes in a matrix with columns for each of the three qualitative
data sources: Preperformance task survey, prevideo lesson analysis, postperformance task survey and postvideo lesson analysis across the 10 dimensions of the IQA observation instrument. I then created individual teacher profiles by integrating data sources by variable in a matrix.

**Sources.** I collected data from each of five data sources: pre/post teacher surveys, pre/post video coaching analysis of a common lesson, mathematical tasks, classroom observation videos, and IQA observation rubrics submitted by teachers, peer teachers and coaches. Numerical data from surveys and IQA rubrics of math tasks and lesson observations were collected for analysis. Ten variables were coded: Task Potential, Task Implementation, Participation, Teacher Press, Teacher Connecting to Concepts and Linking, Student Linking, Student Responses, Student Discussion, Questioning, and Mathematical Residue of big ideas. Descriptive statistics including frequencies and central tendencies were run on SPSS. I looked at the means across the four lesson cycles by variable and as an aggregate. I compared the means for the 10 variables. I ran correlations between variables. Additionally, I ran Crosstabs on comparisons of interest from the research questions. Examples included Task Potential vs. Task Implementation, Task Potential and Task Implementation by teacher category (new, special education, and experienced), Task Rigor vs. Rigor of Questioning, Student Discussion by level of teacher and by Task Implementation, Teacher Linking, and Press by teacher level.

**Descriptive data.** Data were used from open-ended survey questions, teacher video reflections, and classroom observation rubrics. The Video Observation Feedback Survey (Appendix B) asked teachers to identify the target implementation goal, peer
observers to enter coaching suggestions, and teachers to reflect on successes or lesson adjustments. Questions recorded evidence of the 10 variables to describe where teachers were in terms of the domains of Accountable Talk and Academic Rigor during task selection and implementation.

**Qualitative data.** Unsorted data were collected and coded using categorizing strategies of axial and constant comparative analysis to fracture the data (Strauss, 1987). I created the categories as they emerged from the transcripts. My purpose in creating categories was twofold: to facilitate comparison between categories and then to rearrange them under broader themes (Maxwell, 2013). I used Maxwell’s connecting strategies by creating matrices and diagrams to connect and compare data from different sources across research questions (Maxwell, 2013). One strategy I used to look for these contiguous relationships was to create a profile matrix of each participant in order to connect the data from all sources for that one teacher (Maxwell, 2013). I wrote memos of findings, emerging themes and questions that I saw in the data (Maxwell, 2013). Another connecting strategy was to compare evidence from the study for new, special education, and experienced teacher groups to the emerging themes in a matrix. In these ways, I used multiple sources to add layers of description to understand the process of professional learning teachers were engaged in during the lesson study. I triangulated data to confirm findings emerging across different sources in the study. Rather than just confirming, I sought to layer all data from different sources and the 10 IQA variables to answer each research question at a deeper level. In doing so, I focused on the student and teacher
moves that provided evidence of student reasoning made visible through classroom discourse during task implementation.

I grouped the data into three comparative cases to understand how the video lesson analysis study was implemented and perceived by new mathematics teachers, special education coteachers and experienced mathematics teachers. I also studied evidence of how these teachers’ performance task beliefs, task selection choices, and classroom discussions evolved over four lesson cycles from a personal baseline video observation.

**Preliminary analyses.** The ongoing data analysis in my design consisted of viewing all observation videos and coding with the IQA rubrics as an ongoing activity throughout the school year of the study to provide feedback to teachers in my role as coach. Ongoing data analysis took place throughout the 12 months of data collection during the four video lesson study cycles to code the observation rubrics from the videoed teachers, and for peer observers and I to provide coaching feedback. To establish a baseline for each teacher, I read all prestudy teacher reflections on the baseline teacher video reflection of a task along with a presurvey on beliefs about and current practices surrounding use of classroom mathematical performance tasks. I also looked at the target goal each teacher set for their own learning in the lesson study. During the four lesson cycles, I viewed and gave rubric feedback, which included coding and tallying each observation by question types asked by the teacher. After the study, I looked at the postevidence by teacher: the postperformance task survey, common video lesson coaching reflection, and video observation IQA rubrics in SPSS and open-ended
reflection questions over four cycles. I used memos to describe the pre- and postprofiles of each teacher and wrote about connections I saw across research questions from different data sources as well as emerging questions and themes (Maxwell, 2013). I used categorizing strategies to code responses to each of ten rubrics for each video observation submitted by taped teachers as well as peer observers and my own as coaching feedback. I then formed categories to group the different codes to collapse the data into emerging themes. I then used connecting strategies to analyze contextual relationships from five different data sources across the research questions. The qualitative sources served to explain the process of how the professional learning was implemented and which components teachers believed benefited them most. Quantitative and qualitative data together also gave me a description of where each teacher was in terms of growth in task selection and implementation over four cycles. This provided data for next steps for the AFDA program professional development but also for each teacher to inform next instructional improvement steps. Many used this study to reflect deeply about practices, which led to the seeking out of additional opportunities for learning following the study.

**IQA Observation Instrument Results**

This chapter section reports findings on the video analyses done by teachers, peer teachers and a coach to provide feedback and self-assessment by teachers of their task selection and implementation. Nine participating teachers completed four lesson cycles each and submitted, along with 10 peer observers, a total of 114 feedback rubrics during this professional development. This section reports findings from each of the 10 IQA domain rubrics addressing Academic Rigor and Accountable Talk during the lessons.
Academic Rigor domains were: Task Potential, Task Implementation, Student Discussion Following the Task, Questioning, and Mathematical Residue. Accountable Talk domains included: Participation, Teacher’s Linking, Students’ Linking, Teacher Press, and Student Responses. This section includes variables selected, missed data, descriptive statistics, a comparison of means, correlations between the ten rubric variables, and selected cross-tabulations. These data were collected to answer the following research questions:

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
   a. For what purposes did high school mathematics teachers select tasks for students?
   b. What did high school mathematics teachers want to change about task implementation in their classrooms?
   c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?
   d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?
   e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?
   a. How did teachers describe the change process in their own task implementation?
b. Which coaching activities did teachers perceive as leading to growth?

c. How did the learning process compare across three teacher groups: new, special Education, and experienced?

d. What suggestions did teachers offer to improve coaching activities in future models?

**Preliminary analysis.** To begin, I ran frequencies and descriptive analyses on each of the 10 variables by lesson cycle numbers one through four to get an overview of the data and detect any emerging patterns of interest. Only 5 data elements out of 1,140 were missing from the 10 rubric categories for each lesson feedback form. Four occurred when the observer noted that she could not gauge class participation as a percentage due to her vantage point in the video. One observer did not code teacher linking and connecting to concepts in one observation. The following summarizes the distribution of lesson analyses by lesson number.

<table>
<thead>
<tr>
<th>Lesson Analyses Count</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Lesson #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>14</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>18</td>
<td>7</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>7</td>
<td>19</td>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>11</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>4</td>
<td>19</td>
<td>66</td>
<td>25</td>
<td>114</td>
</tr>
</tbody>
</table>
**Variables.** The following variable names were used in the IQA instrument in this study under the domain of Academic Rigor: Task Potential (TP), Task Implementation (TI), Student Discussion (SDiscussion), Teacher Questioning (TQuestioning), Mathematical Residue (MathResidue). For the domain of Accountable Talk, five variable names were used: Participation (P), Teacher’s Connecting and Linking (TConn), Students’ Linking (SLink), Asking Teacher Press (TPress), and Providing Student Responses (SResponses). Each of these stands for a longer variable name described in the corresponding results section. The abbreviations are the ones I used in the data runs in SPSS.

**Frequencies of interest.** While I ran analyses of frequency distributions of all 10 variables, several stood out due to their alignment to the research questions.

**Task potential vs. task implementation.** The cognitive demand of selected tasks and the implementation of those tasks both mattered; students who have been in classrooms where high level tasks have been kept high during enactment, simply performed better (Stein & Lane, 1996).

<table>
<thead>
<tr>
<th>Cognitive Demand Level</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>8.8</td>
<td>8.8</td>
<td>9.6</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
<td>56.1</td>
<td>56.1</td>
<td>65.8</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>34.2</td>
<td>34.2</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 5

Task Implementation

<table>
<thead>
<tr>
<th>Cognitive Demand Level</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>3.5</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>16.7</td>
<td>16.7</td>
<td>20.2</td>
</tr>
<tr>
<td>3</td>
<td>65</td>
<td>57.0</td>
<td>57.0</td>
<td>77.2</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>22.8</td>
<td>22.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 6

Student Participation in Discussion of Task

<table>
<thead>
<tr>
<th>1= &lt;25% 4= &gt;75%</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>10.5</td>
<td>10.9</td>
<td>10.9</td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>21.1</td>
<td>21.8</td>
<td>32.7</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>35.1</td>
<td>36.4</td>
<td>69.1</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>29.8</td>
<td>30.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>96.5</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>

Missing System

<table>
<thead>
<tr>
<th>Missing System</th>
<th>4</th>
<th>3.5</th>
</tr>
</thead>
</table>

Total

| Total | 114 | 100.0 |

Table 7

Rigor of Teacher Questioning

<table>
<thead>
<tr>
<th>Cognitive Demand Level</th>
<th>Frequency</th>
<th>Percent</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>7.0</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>9.6</td>
<td>9.6</td>
<td>16.7</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>44.7</td>
<td>44.7</td>
<td>61.4</td>
</tr>
<tr>
<td>4</td>
<td>44</td>
<td>38.6</td>
<td>38.6</td>
<td>100.0</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>100.0</td>
<td>100.0</td>
<td></td>
</tr>
</tbody>
</table>
Cross-tabulations. I looked at cross-tabulations by key research question. I looked at the relationship between task potential and task implementation by teacher group: new, special education and experienced. Experienced teachers, as a group, were more consistent in keeping the cognitive demand at a higher level during task trajectories.

Table 8

Task Potential vs. Implementation by Teacher Group

<table>
<thead>
<tr>
<th>TchrGrp</th>
<th>TP</th>
<th>TI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced Teachers</td>
<td>3.49</td>
<td>3.28</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.64</td>
</tr>
<tr>
<td>New Teacher or to AFDA</td>
<td>3.03</td>
<td>2.82</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.53</td>
</tr>
<tr>
<td>SPED Teacher</td>
<td>3.17</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.66</td>
</tr>
<tr>
<td>Total</td>
<td>3.24</td>
<td>2.99</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.62</td>
</tr>
</tbody>
</table>

I then looked at TP vs. TI by teacher and added it to the teacher profile matrix by research question.

Comparison of means. First, I looked at the distribution of scores for each of the 10 variables over the four lesson cycles to note patterns. Next, I looked at the means for each of the 10 IQA rubric categories for Academic Rigor and Accountable Talk over the four lesson cycles (Table 9). The domain of Academic Rigor included the following variables: task potential, task implementation, teacher questioning, and mathematical
residue left from the lesson. The domain of Accountable Talk included the dimensions of student participation, teacher connecting and linking to concepts, teacher press for students to explain and justify, student responses, student linking, and student discussion.

In the baseline lessons, Student Linking and rigor of Student Discussion were the lowest-scoring variables. Four of the five lowest dimensions were student measures; the highest entering dimension was Task Potential of selected tasks by teachers. Therefore, Task Potential had the least amount of room for growth. In the final Lesson 4, four of the five dimensions showing the most improvement were in the student realm. Based on difference of means, Teacher Connecting and Linking improved the most followed in order by Student Linking, Student Discussion, Student Participation, Student Responses, Teacher Questioning, Math Residue, Teacher Press, Task Implementation, and Task Potential. All 10 variables showed improvement from the baseline to the final lesson.
Table 9

<table>
<thead>
<tr>
<th>Lesson</th>
<th>TP</th>
<th>TI</th>
<th>P</th>
<th>TConn</th>
<th>TPress</th>
<th>SResponses</th>
<th>SLink</th>
<th>SDiscussion</th>
<th>TQuestioning</th>
<th>MathResidue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>3.05</td>
<td>2.75</td>
<td>2.30</td>
<td>2.35</td>
<td>2.75</td>
<td>2.35</td>
<td>1.70</td>
<td>2.00</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.61</td>
<td>.64</td>
<td>1.08</td>
<td>.97</td>
<td>1.09</td>
<td>.98</td>
<td>.97</td>
<td>1.05</td>
<td>.10</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>3.18</td>
<td>2.94</td>
<td>2.75</td>
<td>2.97</td>
<td>2.79</td>
<td>2.76</td>
<td>2.18</td>
<td>2.58</td>
<td>3.12</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>33</td>
<td>33</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.77</td>
<td>.87</td>
<td>.92</td>
<td>.85</td>
<td>.99</td>
<td>.87</td>
<td>.85</td>
<td>.61</td>
<td>.89</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>3.24</td>
<td>2.91</td>
<td>3.00</td>
<td>3.16</td>
<td>3.06</td>
<td>2.91</td>
<td>2.36</td>
<td>2.58</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>33</td>
<td>33</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.56</td>
<td>.72</td>
<td>.97</td>
<td>1.05</td>
<td>.90</td>
<td>.88</td>
<td>1.06</td>
<td>1.15</td>
<td>.78</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>3.43</td>
<td>3.32</td>
<td>3.30</td>
<td>3.46</td>
<td>3.36</td>
<td>3.32</td>
<td>2.79</td>
<td>3.00</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>28</td>
<td>28</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.57</td>
<td>.61</td>
<td>.78</td>
<td>.69</td>
<td>.68</td>
<td>.67</td>
<td>.96</td>
<td>.82</td>
<td>.58</td>
</tr>
<tr>
<td>Total</td>
<td>M</td>
<td>3.24</td>
<td>2.99</td>
<td>2.87</td>
<td>3.04</td>
<td>3.00</td>
<td>2.87</td>
<td>2.30</td>
<td>2.58</td>
<td>3.15</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td>114</td>
<td>114</td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>.64</td>
<td>.74</td>
<td>.98</td>
<td>.97</td>
<td>.91</td>
<td>.92</td>
<td>1.01</td>
<td>.95</td>
<td>.87</td>
</tr>
</tbody>
</table>

Note. SD rounded to 2 decimals.
To answer Research Question 2.c which compared the learning experience across teacher groups, I looked at the two domains, Academic Rigor and Accountable Talk by group: new, special education and experienced to look for patterns. Experienced teachers performed higher than special education and new teachers across all five dimensions of the Academic Rigor domain and four of the dimensions of Accountable Talk. The exception was the Participation dimension, which was about the same across teacher groups. I then looked at the means for each dimension by teacher and added them to the teacher profile matrix by research question.

Table 10

*Academic Rigor Dimensions by Teacher Group*

<table>
<thead>
<tr>
<th>TchrGroup</th>
<th>TP</th>
<th>TI</th>
<th>SDiscussion</th>
<th>TQuestioning</th>
<th>MathResidue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.49</td>
<td>3.28</td>
<td>2.85</td>
<td>3.51</td>
<td>3.23</td>
</tr>
<tr>
<td>N</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td>SD</td>
<td>.64</td>
<td>.69</td>
<td>.78</td>
<td>.85</td>
<td>.96</td>
</tr>
<tr>
<td>New Teacher or to AFDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.03</td>
<td>2.82</td>
<td>2.55</td>
<td>2.88</td>
<td>2.67</td>
</tr>
<tr>
<td>N</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>SD</td>
<td>.53</td>
<td>.77</td>
<td>.87</td>
<td>.78</td>
<td>.65</td>
</tr>
<tr>
<td>SPED Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.17</td>
<td>2.86</td>
<td>2.36</td>
<td>3.02</td>
<td>2.57</td>
</tr>
<tr>
<td>N</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>SD</td>
<td>.66</td>
<td>.68</td>
<td>1.10</td>
<td>.84</td>
<td>.99</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>3.24</td>
<td>2.99</td>
<td>2.58</td>
<td>3.15</td>
<td>2.82</td>
</tr>
<tr>
<td>N</td>
<td>114</td>
<td>114</td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>SD</td>
<td>.64</td>
<td>.74</td>
<td>.95</td>
<td>.87</td>
<td>.93</td>
</tr>
</tbody>
</table>

*Note. SD rounded to 2 decimals.*
### Table 11

**Accountable Talk Dimensions by Teacher Group**

<table>
<thead>
<tr>
<th>TchrGroup</th>
<th>M</th>
<th>TConn</th>
<th>TPress</th>
<th>SResponses</th>
<th>SLink</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experienced Teacher</td>
<td>2.87</td>
<td>3.28</td>
<td>3.41</td>
<td>3.15</td>
<td>2.62</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>39</td>
<td>39</td>
<td>39</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>.91</td>
<td>1.03</td>
<td>.85</td>
<td>.93</td>
<td>1.09</td>
</tr>
<tr>
<td>New Teacher or to AFDA</td>
<td>2.88</td>
<td>2.88</td>
<td>2.73</td>
<td>2.52</td>
<td>2.03</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td>.96</td>
<td>.98</td>
<td>.87</td>
<td>.92</td>
</tr>
<tr>
<td>SPED Teacher</td>
<td>2.87</td>
<td>2.93</td>
<td>2.83</td>
<td>2.88</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>39</td>
<td>41</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>1.11</td>
<td>.91</td>
<td>.80</td>
<td>.86</td>
<td>.95</td>
</tr>
<tr>
<td>Total</td>
<td>2.87</td>
<td>3.04</td>
<td>3.00</td>
<td>2.87</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td>110</td>
<td>113</td>
<td>114</td>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>.98</td>
<td>.97</td>
<td>.91</td>
<td>.92</td>
<td>1.01</td>
</tr>
</tbody>
</table>

*Note. SD rounded to 2 decimals.*

**Correlations.** In order to examine the paired relationship between different variables coded by observers with the IQA rubrics, bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance. Results shown in Table 12 show the relationships between variables coded from the lesson observation videos. Teacher Connecting and Linking, and Teacher Questioning were the two dimensions that were more strongly linked to the other dimensions. Teacher Connecting and Linking, which teachers focused on as a goal in the majority of the lessons, showed correlations above 0.5 with seven of the other variables, including all of the student variables. Teacher Questioning was correlated above 0.5 with six of the other dimensions, including all of the teacher variables. In contrast, the variables less strongly associated with the others were Task Potential and Participation.
Participation was the most frequent improvement goal selected by teachers yet it was not as strongly associated with increases in quality and rigor of student discussion as the teacher moves of connecting and linking, pressing students to explain, questioning, and student linking. Increasing student reasoning in discussions was the top goal set by teachers once students were actively participating.

As teachers modeled connecting and linking ideas and asked more rigorous questions, students tended to explain and justify more which increased the cognitive demand of student discussions. Correlations above .5 in these four variables with mathematical residue showed that increases in these dimensions were associated with corresponding increases in students understanding of the big ideas.
Table 12

Correlations: Instructional Quality Assessment Dimensions

<table>
<thead>
<tr>
<th></th>
<th>TP</th>
<th>TI</th>
<th>P</th>
<th>TConn</th>
<th>TPress</th>
<th>SResponses</th>
<th>SLink</th>
<th>SDiscussion</th>
<th>TQuestionning</th>
<th>MathResidue</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>r</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>110</td>
<td>113</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TI</strong></td>
<td>.57**</td>
<td>.39**</td>
<td>.53**</td>
<td>.37**</td>
<td>.48**</td>
<td>.46**</td>
<td>.52**</td>
<td>.40**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>110</td>
<td>113</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P</strong></td>
<td>.38**</td>
<td>.39**</td>
<td>.56**</td>
<td>.58**</td>
<td>.47**</td>
<td>.43**</td>
<td>.43**</td>
<td>.49**</td>
<td>.35**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>110</td>
<td>110</td>
<td>109</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TConn</strong></td>
<td>.46**</td>
<td>.52**</td>
<td>.58**</td>
<td>.57**</td>
<td>.67**</td>
<td>.56**</td>
<td>.53**</td>
<td>.71**</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TPress</strong></td>
<td>.478**</td>
<td>.53**</td>
<td>.58**</td>
<td>.57**</td>
<td>.67**</td>
<td>.56**</td>
<td>.53**</td>
<td>.71**</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SResponses</strong></td>
<td>.47**</td>
<td>.37**</td>
<td>.47**</td>
<td>.56**</td>
<td>.67**</td>
<td>.72**</td>
<td>.52**</td>
<td>.57**</td>
<td>.51**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SLink</strong></td>
<td>.38**</td>
<td>.48**</td>
<td>.43**</td>
<td>.48**</td>
<td>.56**</td>
<td>.72**</td>
<td>.56**</td>
<td>.46**</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SDiscussion</strong></td>
<td>.34**</td>
<td>.46**</td>
<td>.43**</td>
<td>.61**</td>
<td>.53**</td>
<td>.52**</td>
<td>.56**</td>
<td>.51**</td>
<td>.58**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TQuestioning</strong></td>
<td>.60**</td>
<td>.52**</td>
<td>.49**</td>
<td>.58**</td>
<td>.71**</td>
<td>.57**</td>
<td>.46**</td>
<td>.51**</td>
<td>.53**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MathResidue</strong></td>
<td>.47**</td>
<td>.40**</td>
<td>.36**</td>
<td>.59**</td>
<td>.44**</td>
<td>.51**</td>
<td>.44**</td>
<td>.58**</td>
<td>.53**</td>
<td></td>
</tr>
<tr>
<td><em>Sig. (2-tailed)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>114</td>
<td>114</td>
<td>110</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: *p < .001. **Correlation is significant at the 0.01 level (2-tailed); rounded to 2 decimals.*
Academic Rigor

The domain of Academic Rigor was comprised of five dimensions: Task Potential, Task Implementation, Rigor of Teacher Questioning, Student Discussion after the Task, and Mathematical Residue, or the degree to which the big ideas stuck with students. In this domain, Rigor of Teacher Questioning was the rigor variable more frequently and strongly associated with higher cognitive demand in all four of the other rigor dimensions as well as in all of the teacher variables.

Task potential. Different tasks offer different levels of cognitive complexity for students to access. Stein et al., (2000) said, “Not all tasks were created equal; different tasks will provoke different levels and kinds of student thinking.” The rubric asked observers and teachers of each lesson to code for the Potential of the Task in terms of cognitive demand from lower to higher levels of thinking: 1 (Memorization Task), 2 (Procedures without Connections), 3 (Procedures with Connections), and 4 (Doing Mathematics). The tasks used in 114 out of 114 video observations were coded for cognitive demand on this scale. The mean score was 3.24 ($SD = .64$). The mean scores for 114 lesson observations of the nine study teachers over the four lesson cycles were as follows: Lesson 1 mean = 3.05 ($SD = .605$), Lesson 2 mean = 3.18 ($SD = .769$), Lesson 3 mean = 3.24 ($SD = .561$), and Lesson 4 mean = 3.43 ($SD = .573$).

Of interest in Research Question 1.b was the relationship between the variables Academic Potential of the Task and Task Implementation. In a bivariate, two-tailed Pearson correlation conducted at the .05 alpha level of significance, Task Potential is significantly related to Task Implementation at $r = .57$; correlation is significant at the .01
level (2-tailed). Task Potential is significantly related to Participation by students in the classroom discussion \((r = .38)\) and to Rigor of Questioning during that discussion \((r = .59)\).

**Task implementation.** The TIMSS video study showed that with the exception of Japan, other countries did not use rich tasks more than United States teachers; however, teachers in the United States more often reduced the cognitive demand of the tasks during implementation (Hiebert et al., 2003). In this study, the IQA rubric asked observers to rate the level at which the teacher guided students to engage with the task in implementation (Boston, 2012). All 114 observations were coded for the level of student engagement in complex thinking or engagement in exploring and understanding the nature of mathematical concepts, procedures, and/or relationships (Boston, 2012).

The mean score for the 114 lesson observations for Task Implementation was 2.99 out of 4.00 \((SD = .74)\). The mean scores for 114 lesson observations of the nine study teachers over the four lesson cycles were as follows: Lesson 1 mean = 2.75 \((SD = .639)\), Lesson 2 mean = 2.94 \((SD = .827)\), Lesson 3 mean = 2.91 \((SD = .723)\), and Lesson 4 mean = 3.32 \((SD = .612)\). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. The mean score for Task Potential was 3.24 compared to 2.99 for Task Implementation. Not surprisingly, Task Implementation correlated significantly with Task Potential \((r = .57)\), Student Discussion \((r = .46)\), Questioning \((r = .52)\), and Math Residue \((r = .40)\).

**Student discussion following the task.** In current practice, teachers are moving to become facilitators of mathematical tasks where students model situations with mathematics to construct their own understanding (NCTM, 1989, 2000; Stein, Engle,
Smith & Hughes, 2007; Stigler & Hiebert, 1999, 2010; Walshaw & Anthony, 2008). In this model, students work through real-world tasks with each other, creating solutions or models that make sense to them. The teacher monitors group discussions collecting formative assessment information, uses evidence of current thinking to question individuals and groups to move learning forward, and selects which student solution strategies to present and sequences different solution strategies presented based on the mathematics goal of the lesson. The student discussions connect and link ideas between representations and different solution methods in order to move student understanding to the big ideas (Stein & Smith, 2011). Key is facilitating the mathematical discussions where teachers press students to explain and justify their mathematical reasoning (Cobb et al., 1993; Cobb, Boufi, McClain, & Whitenack, 1997; NCTM, 1989, 2000; Stein & Smith, 2011). This makes student thinking visible so that both students and teachers can act on that knowledge (Forman et al., 1998; Hattie, 2012). The third dimension under Academic Rigor was the variable Student Discussion Following Task. This dimension measured to what extent students showed their work and explained their thinking about the important mathematical content on a scale of one to four (Boston, 2012). If there was no discussion of the task, then the observer coded a zero or N/A. 114 out of 114 lessons were coded for the Student Discussion Following the Task variable.

The mean score for the 114 lesson observations for Student Discussion was 2.58 out of 4.00 ($SD = .95$). The mean scores over the four lesson cycles were as follows: Lesson 1 mean = 2.00 ($SD = .97$), Lesson 2 mean = 2.58 ($SD = .61$), Lesson 3 mean = 2.58 ($SD = 1.146$), and Lesson 4 mean = 3.00 ($SD = .816$). Bivariate, two-tailed Pearson
correlations were conducted at the .05 alpha level of significance on different variables of interest. Student Discussion Following the Task correlated significantly with Teacher Linking \( (r = .61) \), Mathematical Residue \( (r = .58) \), Student Linking \( (r = .56) \), Teacher Press \( (r = .53) \), Student Responses \( (r = .52) \), Teacher Questioning \( (r = .51) \), Participation \( (r = .43) \) and Task Implementation \( (r = .46) \), and Task Potential \( (r = .34) \).

Student Discussion was the second lowest-scoring dimension in Lesson 1 and also showed the third-highest gains as measured by difference of means in the study. While the percentage of student participation in the discussions increased and students made significant gains in student linking, responses, and discussion, student moves were not as high as the quality of the teachers’ moves. The means for Student Discussion from Lessons 1 to 4 increased from a lower level rating of 2 to a higher-level rating of 3 out of 4. Research suggests that this dimension is difficult for teachers to facilitate, especially connecting the individual student approaches back to the mathematical theory on the fly (Boaler & Humphreys, 2008).

**Questioning.** Key to keeping the cognitive demand of a task high during enactment is the rigor of teacher questioning (Boaler & Staples, 2008). The IQA rubrics asked observers to rate teachers on how often they asked academically relevant questions that provided students with opportunities to elaborate and explain their mathematical work and thinking (Boston, 2012). Observers tallied every question by type while watching the lesson video on a video viewing form, which was created by one of the peer observers. Jo Boaler’s six types of questions were referenced and coded on the form: Probing, Mathematical Meaning/Relationships, Generating Discussion,
Procedural/Factual, Other Mathematical, and Non-Mathematical (Boaler & Humphreys, 2005, as cited in Boston, 2012). Teachers had been trained in the definitions and examples of the types of questions prior to observing. Observers rated lessons on the variable from one to four if there was a class discussion. If no discussion, or the class discussion was not relevant to the mathematics in the lesson, then the observer coded a zero or N/A. 114 out of 114 lessons were coded for this variable. While question types were important, I also coded in qualitative analysis of the pattern of questioning occurring between students and between the students and the teacher. The IQA rubrics coded for a focused questioning pattern where teachers listened and monitored what students were thinking, pressed them to communicate their reasoning, and reflected on both their own and other students’ thinking (NCTM, 2014).

The mean score for the 114 lesson observations for Teacher Questioning was 3.15 out of 4.00 ($SD = .87$). The mean scores over the four lesson cycles were as follows: Lesson 1 mean = 2.60 ($SD = 1.05$), Lesson 2 mean = 3.12 ($SD = .89$), Lesson 3 mean = 3.21 ($SD = .78$), and Lesson 4 mean = 3.50 ($SD = .58$). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. Questioning correlated significantly with Teacher Press ($r = .71$), Task Potential ($r = .59$), Teacher Linking ($r = .58$), Students’ Responses ($r = .57$), Mathematical Residue ($r = .53$), Task Implementation ($r = .52$), Student Discussion ($r = .51$), Participation ($r = .50$), and Student Linking ($r = .46$). Teachers asking more rigorous questions, tended to press students more to explain and justify reasoning with mathematical evidence.
Mathematical residue. The IQA rubrics ranked the extent to which the whole group discussion built new, important mathematical ideas (Boston, 2012). Observers ranked each task for cognitive demand or Task Potential. The dimension of Math Residue codes for the following: existence of a mathematical goal and the alignment of the big ideas in that goal with evidence from the discussion. Additionally, the rubric looks for connections embedded in the task and whether the task serves to solidify or extend student understanding of the mathematical concept goals of the lesson (Boston, 2012).

Observers rated lessons on the variable Mathematical Residue on a scale of one to four if there was a class discussion. If no discussion was held following the task, then the observer coded a zero or N/A. 114 out of 114 lessons were coded for the Mathematical Residue variable.

The mean score for the 114 lesson observations was 2.82 out of 4.00 (SD = .93). The mean scores over the four lesson cycles were as follows: Lesson 1 mean = 2.45 (SD = .10), Lesson 2 mean = 2.73 (SD = .88), Lesson 3 mean = 2.76 (SD = 1.00), and Lesson 4 mean = 3.29 (SD = .71). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. Mathematical Residue correlated significantly with Student Discussion (r = .58), Teacher Questioning (r = .53), Student Responses (r = .51), Task Potential (r = .47), Teacher Linking (r = .44), Student Linking (r = .44), Teacher Press (r = .44), and Task Implementation (r = .40), and Participation (r = .35). Math Residue improved .84 to a higher-level mean rating of 3.29 on the final lesson cycle.
Accountable Talk

Teacher Connecting and Linking was the only Accountable Talk dimension with correlations above 0.5 with all four of the other intentional talk variables as well as with three of the Academic Rigor variables. It is the teacher dimension that is most associated with higher cognitive demand of all of the student variables.

**Participation.** Observers coded each lesson on participation in classroom discussion as a range of percentages from “Less than 25%” to “Greater than 75%.” In 4 out of 114 observations, observers did not code Participation, citing the same reason; that they were unsure, based on the video, of the actual participation as a percentage. As a result, 110/114 observations were coded for Participation in the classroom discussion about the mathematical task. The teacher and the coach or coteacher coding those same lessons would have been able to code Participation, as they had both observed in person and viewed the video later.

The mean for the 110 classroom observations for Participation was 2.87 out of 4.00 ($SD = .98$). The mean scores for 110 lesson observations of the nine study teachers over the four lesson cycles were as follows: Lesson 1 mean = 2.30 ($SD = 1.08$), Lesson 2 mean = 2.75 ($SD = .91$), Lesson 3 mean = 3.00 ($SD = .97$), and Lesson 4 mean = 3.30 ($SD = .78$). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. Not surprisingly, Participation correlated significantly with Student Discussion ($r = .43$), Task Potential ($r = .38$), and Task Implementation ($r = .38$). Participation, or engagement in the classroom...
discussions, had a difference in pre/post means of 1.0 on the 4-point scale, which was third highest dimension, and the most frequently targeted goal set by teachers.

**Teacher connecting and linking.** The IQA rubrics describe the quality of both teacher and student linking moves to connect ideas and positions to each other. The teacher rubric asked if teachers supported students in linking by connecting students’ contributions to each other or by providing opportunities for students to connect (Boston, 2012). Observers rated lessons on the variable Teacher Linking from one to four if there was a class discussion. If no discussion, or the class discussion was not related to mathematics, then the observer coded a zero which was treated as a N/A; 113 out of 114 lessons were coded for this variable.

The mean score for the 113 lesson observations for Teacher Linking was 3.04 out of 4.00 \( (SD = .97) \). The mean scores for 113 lesson observations of the nine study teachers over the four lesson cycles were as follows: Lesson 1 mean = 2.35 \( (SD = 1.0) \), Lesson 2 mean = 2.97 \( (SD = .85) \), Lesson 3 mean = 3.16 \( (SD = 1.05) \), and Lesson 4 mean = 3.46 \( (SD = .69) \). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. The mean score for Teacher’s Linking was 3.04 compared to 2.30 for Student Linking. Teacher Linking correlated significantly with Student Linking \( (r = .48) \). Teacher connecting concepts and modeling linking teacher-to-student and student-to-student showed the most improvement of all 10 dimensions during the lesson cycles. The mean difference between pre- and postlessons was 1.11 on the 4-point scale. This variable was the more frequently and strongly correlated to the other dimensions. Teacher Connecting and Linking was
correlated above 0.5 with all student dimensions: Student Linking, Student Responses in
terms of explaining and justifying, Student Discussion rigor, Participation and the teacher
talk move of Teacher Press as well as Rigor of Teacher Questioning and Mathematical
Residue. Teacher Linking is correlated but not as strongly with Task Potential and Task
Implementation.

**Student linking.** Observers rated lessons on the variable Student Linking from
one to four if there was a class discussion. If no discussion was held, or the class
discussion was not related to mathematics, then the observer coded a zero/NA, which was
treated as a N/A. The rubric asked if students’ contributions linked to and built upon each
other. The rubric scale ranged from no evidence of student linking to consistently linking.
Intermediate benchmarks were one strong linking effort and at least two. The example
statement for the look-for was, “I agree with Jay, because…” (Boston, 2012).

The mean score for the 114 lesson observations for Student Linking was 2.30 out
of 4.00 (SD = 1.01). The mean scores for 114 lesson observations of the nine study
teachers over the four lesson cycles were as follows: Lesson 1 mean = 1.70 (SD = .98),
Lesson 2 mean = 2.18 (SD = .85), Lesson 3 mean = 2.36 (SD = 1.06), and Lesson 4 mean
= 2.79 (SD = .96). Bivariate, two-tailed Pearson correlations were conducted at the .05
alpha level of significance on different variables of interest. The mean score for Student
Linking was 2.30 compared to 3.04 for Teacher Connecting and Linking. Teacher
Linking correlated significantly with Student Linking ($r = .48$). The Teacher Connecting
and Linking variable was the second highest dimension to improve; the difference of
means was 1.09 but at 2.79 was still ranked slightly under the high-level benchmark of 3. Teacher and Student Linking were the top two dimensions showing improvement.

![Means for Student and Teacher Linking over Four Lesson Cycles](image)

**Figure 2.** Means for student and teacher linking over four lesson cycles.

Teachers and students followed a parallel improvement pattern over the four lesson cycles. Both grew at a similar average rate, but students began from a lower initial position (Figure 2). Teacher and student linking were the top two improving dimensions in the study.
**Asking (teacher press).** Another teacher move coded by observers was Teacher Press, or pressing students to support their contributions with evidence and or reasoning. Evidence of “missed press” was also collected, occurring when teachers needed to press but did not (Boston, 2012). Observers rated lessons on the variable Teacher Press from one to four if there was a class discussion. If no discussion was held, or the class discussion was not related to mathematics, then the observer coded a zero or N/A. 114 out of 114 lessons were coded for the Teacher Press variable.

The mean score for the 114 lesson observations for Teacher Press was 3.00 out of 4.00 ($SD = .91$). The mean scores over the four lesson cycles were as follows: Lesson 1 mean = 2.75 ($SD = .97$), Lesson 2 mean = 2.79 ($SD = .97$), Lesson 3 mean = 3.06 ($SD = .90$), and Lesson 4 mean = 3.36 ($SD = .68$). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. Teacher Press correlated significantly with Teacher Questioning ($r = .71$), Student Responses ($r = .67$), Participation ($r = .58$), Teacher Linking ($r = .57$), Student Linking ($r = .56$), Student Discussion ($r = .53$), Task Implementation ($r = .53$), Task Potential ($r = .47$), and the Mathematical Residue left behind by the lesson ($r = .44$). The means for the intentional talk move of pressing students to reason by explanation and justification with mathematical evidence improved .61, bringing the rating into the higher-level rubric range with a 3.36 on the final lesson.

**Providing (student responses).** Along with rating whether teachers pressed students to explain their reasoning, the IQA rubrics assessed students’ responses. The scale measured the level at which students actually supported their contributions with
evidence or reasoning. The overarching question in the observation rubric for the dimension of Student Responses was whether students supported their contributions to the classroom discussion with evidence and/or reasoning (Boston, 2012). The same 4-point scale as in the other Accountable Talk rubrics was used to code student responses. The target behavior was for students to consistently provide evidence for their claims or explain their thinking using conceptual explanations. If no class discussion was held or discussion was not related to mathematics, then a score of 0 was given. At the first level, students did not back up claims or explain the reasoning. At the second level, students provided explanations that were procedural, computational, or were at the memorization level of cognitive demand. Reasoning that was inaccurate or vague was also scored at a 2 level. If students provided evidence once or twice during the lesson, or provided conceptual explanations, then the discussion was coded at a 3 level. Students must have consistently provided evidence for their reasoning or conceptual explanations to have earned the top score of a 4 (Boston, 2012). All 114 lessons were coded for the Student Responses variable.

The mean score for the 114 lesson observations for Students’ Responses was 2.87 out of 4.00 ($SD = .92$). The mean scores over the four lesson cycles were as follows: Lesson 1 mean = 2.35 ($SD = 1.09$), Lesson 2 mean = 2.76 ($SD = .87$), Lesson 3 mean = 2.91 ($SD = .88$), and Lesson 4 mean = 2.87 ($SD = .92$). Bivariate, two-tailed Pearson correlations were conducted at the .05 alpha level of significance on different variables of interest. Students’ Responses correlated significantly with Student Linking ($r = .72$), Teacher Press ($r = .67$), Teacher Questioning ($r = .57$), Teacher Linking ($r = .56$), Student
Discussion ($r = .52$), Mathematical Residue ($r = .51$), Task Potential ($r = .47$), Participation ($r = .47$) and Task Implementation ($r = .37$).

**Research question summary.** Results from the 114 IQA observation instruments showed that all 10 dimensions improved over the four lesson cycles. All 10 variables correlated significantly with each other. The strongest correlations occurred between Student Responses and Student Linking, Teacher Press and Student Responses, Student Linking and Student Responses, Teacher Connecting and Linking to quality of Student Discussion, and rigor of Teacher Questioning and Teacher Press for students to explain their reasoning with conceptual explanations. All of these relationships correlated significantly above .6.

The Accountable Talk variable most highly correlated with all of the student variables in either domain was Teacher Connecting and Linking. Teacher Connecting and Linking was also correlated above 0.5 with teacher variables of rigor of Teacher Questioning, Teacher Press and the leaving behind the Mathematical Residue of the big ideas in the lesson. The cognitive demand of Teacher Questioning was the Academic Rigor variable most highly correlated with the other rigor dimensions. As Teacher Connecting and Linking was associated most strongly with the student variables, Teacher Questioning was linked most strongly with the teacher-controlled variables in both domains. Teacher Questioning was strongly linked to both Task Potential and Task Implementation as well as Teacher Press, Teacher Connecting and Linking, Mathematical Residue as well as Student Discussion.
The next section looks at the pre- and postteacher surveys about beliefs and use of performance tasks.

Performance Task Surveys

The pre- and postsurveys of teacher beliefs about the cognitive demand and use of tasks in classrooms were collected in order to answer Research Question 1 about teacher beliefs about performance tasks and current use, before and after the video lesson study. Survey questions 1 through 11 were written to collect evidence to answer the first of two research questions.

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
   a. For what purposes did high school mathematics teachers select tasks for students?
   b. What did high school mathematics teachers want to change about task implementation in their classrooms?
   c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?
   d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?
   e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

The presurvey contained 11 questions and was administered through Google Forms to nine teachers in the two-day summer professional development days prior to the
school year of the study. The postsurvey contained the same initial questions with three additional questions added to obtain teacher feedback on the efficacy of the professional learning model. The postsurvey was completed via Google Forms by nine teachers on the face-to-face final meeting in June at the end of the school year. Six questions on the presurvey and seven on the posttest were either scale or multiple-choice questions where frequency data could be reported. The remaining questions were constructed response and analyzed using qualitative methods.

Open-ended responses on both surveys were analyzed using two different first cycle coding methods. I used second cycle coding methods to synthesize the data by fitting together categories and finding connections between them (Saldaña, 2013). This chapter reports out both the frequency data from the numerical data and the themes that emerged from the qualitative analysis as evidence categorized by the 10 observation dimension variables from the IQA classroom observation rubrics. The 10 dimensions were organized into two domains: Academic Rigor and Accountable Talk. This organizational structure allowed me to place data from all sources in tables by variable to prepare for the interpretive analysis.

**Survey Results**

**Academic rigor domain.** School divisions have been focusing on aligning their curriculum documents, including daily lesson plans, to the cognitive complexity required of the new standards. This alignment to academic rigor has included the cognitive demand of the selected task, along with its launch, exploration phase and class summary and discussion of big ideas. The variables in this study that form the evidence for data
collection on this academic rigor were task potential, task implantation, teacher questioning, student discussion and the mathematical residue from the lesson.

The pre- and postsurveys on use of performance tasks were given to all math teachers in the division for program evaluation and to use to plan professional learning. This video lesson analysis was one such professional learning project. The following results gave the planning team information as to where the teachers were, challenges to consider, and growth at the end of the video lesson analysis study.

**Task potential.** This section reports task selection criteria and beliefs surrounding the purpose of tasks used to guide the decisions teachers make about which tasks to select. The observation rubric posed the question, “Did the task have potential to engage students in rigorous thinking about challenging content?” (Boston, 2012). The task selection decision was crucial from the standpoint of equity in that this decision determined the mathematics that students will learn (Boston & Smith, 2009; Smith et al., 2012, Stein et al., 1996). A low-level thinking task would never become high; students receive different opportunities for mathematical thinking based on their teacher’s beliefs regarding task selection.

**Definition of a math task.** Both surveys opened with the first question asking teachers to define a mathematical task. In the presurvey, teachers focused on three characteristics: Tasks have multiple uses (7/9), tasks are problems that require math processes (6/9), and tasks require explanation of thinking (5/9). Only two out of nine teachers said that tasks required deeper thinking before the study. 100% of teachers revised their definition in the postsurvey to include statements about cognitive
complexity, and that mathematical tasks require higher-order thinking and/or application and problem-solving (9/9) while 4/9 elaborated by adding justification of thinking with reasoning and proof. I wanted to know how teachers defined tasks, because I felt these beliefs would link to their purpose for using them in the classroom.

**Rich tasks.** Question two asked teachers to describe what made a task “rich.” This question was tied to the IQA rubric for evaluating the Academic Potential of tasks in terms of richness. The IQA rubric is a descendant of Stein and Smith’s Task Analysis Guide (2007), which is the instrument these study teachers have used in their division for the past three years for task selection decisions.

Teachers’ responses to the attributes that make a task rich were distributed into the categories of deeper thinking, application, forming connections, communication, and engagement. The most-mentioned characteristic in both surveys was connections (7/9); teachers described such tasks as “layered,” as having “inter-related concepts,” or connected to prior learning. One-third (3/9) of the teachers in the presurvey said rich tasks required deeper learning while two-thirds (6/9) mentioned deeper cognitive demand. Two teachers in both surveys said rich tasks required real-world application. Tasks that required students to explain their thinking with in-depth communication were deemed “rich” by two prestudy teachers vs. four in the postsurvey. One teacher in the postsurvey explained that a rich task offered multiple paths.

**Classroom use of tasks.** Teachers were asked to select from a list the ways in which they had used tasks. There was an “other” category where additional uses could be added. In the presurvey, the top-scoring use for tasks was to provide differentiation,
including remediation and extension (eight out of nine teachers). Seven out of nine teachers used tasks to introduce a new concept or to activate prior knowledge, apply a concept previously taught, and for formative assessment. Five out of nine teachers used tasks instructionally as the lesson through which students learned the mathematics. Five out of nine saw tasks as useful for culminating unit activities. Only a few teachers used tasks for homework (2/9), extra credit (3/9), or summative assessment (4/9). One special education teacher used tasks as evidence in a state-mandated portfolio collection.

Postsurvey results for how teachers actually used tasks indicated that every teacher used tasks to apply concepts (9/9). Seven out of nine used tasks for formative assessment of student thinking. Six out of nine used tasks to introduce a unit with a problem situation or as a culminating activity. Differentiation dropped from eight to five out of nine teachers. Teaching new mathematical content through a task remained the same (5/9). Infrequent task uses cited included summative assessment (3/9), homework (2/9), and extra credit (0/9). The special education teacher did not list tasks as being used for state assessment data collection as she had in the presurvey.

**Task Implementation.** This section collected evidence on the cognitive level of students during task implementation (Boston, 2012).

*Teacher goals for change in use of tasks.* Teachers were asked in the presurvey Question 10, to name the number one thing they wanted to change in their use of mathematical tasks (Presurvey, Question 10). Following the professional learning, the question was followed up by asking teachers how they had changed the way they used performance tasks (Postsurvey, Question 10). Survey responses on teacher goals were
distributed across seven categories. There was not a majority category in the presurvey. Each category had one or two responses except one. The most-requested change in the presurvey was to allow students to construct their own understanding by becoming facilitators and in doing so, requiring less assistance from the teacher (3/9). Only one teacher mentioned this category as a change in the postsurvey. Following the self-study of their own task implementation, most teachers named improving classroom discourse and questioning as the greatest improvement in task implementation (7/9). Only one teacher out of nine had mentioned classroom discourse as a goal before the study. Four out of nine teachers felt that students’ critical and creative thinking had improved and their understanding formed deeper connections; two of them had identified deeper thinking as a goal. Four teachers said they had increased the frequency of use of tasks so that students felt more comfortable and had increased risk-taking behaviors.

In order to compare self-reported survey data to the goals teachers identified in real-time, I triangulated the teacher survey responses on goal setting to the goals they wrote about in the lesson feedback surveys that teachers completed to self-assess following each lesson. Out of the 36 lessons taught, teachers selected student participation in discussion as a goal of every lesson. The second most selected goal was student discussion after the task (25/36) while teacher linking by connecting student ideas and relationships between was the third (24/36). The least named goal for improvement was Task Potential. It was interesting that teachers recognized that they had already mastered that goal as it was the strongest teacher variable prestudy. The rest of the goals were as follows in Table 13.
Table 13

Lesson Goal-Setting

<table>
<thead>
<tr>
<th>Teacher Improvement Goals</th>
<th>n = 36 lessons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Participation in Discussion</td>
<td>36</td>
</tr>
<tr>
<td>Student Discussion after Task</td>
<td>25</td>
</tr>
<tr>
<td>Teacher Linking</td>
<td>24</td>
</tr>
<tr>
<td>Teacher Pressing students to provide evidence to support their thinking</td>
<td>21</td>
</tr>
<tr>
<td>with conceptual explanations</td>
<td></td>
</tr>
<tr>
<td>Student Linking</td>
<td>18</td>
</tr>
<tr>
<td>Rigor of Teacher Questioning</td>
<td>12</td>
</tr>
<tr>
<td>Mathematical Residue (Big Ideas)</td>
<td>10</td>
</tr>
<tr>
<td>Task Implementation</td>
<td>8</td>
</tr>
<tr>
<td>Task Potential</td>
<td>6</td>
</tr>
</tbody>
</table>

Frequency of task use. Teachers were asked about how often they used rich performance tasks in a typical unit. In both the pre- and postsurveys, one out of nine teachers reported frequency as once or twice per unit. Five out of nine teachers reported once per week while three out of nine said they used tasks more than half of the class periods. These findings were not surprising, since all nine participant teachers have taught in the AFDA experience, an inquiry-based problem-based learning course, where students construct understanding through a series of tasks, projects, and problems as already determined by the division curriculum map.

Student discussion following the task. Both surveys asked teachers to explain their purposes for using whole-class discussions in math classes. As this was a constructed response question, I used qualitative methods to analyze. I recorded all individual ideas from every response, tallied those that were the same, formed categories,
and then collapsed categories. Teachers listed 13 different purposes for classroom discussions in both surveys. However, the purposes changed. In the presurvey, the top reason for having a student discussion was to check for understanding (4/9), introduce a lesson or activate prior knowledge (4/9), and to show different mindsets (4/9). Student collaboration (2/9) and Communication as related to individual student reflection or explanation of thought processes were named by two out of nine teachers. The rest of the reasons each were mentioned once. The postsurvey named collaboration, or sharing of different student ideas, as first (9/9), formative assessment (6/9), summarization of big ideas (4/9), and forming connections to real-world and between ideas (2/9) as reasons for classroom discourse.

Survey Question 11 asked teachers to name the one thing they would like to change about their classroom math discussions. The same question in the postsurvey asked what had actually changed after the video lesson analysis project. Teachers unanimously wanted to increase student participation in classroom discussions (9/9). The postsurvey responses focused on teacher and student questioning as the behavior that had actually improved in classroom discussions. Seven out of nine teachers said they had changed their questioning to ask more probing questions and to press students to explain and justify their reasoning. While most teachers focused on the changes in their questioning, two also cited increases in student-led questioning and student linking and connecting to other student discussion. This aligned with the IQA rubric results that showed improvement in both teacher questioning means (by 0.9) and student response means (by 1.0) in discussion. Three teachers listed that student participation had
increased; three said that classroom discussions were less rushed due to their increased use of wait time. Two specifically said that they now made a conscious effort to stay out of the discussion more and serve as a facilitator.

**Questioning.** Teachers shared their top two question types out of the six coded question types (Boaler, 1999). In the presurvey, Generating Discussion questions were reported used most (4/15), followed by Probing (3/15), Procedural-Factual (3/15), Mathematical Meaning (2/15), and Non-Mathematical (2/15). Poststudy survey results identified the same order for the most-used question types, but the proportions were quite different. Probing and Generating Discussion-type questions were reported in 14 out of 18 responses. Procedural-factual was named in the top two by two teachers while one teacher included Real-World Connections and Mathematical Meaning-type questions.

**Mathematical residue.** Teachers did not mention the dimension of Mathematical Residue in the surveys.

**Accountable talk domain.** Sherin said that teachers listen intently to student mathematical reasoning during classroom discourse to seize important moments that connect to big mathematical ideas and act on them (Sherin, 2001, as cited in Davies & Walker, 2005). The overarching question surrounding the classroom discourse of both student and teachers was, “How effectively did the lesson-talk build Accountability to Knowledge and Rigorous Thinking?” (Boston, 2012). The Accountable Talk domain had five dimensions: Participation, Teacher’s Linking, Students’ Linking, Teacher Press, and Student Responses.
**Participation.** Both pre- and postsurveys asked teachers to rate the participation in their teacher-facilitated discussions on a scale. Before participation in the video lesson study, two out of nine teachers rated participation at 25% to 50%. No teachers on the postsurvey rated their participation below 50%. On the presurvey 4/9 placed their participation at 50% to 75% while 5 out of 9 did on the post. One-third of teachers entered the study rating student participation at over 75% while 4 out of 9 did on the post. Using the same scale, these perceived data points were compared to actual observation ratings from both teachers as well as peer observers from video analysis.

Teachers estimated student participation at a higher rate than the actual participation coded by observers in the 114 lesson observations. Study teachers self-reported their classroom student participation during discussions at a mean of 3.11 out of 4 possible on the prestudy survey. The same teachers reported a mean of 3.44 on the scale of 0-4 for student participation in their classes following the study. The actual mean for student participation in the study was lower, at 2.87. Teachers’ assumptions were correct in that participation increased during the four lesson cycles, but they overestimated that participation percentage range.

**Teachers’ linking.** Teachers did not mention linking by connecting student responses specifically in the pre- and postsurveys. However, linking was mentioned in 24 out of 38 lessons as a goal in the teacher lesson reflections. Boaler and Humphreys (2008) cited the teacher skill of connecting the mathematics in different student strategies as being difficult and hard to anticipate. Yet in quantitative data from the IQA instrument, this variable was more strongly correlated to the other student dimensions than any other
talk move. High modeling of use of connecting and linking strategies of ideas and positions was strongly correlated with increased student participation and linking, student responses to explain and justify, higher cognitive demand in student discussion and increased mathematical residue of the big ideas of the lesson. Strong teacher connecting and linking moves were also associated with teachers who also had higher rigor of questioning.

**Students’ linking.** Five out of nine teachers noted that student involvement in classroom discussions had increased and that students felt more comfortable talking. One described discussion as being led by students, while another noted improvement in student linking and connecting ideas student-to-student. Thus, while most teachers said student participation had improved, only three specifically mentioned student-to-student questioning, linking and connecting.

**Teacher press.** In the postsurvey, seven out of nine teachers said they had improved in their questioning by pressing students to explain and justify their reasoning. Teachers said they asked more “Why?” and “Why do you think…?” and “Why not…” questions. One teacher explained how he had changed the environment of the classroom to be more risk-taking, so that students felt comfortable explaining their answers both agreeing and disagreeing with each other using evidence from their thinking. A teacher explained the shift in the environment as, “creating an expectation that all students are expected to participate in the class also helped me generate more in-depth class discussions.”
**Student responses.** The IQA observation rubrics coded how often students supported their contributions with evidence and/or reasoning (Boston, 2012). The pre- and postsurveys asked teachers to self-report, using the same scale. There was little difference in the pre- and postsurveys. One teacher moved up one level. In both surveys, two out of nine teachers reported that their students consistently provided evidence for their reasoning. On the presurvey, four teachers out of nine, as compared to five, felt that students provided supportive evidence, such as conceptual explanations, once or twice during each lesson. One fewer teacher after the video lesson analysis (2/9) than after (3/9) reported that their students gave computational, procedural explanations or offered incomplete, vague or inaccurate reasoning (Boston, 2012). This finding aligns with the other student dimensions that were very low at the beginning of the study. While they showed the most improvement, they still did not reach the level of the teacher moves by the end of the study.

**“Barbara” coaching reflections.** Examining video footage of classroom lessons for a specific targeted goal has become popular in the past decade as an effective way to deliver embedded professional development to practicing teachers (Borko et al., 2014). In a 2002 study, Jacobs and Morita used video reflections by teachers as a data source to corroborate self-reported answers on a survey by the same teachers. I used a similar approach in this study by having the teacher participants view a video of “Barbara” teaching an Algebra I lesson on quadratic functions. I asked teachers to write a written reflection of what went well and to offer coaching suggestions for improvement. Teachers watched the video and completed the reflection twice: once before engaging in
the lesson study and at the final face-to-face meeting following the professional development.

I recorded each phrase of each response by teacher in a spreadsheet, one row per phrase, creating categories as I went. The emerging categories were: Relationships, Math Learning Community, Lesson Plan, Classroom Discourse, Coaching Suggestions for Improvement, Personal Connections to Practice. I did the same for the “Barbara” postreflections. I entered individual teacher data from the spreadsheet into a matrix along with the pre- and postsurvey data and IQA data to create a rich, multi-layered profile for each of the nine teachers. I then used these profiles to write a comparative case study of three of the teachers. These three teachers were each representative of one the three categories of teachers in the study: New teachers to AFDA, experienced teachers, and special-education teachers coteaching in AFDA. This comparative case study is included in Chapter 5 as part of the interpretation of the results.

**Summary: Data for Research Question 1.** The first research question and its subquestions involved task selection and implementation.

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
   a. For what purposes did high school mathematics teachers select tasks for students?
   b. What did high school mathematics teachers want to change about task implementation in their classrooms?
c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?

d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?

e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

**Task potential vs. task implementation.** The relationship between these two dimensions was examined to answer Research Question 1.c.

**Task selection.** Correlations showed that task potential is positively correlated to task implementation \((r = .57)\). There was a small amount of change from the mean potential rating for tasks in Lesson 1 vs. Lesson 4 (3.05 to 3.24 out of a 4-point scale). A relatively high baseline for task potential was not surprising given the extensive embedded PD that teachers have had using the Task Analysis Guide (Smith & Stein, 1998) for the past three years to evaluate selected tasks. Secondly, all nine of these teachers had collaborated the summer before the lesson study to write and modify tasks for increased cognitive demand as a two-day professional development activity as part of division professional learning. Finally, this group of AFDA teachers had been using tasks as the basis of their course for the past five years; these tasks existed in their curriculum maps and division math website.

Data showed that teachers were typically selecting tasks at the higher levels of thinking: procedures with connections or “doing” mathematics. Data showed that teachers were selecting higher, level tasks about 90% of the time, most commonly
“Procedures with Connections” (Smith, Bill & Hughes, 2008; Smith & Stein, 1998) for about 56% of the tasks. Over a third of the tasks were coded as “Doing” mathematics.

**Task implementation.** While all of the other nine dimensions in the IQA were positively correlated with Task Potential, Teacher Questioning ($r = .59$) and Task Implementation ($r = .57$) were most highly correlated. I had expected task implementation to be the fertile ground for change, since there was more room for growth, and teachers could incorporate tangible teacher moves from the rubric. Task Implementation moved from a mean of 2.75 in Lesson 1 to 3.32 in Lesson 4. These data represented nine teachers and 114 observations of the 36 lessons. This improvement was significant at the .05 alpha level. As with Task Potential, Task Implementation was correlated with all of the other nine dimensions. Task Implementation was most highly correlated with the dimensions of Task Potential ($r = .57$), Teacher Questioning ($r = .52$), and Teacher Press ($r = .53$). In summary, the task selected had to be rich and teacher questioning needed to include generating discussion-type questions to get students talking to make their thinking visible. Teachers then needed to press with probing questions for students to explain and justify their ideas and include mathematical meaning-type questions to get at the big ideas during the discussion following the task. Thus, increasing Task Potential, Rigor of Teacher Questioning and Teacher Press yielded increases in Task Implementation.

**Research Question 2.** The second research question and its subquestions involved a professional learning model.
2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?
   a. How did teachers describe the change process in their own task implementation?
   b. Which coaching activities did teachers perceive as leading to growth?
   c. How did the learning process compare across three teacher groups: New, Special Education, and Experienced?
   d. What suggestions did teachers offer to improve coaching activities in future models?

**Video lesson analysis as a professional learning model.** The teacher postsurvey posed three additional questions from the prestudy survey to solicit feedback on the model as professional development. Question 12 asked teachers, “Which PD activities or tools did you feel were useful to improve something in your teaching?” Teachers chose from a checklist of those offered in the professional learning. Question 13 asked the question as open-ended, “Which part of the professional was most useful?” The final question asked participants to suggest improvements to the video lesson analysis model. These questions were included to collect data to answer the second study research question.

**Feedback.** Teachers selected the self-analysis of their own teaching videos as the most helpful activity to improve teaching. Eight out of nine chose video self-analysis from a checklist and six out of nine also mentioned self-assessment in the open-ended question. Four noted the importance of comparison criteria with the IQA rubrics. Three
teachers noted the cognitive dissonance created by their own perception versus the video reality of their teaching. For example, “It was enlightening to see what was going on in my classroom in reality, not what I thought was going on.” Teachers wrote about the progress they had made as being helpful. As one teacher stated, “I also liked being able to watch my own videos and self-assess how I got better from the first lesson to the final lesson.” Self-analysis was ranked first in effect size in John Hattie’s meta-analyses, (Hattie, 2012).

Getting direct feedback from peer observers was rated second in usefulness in both questions 12 and 13 in both formats. Six out of nine selected peer feedback as being helpful and five out of nine also said that talking face-to-face with the observer was important. Five participants reported that observer feedback was helpful in leading to change in practice. Two teachers talked about getting over the fear of being recorded and allowing themselves to receive feedback.

**Resources.** I had set up digital folders with about 200 resources grouped by topics such as classroom discourse and cognitive demand. Resources included video clips, journal articles, case studies, websites, and books. One-third of the teachers mentioned digital resources as being useful with another two preferring the print versions I had shared with them. Thus, 4/9 participants felt the resources helped while 8/9 felt their own self-analysis of the videos helped the most. This aligned with the work of Dylan Wiliam on formative assessment that highlighted the importance of student self-analysis in the feedback cycle (Wiliam, 2011). Teachers as learners benefited from timely feedback from observers, just as students would. Teachers also valued opportunities to self-assess
based on the criteria from the IQA rubrics. “Reviewing my own teaching by video” was thus the most useful component of the professional learning model noted by participants as leading to change. Research by Hattie (2012) and Wiliam (2011) both highlight the importance of self-assessment by learners as being the most critical component in the formative feedback cycle.

**Suggested improvements in the model.** The final postsurvey question asked participants to offer suggestions for improvements to other teachers studying their classroom tasks and discussions. As a researcher-participant, I needed this feedback as part of professional learning program evaluation for the division mathematics curriculum study team to use to adjust this model before offering it again. Teachers offered 14 different suggestions which could be categorized under two overarching categories: structural/logistical and affective/coaching comments.

**Structural and logistical modifications.** One participant suggested posting a model “comment” lesson to use to compare and contrast against the rubric criteria. Another found useful the perspective of feedback from a nonmath teacher and thought this should be a required aspect when selecting peer observers. This aligned with the idea of identifying noncontent-specific instructional strategies during coaching particularly for increasing student engagement and the cognitive level of discussions through improved questioning techniques. The most-cited suggestion was to add more face-to-face conversations with peer observers about what worked well and what else can be done to help students (3/9). Secondly, one-third of the teachers suggested modification of the lesson based on feedback as a follow-up activity.
Coaching/Affective suggestions. Three teachers suggested making small incremental changes in practice: “Look for small things to change first and make those needed modifications.” One referenced “review a video tape of yourself and reflect.” In this video lesson analysis model, each teacher sets a goal from their baseline video to work on within the IQA domains. Three participants coached peers to “not be afraid,” even if it meant making changes. Two also motivated teachers “not to give up” because, “this video lesson study will help.” Another advised that it was important to be honest with yourself and others in self-analysis and when giving peer feedback. Other wisdom: Choose peers who are as vested in improving their teaching as you, let go of the reigns and continue to work on classroom discourse, ask more questions for students to explain their reasoning, and spend more time on class discussions to “help students increase understanding.”

Summary

Findings in this chapter suggested that the cognitive demand of task implementation was related to the cognitive level of the selected task. While all 10 of the Academic Rigor and Accountable Talk variables correlated significantly with each other, the variable of rigor of teacher questioning was correlated the highest with the other variables. Experienced teachers selected and implemented tasks at higher levels than new or special education teachers. Teacher self-reflection following lessons along with pre- and postsurveys suggested that teachers most wanted to increase student participation in and the quality of student classroom discussions. The next most selected goals for
improvement were teacher linking and pressing students to explain and justify their reasoning.

All student dimensions were rated lower than teacher dimensions at the beginning of the study. However, student participation, linking, student responses, and quality of student discussion were among the top five dimensions in improvement.

Chapter 5 presents a comparative case study of three teachers representative of the three groups: new, experienced, and special education. Their experiences in this video lesson study analysis are explored in order to understand the process of teachers learning to increase the cognitive demand of their task implementation through student reasoning during classroom discourse. Findings are also discussed with limitations of the study and next research questions explored.
Chapter Five

Chapter 5 situates the quantitative and qualitative results from Chapter 4 in the context of the teachers’ experience. I wanted to understand the learning process during the video lesson analysis study from the teacher perspective. I also wanted to compare and contrast the professional learning process of new teachers to special education teachers and experienced AFDA teachers. I had compared the results of the three groups in terms of the IQA instrument across the dimensions of Academic Rigor and Accountable Talk. Next, I wanted to understand the numerical data and themes that had emerged from the data by studying them in the context of the teachers’ experiences over the four lesson cycles. The themes included student engagement, patterns in classroom discourse, questioning, and coaching/professional learning. To do this, I conducted a comparative case study of Will, Sarah, and Ashley, teacher representatives from each of the groups.

Division Impetus for the Design of the Professional Learning Model

Cohen and Ball stated in their 2001 study that teachers needed to become serious learners of practice rather than learners of strategies and activities. To this end, I implemented the video lesson analysis project with the entire AFDA team of teachers as embedded professional learning for one calendar school year to study their practice of task implementation. The nine teachers broke out neatly into three groups: New, special
education and experienced. I defined new teachers as those in their first or second year of teaching AFDA and those experienced as having more than two years of experience teaching this course. Special education teachers were coteaching in an AFDA course with the new or experienced teachers. Our division needed a format for coaching teachers in how to teach mathematics through rich tasks, projects and problems. The AFDA curriculum required this teaching approach, and as teacher assignments changed, the division needed a way to train new AFDA teachers in a cost and time-effective way. Secondly, since AFDA serves struggling learners who need the course credit for graduation, including many special education students, it was imperative to focus on special education coteaching teams in the project. Training a highly mobile group of teachers of a niche-course (about 24% of Juniors) to teach a task-based mathematics course without a textbook to struggling learners required a new type of professional development model. I designed a personalized, digital coaching model to fit this need.

**Teacher Selection**

There were three new AFDA teachers; one was an experienced teacher in her first year of teaching AFDA, and the other two were second year teachers. I selected Will, using purposeful selection, because I wanted to understand deeply the experience of a teacher new to the profession, and he was remaining in the course following the study, while the other new teacher had been reassigned to other courses. Of the three veteran teachers, I selected Ashley for the case study, because she had the most experience teaching the course and was the only remaining founding member of the AFDA team. I selected Sarah for the special education teacher case, because she had been videoed both
as a coteacher and in one of her self-contained classes for all four cycles, so I could study both of the typical settings of a special education teacher. In all three cases, the teacher was videoed with the same section of students across the four lesson cycles with one exception. Will’s baseline lesson was recorded from a different section of the same course. His next three cycles were in the same course and the same section. I observed all 12 lessons in person for the case study teachers over the four teaching cycles.

**Teacher Background**

Will had been hired in the division as a first-year teacher the year before the study. He was a secondary mathematics education major with a statistics minor from a college in Western Pennsylvania. During his first year at the high school, he immediately became a contributing member of the school community as both a mathematics teacher as well as a soccer and basketball coach. He taught Algebra I and AFDA, which were courses for nonaccelerated students. He also did test prep and remediation for the Algebra I state test during math lab. This study began during the summer following his first year. Will was 24 years old and was scheduled to teach AFDA and Probability & Statistics during his second year of teaching.

Ashley was in her early forties with close to 20 years of teaching during this study. Her teaching career had begun as a special education mathematics teacher and following certification courses in secondary mathematics, she had been teaching math at her high school for ten years. She had been a part of the original design team for the AFDA course and was in her fifth year of teaching AFDA during the timeframe of this study. As part of that team, she had taken two graduate mathematics courses in the MKT
of functions and probability and statistics, offered by a Mathematics and Science Partnership grant from the state through a partnering university. Following the courses, she had worked with the team to design the curriculum over a five-year period, involving six paid collaboration days for embedded professional development per year with the AFDA team. Although she did not teach AFDA until the second year, she had been part of the design team during the first year and was considered to be very knowledgeable of the course and its evolution. She had coached the other AFDA teachers at her school and during division meetings. Additionally, she had taken the initiative to invite AFDA teachers from neighboring divisions to join in the collaboration days. Ashley had been teaching five sections of AFDA for several years at her high school at the time of the study.

Sarah was in her late forties and in her 14th year of teaching special education after switching careers. She had been coteaching AFDA for two years in two sections with Ashley and was also teaching Math Applications to self-contained students. Math Applications was a course for special education students working toward a Modified Standard Diploma, the equivalent of eighth grade mathematics. This diploma option was being phased out in the state, with the Standard Diploma being the minimum diploma option. The significance of this diploma change was that similar students would soon be taking AFDA, which had a much higher degree of mathematical content, for their third credit for a Standard Diploma. Sarah had been teaching special education English prior to being assigned to AFDA two years prior to this study. Since she had begun teaching math, Sarah had taken advantage of two major professional development opportunities
including a week-long NASA course for mathematics and science teachers integrating curriculum through investigations and the use of simulation software.

**Goal Setting**

Goal setting was a key component to this professional learning model. During the two summer collaboration professional development days, Will worked with two fellow AFDA teachers from his school to view and rate his baseline video lesson with the TAG and IQA rubric for Academic Potential (Boston, 2012; Stein & Smith, 2007). Will’s team gave face-to-face feedback to him against the criteria from the IQA instrument for the 10 dimensions. Will gave the other two teachers similar feedback on their baseline video of a cotaught lesson from the same probability unit. Then each teacher set an improvement goal for the study time period from the dimensions of Academic Rigor and Accountable Talk during task implementation. Will chose “student participation in discussion and task implementation” as his goals. Ashley worked with Sarah and a veteran AFDA teacher, Brian, to view her baseline video and to set her goals. She decided to set four goals: Student participation in discussion, teacher linking, teacher press for students to explain and justify and rigor of teacher questioning. Sarah chose three goals: Student participation in the discussion, student linking (connecting and relating) and student discussion after the task. Teachers set goals for each of the four lesson cycles over the course of one school year, based on feedback data from their lesson videos. Results indicated that teachers improved in goals they focused on for an individual lesson as well as for the larger study.
Case Studies: A Comparison of the IQA Dimensions

In this section, I will compare the evolution of the 10 dimensions of the domains of Academic Rigor and Accountable Talk in the tasks across the three representative teachers. Key themes that emerged from the data analysis in Chapter 4 included: task selection vs. task implementation, student engagement, participation patterns in classroom discourse, rigor of teacher questioning, and coaching during professional development. I will describe how each teacher fared in meeting instructional improvement goals.

Academic rigor: Task selection. The quantitative data from the IQA instrument indicated that study teachers improved in rigor of selected tasks from means of 3.05 to 3.43 over the four lesson cycles. The mean for experienced teachers across all tasks was 3.49 as compared to means of 3.03 for new and 3.17 for special education teachers. This meant that AFDA teachers generally entered the study already selecting high-level tasks, most-frequently Procedures with Connections (Boston, 2012; Smith, 2007). This was to be expected since the whole AFDA team had been evaluating the cognitive demand of their tasks for the three school years prior to the study using Stein and Smith’s 2007 Task Analysis Guide.

Baseline lesson. Similarly, two of the case study teachers, Ashley and Sarah, selected high-level tasks in the first lesson cycle at 3 on a 0-4 scale; Will’s first task was rated as requiring lower-level thinking at a 2: Procedures without Connections. Will, as a new teacher, had had less experience with the AFDA team in task selection.
Will selected the Birthday Problem, which asked the students to calculate the probability of two people in a room having the same birthday, by substituting into a given formula. His baseline video was taken during the probability unit at the end of his first year of teaching. Ashley also gave the students a problem question to investigate: How many great + grandparents does this class have in any given generation? She gave this task in the early fall during the functions unit. She asked her students to generalize to a rule to figure out how many great + grandparents one would have per given generation.

Sarah’s baseline lesson was taught to her self-contained class on the eighth grade standard of interpreting graphs. Instead of one rich task being the lesson for the 90-minute block, Sarah’s lesson contained a series of smaller rich tasks. She began by activating prior knowledge by giving them an unlabeled or scaled graph and having students guess the context. As they asked only “yes” or “no” questions, she gradually digitally revealed more information on the graph displayed on her electronic whiteboard. During the new learning, students analyzed graphs to form inferences and held whole-classroom discussion of what each graph meant. To try out the new learning, students analyzed different graphs in a given context to form inferences and make predictions. In comparison, the experienced AFDA teacher used the selected task as the lesson and worked through the launch, explore, student presentation and summarization phases, while the new and special education teachers chose a traditional lesson format.

Observer feedback. Observers, including the teacher and researcher, rated Will’s Birthday probability task as having an academic rigor of 2 (lower-level), Ashley’s Grandparents Problem as Procedures with Connections at 3 (higher-level) and Sarah’s
graphing lesson as Procedures with Connections 3 (higher-level). Feedback to these three teachers over the next three lesson cycles is summarized in Table 14.

Table 14

*Academic Rigor: Potential of Task*

<table>
<thead>
<tr>
<th>Academic Potential of Task Means</th>
<th>Lesson 1*</th>
<th>Lesson 2*</th>
<th>Lesson 3*</th>
<th>Lesson 4*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will (New)</td>
<td>2</td>
<td>3</td>
<td>3.2</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>What is the probability that two people in a room will have the same birthday?</td>
<td>What is the relationship between height and time for a balloon to drop?</td>
<td>About how many licks does it take to get to the center of a lollipop?</td>
<td>What are the chances that you would select a particular color of M&amp;M and a Skittle?</td>
</tr>
<tr>
<td>Ashley (Experienced)</td>
<td>3</td>
<td>3.6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>How can you figure out how many great+-grandparents you have in any given generation?</td>
<td>How did the extinction of the wooly mammoths look like in terms of change in population over time?</td>
<td>What is the relationship between a quadratic equation and its table and graph?</td>
<td>Predict the graph following the transformation of the parameters of a quadratic or exponential equation.</td>
</tr>
<tr>
<td>Sarah (Special Education)</td>
<td>3</td>
<td>3.5</td>
<td>3.3</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>What could this (unlabeled) graph represent in real-life?</td>
<td>How can I use angle relationships to justify missing angle measures?</td>
<td>How can I balance an equation using a scale?</td>
<td>How can I write and solve equations from pictorial representations?</td>
</tr>
</tbody>
</table>

*Note.* *0* = No task; 1-2 Low-Level Task; 3-4 High-Level task.
Will: Evolution of task selection. While Will began his baseline lesson with a lower-level task, his ensuing three tasks were higher-level requiring Procedures with Connections. In his baseline, Will selected a whole group task with a driving question to answer with probability as the content. The task itself was presented on a worksheet with scaffolded questions built to lead students to the correct answer. Will kept this general task format throughout the school year but changed the implementation of the task greatly over the course of the study by turning over much of the decision-making, exploration, explanation and justification over to his students. This will be explored further under task implementation. What did change in Will’s task selection was a move from a canned textbook-type problem in the first lesson as the driving question to solve from substituting into a given formula, to a real scenario in lessons two, three and four where students collected data and then built a mathematical model to answer the question. In doing so, Will’s tasks were rated as requiring Procedures with Connections. Will’s essential questions were strong in terms of the essential understandings of mathematics, and he connected those ideas to the scenario in all four tasks.

This new teacher’s strong academic background in mathematics showed in his writing of the task questions. Will typically took a problem or task from a textbook, the curriculum map or online source for the driving question. However, he wrote all of the scaffolding questions and designed the worksheet so that students would build understanding towards a logical, efficient solution. The student was led down a path that, if followed, would solve the problem without having to struggle to think of the solution path. These questions were preplanned to be the focus questions for the classroom
discussion following the task. What opened up in the task questions over time, was the acknowledgement that there may not be just one path or one method to reach the solution. Will also moved toward writing his own tasks from real-world data or scenarios of interest to his students. “When I moved here from the city, I knew nothing about NASCAR or hunting, but I have learned what my students are interested in and tried to write tasks that reflect that” (Coaching Conversation, Task 2). Finally, Will’s final lesson questions pressed students far more to struggle with explanations of reasoning and to justify solutions than in the initial Birthday Problem which had been heavily scaffolded.

**Ashley: Evolution of task selection.**

*Use of context.* Ashley’s first two tasks were similar in that in each she posed a driving question which students explored in small groups by collecting data, organizing the data into a table and graph and then discussing the results and inferences during a whole class discussion led by the teacher. Both had a context such as the number of grandparents in the class or extinction of mammoths over time; this context gave the students a question to ask and a purpose in collecting data to answer. These tasks were ranked as higher level with scores of 3 and 3.6 based on the real-world connections. Her last two tasks lacked any real-world context, yet they were rich examples of tasks that explored intra-disciplinary connections between the representations of a function. In the fourth lesson, by predicting the behavior of the graph from the parameters of a transformation in the equation, students had to use nonalgorithmic thinking to prepare a poster comparing their prediction of different graphs based on how their equations had changed. Some chose to draw a table first and then a graph. Others went straight to the
graph. All had to explain their reasoning for how they knew how the graph was going to move based on either their equation, table or both. Observers evaluated this task at a 4, or “Doing Mathematics.” Thus, the first two tasks were scenario-based while the final two were Procedures with Connections that were intra-disciplinary, featuring students forming connections between representations and predicting changes in one based on changes in another representation.

**Thinking level.** Ratings for Ashley’s four tasks were higher-level in 13 out of 15 observations as shown in Table 15.

### Table 15

<table>
<thead>
<tr>
<th>Teacher</th>
<th>TP Level Low-1</th>
<th>TP Level Low-2</th>
<th>TP Level High-3</th>
<th>TP Level High-4</th>
<th>Total Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Ashley</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td>Sarah</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>12</td>
</tr>
</tbody>
</table>

The Grandparents Problem was posed as a question with a model worked through as a class where students first figured out together in a whole group how many total grandparents the class had. The purpose of the investigation was to use strategies to figure out how many great-grandparents or great-greats, etc. The higher-level thinking came in when students were asked to generate a rule to know how many great-grandparents the class would have in any given generation. They were not shown how to do this. Students worked in groups of three or four to create a solution and then presented
their solutions on poster paper. Students chose how to represent their thinking; the teacher used the student work to know where in the learning progression each student was. Only a few students were able to generate an algebraic expression or equation for the rule. In the Mammoth task, the launch contained a discussion about what extinction meant and then students were given a procedural data collection simulation with dice to model the extinction of the mammoth population over time. The cognitive demand of the task moved to a higher level when students engaged in mathematical modeling. Students predicted and sketched the graph of the mammoth population over time from the simulation data and determined the type of function that would best model those data. In summary, all four tasks offered rich connections. Ashley moved from real-world connections to intra-disciplinary connections between representations largely due to the content in the units. She also was trying to better align with the more theoretical Algebra II course that she was preparing the students for as the year progressed.

_Sarah: Evolution of task selection._ Sarah’s tasks were all rated as high-level thinking (Table 16); eight were coded at a 3 and five on a 4 on the scale. Peers rated her first lesson with a mean of 3 and the rest at 3.5, 3.3, and 3.7. Each task was open-ended and had a twist to it to increase decision-making by the student. For example, in Lesson 1, Sarah removed all labels and scaling on a graph and asked students to form hypotheses as to what the graph represented. Students had to explain and justify their ideas. As they asked “yes” and “no” questions, Sarah gradually revealed the labels and scale. The main difference noted between Sarah’s tasks and the other teachers was that she presented a series of small tasks rather than one large task for the entire lesson. Her reasoning was
that her students needed to switch gears and do something different about every 12 to 15 minutes to hold their attention and she needed to scaffold more with a series of related, sequential tasks.

**Academic rigor: Task selection vs. task implementation.** Across all nine teachers and through the progression of the four lesson cycles, Task Implementation followed the same pattern of increases that were seen with Task Potential but always with slightly lower ratings. Overall, participating teachers selected higher-level tasks (3.24) and implemented them just under high-level implementation (2.99). Peer observers and the teachers coded the four lesson cycles for Will, Ashley and Sarah as shown in Table 16.

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Level Low-1</th>
<th>Level Low-2</th>
<th>Level High-3</th>
<th>Level High-4</th>
<th>Total Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Ashley</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Sarah</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>12</td>
</tr>
</tbody>
</table>

Overall, the experienced teacher implemented tasks at a higher level of thinking than the new or special education teachers. Will solidly fell into a Level 3 category, Ashley between a 3 and 4 and Sarah at a low Level 3. Ashley and Will’s ratings showed more variance. Ashley’s mode was Level 4 “Doing Mathematics” while Will and Sarah’s were Level 3 “Procedures with Connections.” The task trajectory showed where the
higher-level tasks became lower during implementation and where they remained high (Table 17).

Table 17

Task Potential vs. Task Implementation

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Frequency Lower Level TP</th>
<th>Frequency Higher Level TP</th>
<th>Frequency Lower Level TI</th>
<th>Frequency Higher Level TI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will</td>
<td>1/14</td>
<td>13/14</td>
<td>4/14</td>
<td>10/14</td>
</tr>
<tr>
<td>Ashley</td>
<td>2/15</td>
<td>13/15</td>
<td>3/15</td>
<td>12/15</td>
</tr>
<tr>
<td>Sarah</td>
<td>0/12</td>
<td>12/12</td>
<td>4/12</td>
<td>8/12</td>
</tr>
</tbody>
</table>

In no case did tasks requiring lower-level thinking become higher-level. With new teacher Will, three of his observations showed a drop where tasks with potential for higher-level thinking dropped to lower level during implementation. One of experienced teacher Ashley’s observation ratings dropped to lower-level demand during implementation; all other selected high tasks remained high during implementation. The special education teacher Sarah had four out of her 12 observations show a drop in rigor during the lesson trajectory. In summary, once I knew whether the teacher was selecting tasks requiring a high cognitive demand, I looked at which ones remained high during the task trajectory. After figuring out which tasks remained higher level and which did not, I wanted to find out why.

The following case study stories describe how task implementation looked over the four lesson cycles during the school year in Will, Ashley and Sarah’s classrooms. The purpose of delving deeply back into the video analyses was to better understand the
themes that had emerged from the data sources as outlined in Chapter 4. To understand how the task trajectory of a higher-level task remained higher-level or was reduced to a procedural task, I looked at four themes in each case study: student engagement, participation patterns in classroom discourse, rigor of teacher questioning, and coaching during professional development.

**Will’s Experience as a New Teacher**

**Will’s baseline task.** Will’s baseline task, the birthday problem, was rated as a lower-level Procedures without Connections-type problem (level 2) and remained so during implementation. The task asked an open-ended question from a worksheet that may have had some potential for problem-solving had it remained open. “We are going to work through the problem and hopefully by the end you will be able to tell me how many people you would have to have in the same room to have the same birthday.” However, it was heavily scaffolded both by the worksheet questions and even more by the teacher as the task unfolded. Will projected the worksheet on the electronic whiteboard and wrote in all answers to each problem, which the students copied onto their worksheets. He told students which principle they had studied to use and gave them the formula, and then told them which numbers to plug into the calculator to first calculate the probability of not sharing the same birthday. The dialogue exchange follows:

\[
\bar{p}(n) = \frac{365Pn}{365^n}
\]

“where n represents the number of people. On ‘top’ [numerator], we will look at what’s called a permutation and in a minute I will show you how to put it into the calculator.
Any questions?” Then he showed students how to find the complement in order to find the probability of a duplicate birthday in the group of three people in the example, or 

\[ n = 3, \text{using the formula:} \]

\[ p(n) = 1 - \bar{p}(n). \]

Will checked each student to make sure he or she had entered the formula correctly into the calculator. Then he told students when they would know they had the answer. “Fill out the rest of the chart; enter the formula into the calculator exactly as I did. As soon as the probability gets greater than .5, that will be our answer.”

Will typically asked if there were any questions, and there were none in this lesson. Will then wrote on the board the numbers for students to divide in the example of three people. He never asked any mathematical meaning-type questions about what permutations were or why he was using one to solve this particular question. A peer observer described the lesson:

The teacher began by translating he problem, explaining what the problem wanted students to find out and then proceeded to tell them exactly how to solve it step-by-step by giving the formula and modeling how to substitute into the formula with steps on a calculator. For example, students were told four times the task purpose, to replicate the process and when they would know they had the answer. “We wanted to see how many people it would take to have in a room to get at least a 50% probability of two people having the same birthday.” Students completed a chart with the calculations as modeled to find \( n \) number of people where there was at least a 50% probability of sharing a birthday.
Engagement in the learning process. During Will’s baseline task, students sat silently in rows and followed directions by copying down the answers to the scaffolded questions on the worksheet from the whiteboard and then completing a chart with calculations from their calculators. Students did not speak to one another or ask any questions in the discussion. Two never spoke and appeared to not be engaged. Most participated when asked questions and appeared to be used to letting one student answer each question. They checked their answers in the discussion following the task. Will summarized the big idea of the lesson.

Classroom discourse patterns. The questioning pattern was a didactic model: teacher to class with one volunteer student answering each question. Will really did not know what the other students were thinking; if they could not answer a question, he answered it himself and explained. There were no opportunities for student-to-student discourse, therefore students did not link or connect ideas to one another; neither did Will. Will only pressed a student to explain his thinking twice, and both instances occurred when the student had a correct answer. After giving the assignment, Will did offer for students to collaborate, but the stated purpose was to “check answers.” It was interesting that Will was very focused in his questioning. He never asked nonmathematical or other mathematical-type questions such as other applications of permutations. He was “all business.”

Questioning. During this lesson, Will asked 24 procedural/factual questions and students answered 16 of them. For the remaining ones, Will answered them himself. Examples of these included recall knowledge like, “How many days in a year?” to
procedural steps such as “Multiply 365 x 365 x 365…someone tell me what it is.” Will used the teacher-to-one student response pattern, asking six Generating Discussion questions but in only one case asked more than one student to answer. In the opening question, he asked students to form a hypothesis as to about how many people there would need to be in a room to get a shared birthday. In this problem, he asked most of the students what they thought. This was the only break from the pattern of “teacher asks a question and takes one student response before moving on.” Twice, Will probed a student to explain his answer and twice he used a mathematical meaning-type question. In one he asked the range that the answer must fall between since it was a probability question, “since this number is dealing with probability, it will be between ___ and ___?” Had the question been asked in a more open-ended way, it would have moved from procedural/factual recall to mathematical meaning. The other issue was that knowing that probability ranges between 0 and 1 is a middle school concept. There was no meaningful discussion of permutation and complements, which would have been new learning for these students. In all, students answered the questions they were asked; they answered 24 out of 34 posed questions. Will kept the questions at the recall, memorization and procedures-without-connections level. Using Webb’s Depth of Knowledge, this task implementation would have been at a Level 1, or recall level of thinking.

**Professional development.** Will began the video lesson study with two days of professional learning with his AFDA team studying the rigor in task selection and implementation. All three schools did the task implementation analysis of baseline videos together on the same day; one school did the task selection day on the day after rather
than before due to a conflict. During one day, teachers studied the IQA rubric for Academic Potential of the Task and worked individually and in teams to evaluate existing tasks from their AFDA curriculum site. They worked as a team to modify and rewrite some of these tasks to make them high-level. They wrote new tasks and eliminated others that were lower-level. By the end of the day, AFDA teachers had come to agreement on about 21 tasks for use in their six units for the upcoming year.

On a second day, Will and his team worked on task implementation. First they viewed a video of an Algebra I teacher, Barbara, teaching a lesson on quadratic functions from a professional development site. Each teacher wrote a coaching reflection of what went well and offered suggestions for improvement. This reflection served as a baseline data point. Next, I shared the IQA instrument with all ten rubrics, explaining the way feedback would be collected from the IQA rubrics by peer observers, self-assessment by teachers and by me as a coach. The AFDA teachers then broke into self-selected groups of three by school to observe their baseline videos for the first time. Each school had one teacher who taught alone and one coteaching team of a special educator and a general education mathematics teacher. Each teacher watched his or her video individually at first, self-assessing with the IQA rubrics and then watched the video from the other team. They had a conversation in groups to compare three individual ratings and come to agreement. The groups revisited the videos to check out certain evidence in footage when discussing. The groups came to agreement on a rating for each of ten rubrics. They discussed what they rubrics meant and became familiar with the protocol for later video analyses.
The last activity was to discuss areas of strength and improvement in order to personally set an instructional goal for the next lesson. Teachers shared their goals with their small groups and in a whole group discussion. Most goals centered on increasing student participation and engagement in class discussions. Several teachers chose improving the rigor of teacher questioning. From his baseline video of the Birthday task, Will chose to improve both student participation in classroom discussion and rigor of his questioning. One experienced teacher chose to work on increasing mathematics residue or students taking away the big ideas in his lesson. Teachers later entered goals for each lesson into a Google form following each of the self-reflections during the lesson cycles. The goals they set in the summer aligned with the goals they kept once the lesson cycle began during the school year.

*Resources.* Because most teachers had selected improving classroom participation in classroom discourse. I facilitated teachers reading and discussing Reinhart’s 2000 article, “Never Say Anything a Kid Can Say.” This activity kicked off an example of the available online resources I had compiled for teachers to use during the study. I talked with them about self-selecting resources that interested them given their goal. I was curious how they would use this collection of about 200 online resources. I had envisioneded them using these resources as a research intervention following each self-assessment and analysis of feedback from peers.

*Will’s second task: Balloon drop.* Lesson Cycle Two occurred during the school year. Will’s second task, the Balloon Drop Regression Lab, was a math modeling data collection investigation where students worked in teams of three or four to collect data to
find the relationship between two variables, height and time during a balloon drop. This task was in the folder on the AFDA curriculum site and was used by all AFDA teachers. Peer observers rated the Balloon Drop Regression Lab as a higher-level task, Procedures with Connections, which is a Level 3 out of 4 on the scale. While there was a set procedure for the lab, the students had to work together to move fluently between the representations of data in a x/y table to create a graph and determine the equation that would best fit these statistical data. Reps had to go to the board to record group data and combine to find class means for the time it took the balloon to drop from each foot of the balloon drop. Students had to write an explanation of findings, including an explanation of the relationship between height and time and explain the physical representation of slope as a rate of change in terms of physical units of change in time per foot. They needed to express the meaning of the y-intercept orally and in writing as the time it would take to drop zero feet and why their intercept was not (0,0). Thus, students had to form inferences in writing based on the questions Will had prepared for the class discussion from the lab sheet.

Changes. From the data analysis, I noted the following changes in Will’s task implementation moves: allowed individual think time by placing thoughts in writing before asking for answers from whole class, utilized student groups of three or four instead of individuals, set group roles for the lab, had students record group data on board and calculate class means instead of doing it himself, encouraged students to estimate the line of best fit first before running the regression, rerouted some student questions back to their group for discussion, asked more mathematical-meaning-type questions and fewer
procedural-factual questions, restated student inferences to probe deeper into students’ intended meaning, and pressed all students to justify the relationship between variables time and distance in writing.

**Engagement in the learning process.** Will’s second lesson showed 100% student engagement. All students participated in the data collection lab with the balloon drop lab by actively creating a data table and graph as well as discussing the questions together before writing multi-sentence responses to the essential questions on the lab sheet. Will used several key moves to make this total participation happen. The first was that he arranged for students to work in groups of 3 or 4 rather than individually and assigned roles. Groups decided which role each person would have in the lab: scorer/recorder, timer with cell phone, and balloon measurer and dropper. Secondly, Will had student reps from each of the six groups record group data on a class chart on the whiteboard rather than doing it himself as he previously had done in Lesson 1. Students seemed comfortable comparing results and asking questions of each other. “I got 6.5 for 7 ft. drop and mine is way higher than yours.” Will redirected several students who asked him questions with, “I want you to double-check partners and check with each other.” Similarly, during the inference questions part of the discussion, students helped each other to write summary statements for the meaning of slope as a rate of change and the y-intercept from the balloon drop context of time and distance. Later, they helped each other run the regression on the calculator and use the values of slope and y-intercept to write the line of best fit. Unlike in Lesson 1 where every student worked individually and directed all questions to Will, in this lesson students directed most questions to each
other. Very few went to Will. It was obvious that this culture permeated the classroom and was not just apparent on this day in this one lab. The AFDA course was designed to be an inquiry course where students worked in groups most days on tasks and projects. Will had been planning with his AFDA school and division teams and was using the tasks the team had selected to go in each unit during the summer professional development days.

**Questioning.** The most obvious change in Will’s questioning was the balance of procedural/factual and mathematical meaning-type questions. In both Lessons One and Two, Will posed about the same number of questions, 34 vs. 30. However, in Lesson 2, Will asked about 47% procedural/factual questions as compared to 71% in Lesson 1. He spent more time exploring deeper cognitive level, mathematical meaning-type questions in Lesson 2 (27% vs. 6% of total questions). He asked about the same percentage of Generating Discussion questions at 18% in Lesson 1 as compared to 20% in Lesson 2. Will still did not generally ask probing questions to press students to explain or justify their thinking. He asked probing questions in only 6% of questions in Lesson 1 and 7% in Lesson 2.

Will’s higher-level thinking questions in Lesson 2 were about the new mathematics in the current investigation as compared to prerequisite recall or skill-level questions from previous courses. When asking students to form inferences in writing from the class data displayed in the table and graphs, he posed questions about the big ideas embedded in estimating a line of best fit from investigation data. Examples included: “How can you interpret the meaning of the slope value in terms of the lab?”
“How can you justify why you have a strong positive correlation?” and “What is the physical meaning of the y-intercept in this lab?” The way that he got all students’ thinking on this was to give them each ample time to write down their individual responses before discussing whole group where only those articulate students who were quick to answer typically shared their ideas. Will circulated the room, listening to students and reading their written responses. He also collected these written responses from their lab worksheet to assess later. He did not yet select and sequence student work to highlight the mathematics in the summary discussion (Smith & Stein, 2011).

**Classroom discourse patterns.** Will had improved the rigor of his questions and increased student participation in the discussion. However, the questioning pattern was still teacher to one student and he rarely asked more than one student what they thought. Even though he had all of the student answers to the essential questions to choose from as he circulated, he still took the first volunteer to answer in the whole group discussion rather than selecting and sequencing student work based on the mathematics or approach. Rather than hear several responses to the key questions, he took only one and modified it himself by asking that student leading questions to “fix” it in order to hit the mathematics he wanted. He did not yet form the class response from several different approaches/answers to the question that together could be edited by the class to be a complete response.

Will was restating key student responses to several essential questions in order to focus the class on their importance and used the talk move of “adding on” to edit to be more accurate mathematically. I had not observed any restating in Lesson 1. He also
asked a couple of deep reasoning questions that emerged in the “messy” data that were not on the lab worksheet or planned in advance. He pointed out to the class that the y-intercept was not (0,0) in the regression equation, as would have made sense from the balloon drop context, but had a drop time of 2.7 seconds for a distance of 0 inches. On the fly, he asked the class, “When the height is zero and the time is .27 (seconds), how does that make sense?” He also addressed their cognitive dissonance about time being the dependent variable in this lab. Additionally, he asked an extension question about the concept of residuals that was beyond the content expectation for this course after noting a related student comment. In these ways, Will’s questioning had become more responsive to students’ thinking as he worked through the lab. In contrast, during Lesson 1, Will never varied from the written questions on the worksheet, which formed his discussion questions.

**Coaching/Professional development.** Aggregate peer observer feedback to Will for Lesson 2 essentially showed him that his rigor dimensions had increased as had his Accountable Talk indicators. However, like the other teachers in the study, classroom discourse moves lagged behind the Academic Rigor dimensions of task potential, implementation, questioning, and quality of student discussion. The Accountable Talk indicators had means of mostly 2 on the 4-point scale while the Academic Rigor components had means of 2.75 except for Mathematical Residue, which was rated at a 2. Notably, the rigor of student discussion moved from a mean of 1 to a 2.75 and student responses increased from a 1 to a 2.25. Student linking was still the lowest dimension at 1.75.
Following the lesson observation, I met with Will to discuss the lab. He was pleased with the increased participation in the lab and in the discussion. He felt the rigor of his questioning had increased to a 3 on the scale yet there were missed opportunities. Having shown improvement in his goals, we talked about what he wanted to improve in his class discussions. I shared some video links from the resources related to allowing individual think time first for all students in order to engage in the problem. Following individual think time, I asked Will to consider using a strategy with students to compare individual ideas in a table discussion before going whole-group. Peer coaching advice also included increasing generating-discussion questions and think/pair/share structures to get more students involved in the discussion rather than selecting one volunteer for each question. While he had allowed ample time for written responses to the questions, he wanted to see how to connect and link student ideas to each other to build group understanding from individual contributions. I shared a link of a video of a middle school classroom led by a mathematics specialist as she launched, explored and summarized a task in a discussion. Will and I informally discussed how his classroom tasks were going as I stopped by for walk-throughs between the formal video lessons, since they were several months apart. The coaching intent was that he would practice with all of his classes over time and look for improvement in his goal areas in the next video lesson. He also planned to solicit feedback online from his former AFDA partner who had taken a job reassignment to a different location. The benefit of sharing digital lesson links made available on a Google Site with the rubric feedback data gathered through a simple Google Form, allowed peer-coaching feedback across the division. As part of the model,
Will viewed one video per cycle as a peer observer to provide feedback using the same IAQ rubrics. In doing so, teachers in the study had internalized the IQA rubric look-fors by the end of Lesson 2. In postsurvey feedback, teachers said rating themselves against set criteria was the most helpful of all of the activities.

**Will’s Lesson 3: The lollipop lab.** Will selected another hands-on data collection task for his third lesson. By this time, he was in the statistics unit where students were studying applications of one-variable descriptive statistics including z-scores in real-world scenarios. The task asked students to estimate about how many licks it would take to get to the center of a lollipop that had a chocolate nougat center. Will set four goals for the lesson in his reflection: increase student participation in discussion, increase teacher linking and connecting student ideas and relationships between students, improve the quality of the student discussion after the task, and increase the rigor of his teacher questioning. Data from the IQA instruments by seven peer observers, including the teacher, indicated that Will had met his goals. His top two increases overall across all 10 dimensions were in the goal targets of teacher linking and connecting (2 to 3.6) and rigor of teacher questioning (2.75 to 3.14). Student participation was a continued focus from the two previous lessons and continued to increase from 2 to 2.6. The quality of the student discussion following the task remained essentially the same at 2.71, only .04 lower than in Lesson 2. These quantitative data were confirmed by several of the peer observers,

This lesson went well. There were more questions asked of the students than the last lesson viewed (Lesson 2). I can tell where there has been some growth in
asking more probing questions. Students were engaged in the lesson and participated consistently even when they were unsure of the material.

*Engagement in the learning process.* Will’s reflection showed that he was pleased with the student responses to his questioning:

Every activity, from the first lesson to this third lesson continues to build student engagement and help(s) them connect ideas to what will happen in the real world. I liked the thought process of some of the students during this lesson and how they were thinking “outside the box” to help relate to real world. I could tell many of them were interested in the topic by some of the individual and group questions that they asked me while doing the experiment. This really was a better lesson and I believe I continue to build upon my questioning techniques each and every time I do a lesson.

A peer teacher from another school described the engagement from his vantage point as observer:

Participation was evident. I think the teacher led them a little too much. He showed connections but didn't let them discover them on their own, by presenting, asking questions and letting them work.

*Questioning.* Will recognized the need to expand his use of probing questions: “enhance my questioning techniques to ask probing questions and make the students do even more of the thinking/reasoning/and connections.” Another peer observer also urged Will to press students to explain: “Ask why. I believe that if a student can tell you why, they have learned it or at least at a level of cognitive thinking you want them at.”
Following this feedback from Lesson 3, Will set one of his target goals in his reflection to increase his ability to press students to explain by using probing questions. Will continued to increase the percentage of mathematical meaning-type questions that his students considered. I noted the following in my survey reflection, “…increased participation by students in classroom discussion. Some students were able to give a physical meaning of z-score with (the) number of licks of lollipop ex. what a neg[ative] z-score meant in this context.”

**Classroom discourse patterns.** A peer observer described the discourse pattern in his feedback:

Lesson was more teacher prompting students to answer as opposed to students explaining why the answer they got is correct. Maybe have the students write down their reasoning on why they got the answer that they did and have them compare them with other students. Maybe a Think-Pair-Share sort of deal.

**Professional development.** Will received feedback face-to-face from me and from the digital forms. Both fellow teachers and I urged Will to give opportunities for all students to respond to the questions other than the one who answers whole group. Think-pair-share was mentioned by some. My feedback to Will included the following notes:

Only 1-3 students were able to talk about each question in the class discussion with the interaction being teacher to student. Try a Think-Pair-Share...Try individual think time followed by sharing with either a partner or group of 3 and have each group form consensus and share out those groups you select based on what you hear them discussing when you circulate. Ask questions that require one
group to respond to the thinking or solution of another group--to link ideas
together to build a more complete understanding by whole class than just from
any one individual.

Another peer observer urged more wait time, which I, too, had noted as early as the first
lesson.

Just like the last lesson there was not quite enough (wait) time between asking of
the question by the teacher and an answer from the students. In some cases it
looked like the teacher was working harder than the students. Some suggestions
for next time: have a round robin discussion prior to the lab to discuss the
prerequisite knowledge.

In Will’s postlesson reflection, Will wrote that he would like to learn how to
differentiate instruction. Up until this lesson, Will had always run the labs in a lock-step
fashion where he asked the same questions to all students whole-group from the lab
sheet. Students wrote down the answers and then discussed in a whole-group summary.

He was beginning to struggle with giving up some control to the students.

In AFDA I continue to work at helping differentiate instruction throughout the
lesson. Sometimes it is good to have “checkpoints” where we come back for
discussion of the topic and make sure everyone is on the right track. However,
other times it is important to allow students to continue to work ahead and
investigate topics further and I would like to allow some of my activities and labs
to be able to do that.
Will and I had had the conversation about how to differentiate in mathematics through open-ended questioning and parallel tasks (Small & Lin, 2010). In Lesson 3, Will had tried this strategy in a small way by giving a choice of two warm up tasks. He showed how they were about the same topic and tried to connect the concept of standard deviation in both. However, he made the connection, not the students. He told me that he had tried to write one about cars for the students who liked to work on cars and the other task about SAT scores for students currently applying to college. Will was still direct teaching and dispensing the information. In his reflection, when asked about which questions he would research to find strategies to implement now before the next video, Will wrote about differentiation. He also responded to the peer feedback data on his need to increase use of probing questions to press students to explain why and to justify their thinking.

How can I differentiate instruction further? How can student(s) relate the ideas and topics in class with what they will see in the real world? How can I facilitate interest among students to get them to answer their own questions? Enhance my questioning techniques to ask probing questions and make the students do even more of the thinking, reasoning and connections?

In Lesson 4, Will made a large shift in practice as he tried to implement strategies to increase classroom discourse by increasing student-to-student discourse about mathematical thinking in small groups during the task and by letting student teams conduct the lab at their own pace. His role shifted in Lesson 4 to one of facilitator for the first time in this study.
Will’s Lesson 4: Conditional probability. The room looked and felt different when I entered Will’s classroom to observe Lesson 4. Student desks were now in groups of four and had been for a few weeks since the last video. Once the task started, there was a buzz in the room of students talking to one another about how to set up and conduct the lab and about the mathematics during their investigations. Students were engaged in the mathematics in that they were asking relevant questions, questioning each other, asking peers for help and finally utilizing sense-making. For example, one girl compared her results to a partner and asked if “…it made sense? Mine seems too high. What did you get?”

Will started off by giving his class three warm up probability tasks from a bag of 20 Skittles. I noticed that he did not tell students how to solve the three tasks. Previously, he would have heavily scaffolded the tasks both orally and on the worksheet. This time, Will posed the questions and then circulated listening to each group, asking probing questions, using wait time before answering student questions without telling how, and often restating and redirecting them back to discuss questions as a group. For the first time, I heard him say over and over, some idea or question followed by…”Think about it.” Students seemed excited, puzzled and curious by the tasks. Many were unsure of the mathematics, but they showed persistence by tapping neighbors and their teacher to answer questions not on the answer but on the process and what “or” and “and” meant in a probability task, for example.

Will pulled the groups back to discuss the warm up. At this time, he reverted back to his usual direct teaching stance where he asked a question, took one volunteer answer,
elaborated on the student response until it was complete and then moved to the next question. The class became silent again with just the volunteers speaking.

After the warm-up, Will launched the Skittles and M&M conditional probability experiment. He had activated prior knowledge with the Skittles warm up tasks. Students had practiced the prerequisite skills needed to conduct the experiment in the warm up with Will’s guidance. All Will did to launch was to pose the driving question and to discuss the protocol for getting lab materials of cups, napkins, M&Ms and Skittles. Then, amazingly, he turned the teams of four loose to conduct their own experiment and record individual, table group, and class data. He announced that he planned to pull them back in to discuss inferences at the end.

**Questioning.** From the beginning of this lesson study, Will had set a goal to increase student participation in discussions. In Lesson 2 even with a physically active, hands-on balloon drop lab, Will lamented, “How can I get all [not 75%] of my students [cognitively] engaged and in discussion about mathematics?” Will’s questioning changed in two ways in Lesson 4 as he moved into a facilitator role to try to increase participation. First, whole group discussion questions were cut about in half. Will typically would ask over 30 questions in a whole-group discussion of a task; he asked only 14 in the whole period in this lesson. His question and answer venue shifted to small group discussions, which could not be as easily captured by the microphone.

**Classroom discourse patterns.** Secondly, the questioning pattern changed from just teacher-to-one student to more student-to-student questioning with the teacher linking student comments together. Will let the student groups conduct the lab at their
own pace and did the questioning with small groups who were posting in their investigations. He asked more open-ended questions. Students actively discussed the same questions in a smaller group that they would have in a large group. The difference was that there were more “hits” per student in terms of speaking turns. The way that it differed from a whole group discussion was in the increased amount of error correction and misunderstandings being eagerly discussed in the safer environment that a small group offered. However, all of that talk focused on how to calculate the probabilities to get a correct answer and not so much on what it meant and why. Will posed mathematical reasoning and relationship questions when he would press a group to explain. He went after mathematical meaning when the opportunity presented itself from a student. For example, he pressed a student to explain in the context why “or” meant to add in a probability situation.

A peer observer noticed that the groups did not come back to discuss the big ideas together at the end. Will did not purposely select groups and certain solutions to share at the end to have a closure discussion about the big idea of the lesson. Not having a summary discussion meant that students did not have an opportunity to discuss the big ideas of the lesson as a whole group.

**Engagement in the learning process.** Will also noted that the cognitive demand stopped short of “doing mathematics” even though his students were participating and engaged in classroom discourse.

This lab was engaging for a majority of the students, however I would like to help them connect this idea with the bigger picture better. Answer the question: Where
is Conditional probability seen in the real world? Help them understand how this comes into place outside of class and outside of just eating candy. I also would like for all the students in the class to be able to determine how to talk about the numbers that they are calculating. Most of them see to understand how to get the numbers but the explanation part really matters as well.

Observers rated Will’s student participation in classroom discussions with a mean of 3.67 and quality of student responses in terms of explaining their reasoning also at 3.67. These results aligned with the rigor dimensions of task implementation at a 3, student discussion at about 3.3 and total mathematical residue of big ideas left from the lesson at about 3.3. In summary, participation was up, but the rigor of justification and conceptual understanding leading to real-world application and extended thinking was not yet at a 4-level. Will reflected on his next goal, “How can I get students to easily connect more difficult and generic topics in math to what they will see in their future? Out in the real world.”

**Professional development.** Following the final lesson, I met with Will during a study hall period where he shared strengths and weaknesses of his last lesson. He seemed pleased overall in reaching his goals of increasing student participation in the tasks and especially in the discussions of the mathematics. By moving to small group discussion, he felt student participation in explaining reasoning and working through misconceptions improved. He had tried several strategies during the year such as giving students individual think-time on a task first before working as a group and by adding an open-ended exit reflection for formative assessment. He also had successfully turned over the
individual, table, and class data collection over to the students. However, he stopped
short of highlighting, selecting, and sequencing student work to get closure on the big
idea of conditional probability as part of a summary discussion.

Will was at the point now of moving his students to a higher cognitive complexity
by asking them to move beyond application in the immediate context to beyond on the
real-world. He ended up doing this through his Probability Games Showcase projects that
followed this fourth lesson.

**Final study professional Development.** At the end of the school year, all of the
current AFDA teachers, as well as some scheduled to teach AFDA next year, met for a
day of as the final video lesson study event. This day was funded by professional
development funds by the division. For a half-day teachers collaborated, giving feedback
to each other on final lessons. Some teachers requested certain AFDA teachers to watch a
particular video segment and respond. I asked others to respond to those who had gotten
fewer “hits” or peer observations. Participants wrote their final personal reflections, took
the final postsurvey on the purposes and uses of performance tasks, and viewed the
“Barbara” video to write a coaching reflection as they had done as a prestudy activity.
Study teachers gave specific feedback on the effectiveness of different components of the
professional learning model, including adjustments. A big highlight for me was when
teachers shared tips with each other about how to do certain strategies such as use of the
calculator and digital formative assessment. I invited our Instructional Technology
Supervisor to attend the meeting and observe our process. This administrator had seen the
videos through the Google Site and Google + posts so was familiar with the project and
had helped since its inception with the logistics. He shared his insight and advice on our model saying that it represented where we wanted to go as a division to personalize professional development and have teachers drive what they wanted to learn.

**Will’s role.** Will took a very active role in the final PD day, emerging as a leader. I noticed that one of the other new AFDA teachers sought him out explicitly to give feedback on this lesson on expected values. After receiving the digital survey feedback, the teacher responded, “Thanks. That is very useful. I never would have thought of that. I will try that next time.” Will also explained why we were doing certain VLS tasks such as the survey and peer observations to those who were not clear. He helped one of the special education teachers navigate both the event and the Google tools, with the Internet being irregular that day. He continued to serve as a leader in the second part of the meeting, where schools planned action plans for AFDA next year.

**Summary of Will’s experience.** Will began the study as a second-year teacher working from heavily scaffolded tasks that did offer varied thinking levels in the questions. While most questions were procedural/factual and closed, Will did offer some open-style key essential questions. These questions were open so that students could explain their thinking. The problem was that they were not answering these questions in a whole-group discussion. Will’s student participation was lower than he wanted with all discussion funneled through a teacher-to-one volunteer student response pattern. Student responses to open questions requiring mathematics reasoning about relationships were incomplete and often offered as brief, computational answers. Data showed that students
often did not reply to these type of questions, waiting for Will to “tell” them the answer. At best, one student would respond, leaving Will to elaborate on that incomplete answer.

Will used several strategies to increase participation and engagement in the lesson. First, he began using hands-on data collection-style tasks with real-world contexts. He moved from students working individually in rows with a worksheet to students working in a small group to collect data and then moving back to individual answering of worksheet questions. In his final task, Will tried to have groups move past the data collection phase as the reason they were working in a group. After collecting individual and group data together, groups worked to discuss the mathematics resulting from analysis of the class data and essential questions. Will had moved past his fear of students “copying” others’ work to seeing the benefits of students constructing understanding by talking through the task questions with each other. Will reflected in his postsurvey on the most important thing he had changed in his classroom discussions:

The most important thing I have done to change classroom discussions is firstly creating an environment that allows students to feel comfortable answering questions and taking a chance on giving their answer even if they feel they might be wrong. Allowing students to disagree and correct one another and creating an expectation that all students are expected to participate in the class will also help me generate better more in depth classroom discussions.

Will also began to use formative assessment strategies to gauge whether students understood in the moment rather than relying on the thinking of the one volunteer who answered his questions or waiting until he collected the worksheet at the end of class. He
served as a facilitator by listening to individual and group discussions and offering questions to restate, link and connect student ideas rather than to explain the concept and solution himself. His questioning was strong both on the task sheet questions and on the fly. He was improving his ability to ask the right “next question” by listening for student misconceptions and correct strategies and knowing when to interject to move learning forward. Will summed up in his reflection how his role had changed in his use of performance tasks:

I think my performance tasks have changed in that I now let the students facilitate the lesson. I try to be a mediator but I let them come across the problems themselves and actually discuss with each other what they could do in order to solve the problems. Before I might have warned them [about] the problems that would arise and what they possibly could do to fix it but now I let them discuss with each other what they could do to fix the problem.

As a new teacher, Will had set two initial study goals: to improve both student participation in classroom discussion and rigor of his questioning in task implementation. Both quantitative and qualitative results showed that he had reached his goals during this video lesson analysis study. Will felt by observing and self-analyzing his own practice, using feedback from peers, coach and other resources, he could identify what to change. He then selected a few strategies to use and then observed in the next video whether student participation and rigor of teacher questioning had improved.

Will listed what had worked for him in his reflection: “Peer observations, self-analysis of your teaching videos, talking about your lesson with other teachers/researcher,
video references about instruction.” The next case study describes the experience of a veteran AFDA teacher participating in this same video lesson study professional learning experience.

**Ashley’s Experience as an Experienced Teacher**

**Ashley’s baseline task.** Ashley set four goals from her baseline task video analysis: student participation in discussion, teacher linking-connecting student ideas and relationships, press students to provide evidence to support their thinking/conceptual explanations, and rigor or cognitive demand of teacher questioning.

As described earlier, Ashley posed an open-ended question to her students: How can you figure out how many great-grandparents you have in any given generation? Use your rule to figure out how many of these ancestors are in this class in any given generation. She expected students to create a table of values for number of grandparents per generation and then look for patterns in the table to write the rule as an algebraic equation. Ashley used the launch, explore, summarize task implementation model for this lesson. She launched the task by posing the question, making sure that students understood the question by modeling how to figure out how many grandparents and great-grandparents existed in the room of 16 students. She broke the students into groups and asked them to explore the problem and design a solution to share with the larger class on chart paper.

**Engagement in the learning process.** By the end of the lesson, Ashley pulled the groups back to discuss their rules, but she felt fewer than 25% participated in the large class discussion following the task. Other raters rated participation as 25% to 50%, but it
was definitely low. My coaching suggestion was to have selected groups post and share their solutions by explaining their tables, graphs and equations on the chart paper in an oral presentation before sequencing selected student work to summarize key mathematical ideas in a whole-group discussion.

When asked what she was happy with in terms of the cognitive demand and student reasoning during the class discussion of the task, Ashley gave a brief response: “Like the picking of the groups. Like the discussion in some of the groups.” She wanted to change the student output in terms of written solutions by each student: “I would like to provide more independent time to get answers. Would like to see more written answers and responses.” This aligned with Ashley’s concern about Barbara’s lesson in the prestudy video reflection where she felt the teacher should make the students show their written work. Ashley did not reflect on the discussion itself in terms of whether students could explain their reasoning and whether they got the big idea of her intended learning target. When asked what her next goal would be for the next lesson with this group of students, she said, “Better engagement of all students.”

**Classroom discourse patterns.** However, when I checked other data sources such as the survey ratings from the IQA rubrics, I was better able to understand how Ashley viewed the cognitive level of student responses and her own rigor of questioning. She described the discussion: “Speakers do not back up their claims or explain reasoning. (Students)…provided brief answers OR nonmathematical responses and that she asked procedural /factual (PF) questions that elicited facts or procedures or require brief, one-word responses. At one or more times, students link ideas, but did not show how ideas
relate to each other or only one strong effort is made to connect.” (rubric language adapted from Boston, 2012).

**Questioning.** Boaler and Humphreys (2005) recommend preplanning the questions that will lead to the mathematical big ideas of the lesson and focusing questioning on those. The reviewers agreed that the mathematics was only partially developed, and that the students did not come to closure on the big ideas. The main reason was that Ashley spent a long time on the launch and did not have time to select and sequence students to present their solutions based on the mathematics she wanted to bring out in the lesson. Therefore, she quickly summarized at the end of class what they should have found as the rule in an equation or formula to figure out the total number of grandparents. She tried to pull out ideas from students, but as a group, they did not largely participate with deep explanations of their thinking. A second observation was that she spent time asking about 15 nonmathematical or other-mathematical questions that caused both the launch and summarizing discussions to lose focus. I coached her to tighten up and focus her questioning by preplanning the essential questions she would ask during each phase of the lesson. I coached several of the study participants to predict student misconceptions and then to write down key questions to address these confusions in their lesson plans. Wiliam argues that asking high-quality questions may be the most important action we can take to improve the quality of student learning (Wiliam, 2011).

Based on rubric data, at least two times, Ashley asked questions that explored mathematical meaning of relationships, probed students about their initial thinking, and generated discussion-type questions to bring in more students in to the class discussion.
However, she did not link ideas or consistently press students to explain their thinking. In contrast, Ashley did not think that she had asked any deeper questions, rating her questioning as only procedural and factual. Throughout the study, Ashley was critical of her practice whereas Will and Sarah rated themselves higher than peer observers in almost every case. In reality, Ashley’s ratings were highest for cognitive depth in both task selection and implementation.

**Coaching/Professional development.** Following her baseline lesson, Ashley became excited about the Jo Boaler videos on classroom discourse from the online professional development resources on the division math site. I met with her during a lunch period to discuss the strategies in the video and to hear which she might try in her classes. She also asked a special education teacher in the building named Veronica to observe her videos and to provide feedback. Ashley respected her opinion, because she used to coteach with this teacher and knew Veronica to be a teacher leader. Veronica had a strong reputation in the school as National Board certified-teacher who had high expectations for special education students. This ability to choose peers to give feedback turned out to be one of the real strengths of this PD design from the teachers’ perspective. This choice kept teachers in control of their own learning and was balanced by the consistent feedback from one coach or peer throughout the cycles. Peer observers from the AFDA team rounded out the feedback team. Multiple sources of feedback increased trustworthiness of findings for teachers.

Either the AFDA team or I trained peer observers like Veronica on the video feedback process and rubrics before each began. Veronica commented to me that this
process was exciting, because it reminded her of the process she had completed for National Board Certification. Following her first observation, Veronica improved the process by creating a video viewing form to use while tallying question-types and recording examples of questions and suggestions. This evidence became very handy for observers to use when completing the rubrics on the online Google form following the video observation. I placed the viewing form on the lesson study site and emailed all participants to let them know how it had been developed. Its use caught on quickly as it provided a focused way to capture the needed evidence. This is an example of how the participants shaped the professional learning model as a prototype in an iterative design. Ashley also gave feedback on a peer’s lesson through the same video lesson study process thus becoming very familiar with the expectations of the 10 dimensions on the rubric.

**Ashley’s Lesson 2: Wooly mammoth modeling task.** Ashley set the same goals for her second lesson: student participation in discussion, teacher linking by connecting student ideas and relationships between, student linking by connecting/relating, and rigor of teacher questioning. She added mathematical residue, or the degree to which the lesson left behind the big mathematical ideas with students. The task was modeling an exponential decay by conducting a simulation using dice of the wooly mammoth population over time before extinction. Ashley launched the task with a long discussion about extinction in general and mammoths in particular for six minutes. She then explained how to collect data in the simulation with the 20 dice representing the herd. In her reflection, she recognized that she had asked too many nonmathematical questions,
about 37, or 55% of her 67 questions in the whole class discussion portion of the lesson. This launch did not leave her adequate time for closure on the big idea of an exponential decay fitting the data. Students collected data, tallied results, made a table, graphed the data and decided which function would best fit the data. Pairs rolled the dice to collect the data and then completed the data analysis individually with the task sheet. Ashley conducted a large-group summary discussion for seven minutes at the end of class to formatively check which type of function’s graph they felt fit the statistical data and why.

**Engagement in the learning process.** Engagement varied based on the activity. Forty percent of the observers said there was 25-50% student participation in the discussion and another 40 percent said 50-75%. During the launch, participation was less than 25%. During the simulation, nearly all students were engaged in rolling the dice, completing the table and plotting the points on the graph. During the last seven minutes, Ashley selected two students to draw their different graphs and then asked the class to decide which graph would best fit the statistical data. 100% of students voted when polled as to whether the curve of best fit would be quadratic or exponential. In this way, Ashley formatively checked for understanding of which function would best fit the mammoth data.

**Classroom discourse patterns.** In the launch, Ashley followed a didactic pattern of communication by asking students questions about mammoths, extinction and endangered animals to activate background knowledge. She asked questions to generate discussion, and then called on the volunteers. None of the questions in the launch were mathematical. In contrast, during the paired exploration phase, Ashley asked individual
students questions about their data collection protocol as she circulated, pressing them to explain what they had noticed about the shape of the scatter plot. In the whole class discussion during the summary phase of the lesson, Ashley selected two students who offered different ideas on the shape of the graph. She probed their thinking and pressed them to sketch the plot and to explain why they thought it behaved as a linear, exponential or quadratic function. When one student could not justify his inference, the teacher called on the class to confirm by vote to show what they thought. In this way, Ashley linked and connected students’ ideas and involved the group in reasoning to add their own ideas to link to add on to one student’s idea. However, the discussion was not yet student-centered as the students did not link or connect to each other but always answered back to the teacher. The two students who had different hypotheses were not speaking to each other to explain and defend their ideas but through the teacher as interpreter. Thus, Ashley was modeling the following talk moves: revoicing, repeating, reasoning and adding on and wait time until students could deploy the moves student-to-student through linking (Boaler & Humphreys, 2005; Chapin, O’Connor, O’Connor & Anderson, 2009).

**Questioning.** Ashley asked 55% nonmathematical questions when she got off-topic during the launch and a second time for a total of seven minutes. She asked 15% generating discussion questions and then followed up by probing with 13% of her questions. Only 4% of her questions were reasoning about mathematical meaning and relationships and these occurred in the last seven minutes of the period. As a result, student mathematical understanding of the learning target was incomplete. Had she had
regained the seven minutes lost added to the discussion at the end, the mathematical residue might have been higher. Ashley agreed in her reflection following the lesson: “Work on closure of lessons. Add a journal entry to this lesson. Too many nonmath questions.”

**Coaching/Professional development.** The peer observers offered advice on this lesson; common themes were to conduct small-group talking at the table groups before attempting a larger group discussion. Advice included: “Perhaps a table share/pair share before opening the discussion up to the whole group; this gives the quieter ones more self-confidence before sharing an answer in front of the larger group. Try A/D agree/disagree cards to get 100% participation in your formative checks. Try Think-Pair-Share for the 1-2 big essential questions of the lesson to engage all the students in thinking and discussing it and getting higher quality answers because they have had to process, discuss and combine ideas to get a more complete understanding.” Other advice was more content-focused: “Connect the table to the graph sooner...what patterns do they notice in the table? Generalizing to a rule in words if not equation...can guide them to fit the data with a function with calculator if they can't yet do this curve-fitting independently.” Ashley was getting thus getting feedback from five teachers solely through digital means. I had observed and recorded this video in person, but the other observers could give feedback without having to get substitute coverage. Ashley could self-assess following her lesson by viewing the video when normally she would have had to reflect on her practice from memory. The study teachers ranked self-assessment highest in terms of leading to change in classroom practice. The digital nature of this
model allowed teachers to self-assess and review change over time along with receiving more formative feedback than in a regular evaluation observation from an administrator.

**Ashley’s Lesson 3: Patterns in quadratics.** In Lesson 3, Ashley selected a different type of task. The first two tasks were real-world scenarios that students were investigating by collecting and interpreting data. In contrast, in Lesson 3, Ashley was having students investigate the intra-disciplinary connections between representations of quadratic functions. The essential question was: What is the relationship between the table, graph and equation in vertex form of a quadratic function? Students worked in small groups of three or four to form a hypothesis and then explore the relationship between patterns in the table and the resulting graph given an equation. Groups sketched their graphs on chart paper and confirmed with the calculator. Ashley asked students to analyze the parameters of an unfamiliar equation to predict the resulting effect on the graph. The higher cognitive demand came from this inquiry approach to understanding the connections between the table, graph and equation in vertex form of a given quadratic function. Students wrote a journal entry on the connections they had discovered. This lesson offered an example of where connections do not always have to be real-world in order to be rich.

**Engagement in the learning process.** Observers rated Ashley with a mean of 2.6 for participation due to her focus on one table during the bulk of the classroom discussion. This was down from 3.0. While all student groups were included in the questioning in terms of formative checks, the probing questions, where she pressed students to explain their reasoning beyond their first replies, were focused on one table.
group who tended to volunteer first. One limitation of the video observation analysis process was that the participation or engagement rubric was only applied to the whole-group classroom discussion following the task. Given that constraint, 25-50% of students participated in the large-group summary discussion. However, I physically observed the entire classroom period, noting that greater than 75% participated in the small group discussions of the same essential question. The variance in observer ratings for this particular lesson on participation showed that they, too, struggled with whether to include the small group discussions. Since more students participated in the discussion in a small group, Ashley could have used other structures to connect those conversations to the larger summary discussion.

**Classroom discourse patterns.** Accountable Talk measures either dropped slightly or remained unchanged from Lesson 2. Teacher moves of linking, connecting and pressing students to explain and justify were down slightly from 3.3 to 3.0. Student moves of linking remained unchanged with a mean of 2.8, and the quality of explanation of thinking in student responses dropped slightly from 3.3 to 3. These results showed alignment between the quality of Ashley’s talk moves and the resulting student responses. This aligned with results from the larger study where teacher moves are correlated with resulting student moves. In the larger study, student moves were rated slightly lower than the quality of teacher moves. In this lesson, teacher and student moves were at about the same level.

**Questioning.** One of Ashley’s stated goals in her lesson reflection was to increase the rigor of her questioning. In this lesson, Ashley shifted from asking a majority of
nonmathematical questions to a more procedural approach with 55% procedural or factual questions. She increased her probing questions to 24% while questions about mathematical meaning and relationships increased slightly to 10%. The remaining 10% were questions that Ashley used to generate classroom discussion. A shift from the first two lessons was that she did not ask any nonmathematical or other-mathematical questions. Ashley was more focused on the mathematics for the entire period with fewer nonmathematical discussions. However, she replaced them with procedural and recall-type questions. While the procedural questions were more numerous in the distribution, Ashley did increase the percentage of mathematical meaning and probing questions. The complex computational aspects of the lesson generated the large number of procedural questions. Ashley’s questions did get at the big idea of the connections between the multiple representations of the function. She pressed students to explain the meaning of the coefficients and variables in the equation by connecting them to the table and particularly to the graph. Veronica noted the varied levels of rigor:

> Ashley leveled her questions to begin at a lower cognitive level and raised them higher as the discussion went on. She asked students if they agreed or disagreed with each other and with her and then to justify their feelings. I thought that was great as it requires the students to be able to perform the math and then justify it—very high up on Bloom's Taxonomy.

The quality of student explanations aligned with the rigor of Ashley’s questioning (3.0 to 3.2), mirroring the larger study results where the rigor of teacher questioning
positively correlated with student responses ($r = .59$) and the quality of student discussion ($r = .51$).

*Coaching/Professional development.* Ashley was quick to note that she would have improved this lesson by incorporating, “More real-world examples for quadratics.” The entire lesson’s essential understanding was outside of a context. Instead of being real-world, the connections were between the function representations. Not all connections needed to be real-world as intra-disciplinary connections required rich cognitive complexity and aligned to the content target of transformational graphing. Ashley had been trying to better align her course with the expectations of Algebra II, so incorporated a balance between hands-on, real-world data collection activities and more theoretical procedures with connections in her unit plans. In this quadratics lesson, the connections were very rich, but they were intra-disciplinary rather than real-world. Observers may not have recognized the value of the intra-disciplinary connections. I would contend that this is why Ashley’s observer ratings fell slightly or remained unchanged as compared to the Mammoth lab that was an obvious hands-on simulation of a real-world scenario. Had she done a ball-bounce activity with a motion sensor to collect distance vs. time data to fit with the same vertex-form of a quadratic function, I would have predicted higher ratings. She could have used that context as a springboard to go after the meaning of the parameters in a modeling approach by curve fitting, or creating the equation from the data results from the table and graph to fit the parabola. The large number of procedural and recall questions would have been reduced by use of the calculator’s lists, graphing and trace functions.
At this point in the study, Ashley contacted me, asking for a half-day of sub
coverage so that she could review her first three videos in a sequence and compare IQA
rubrics and survey reflections from all of her observers and compare to her own self-
analysis. She explained to me that, “I want to compare in one sitting, how Veronica’s
feedback compared to yours, my own and the others (peer observers).” On that day. I
stopped by and observed her in the school library immersed in this process of examining
her own practice from video, rubrics and survey reflections. I asked her what she was
thinking about. She said she saw why her third quadratics lesson was not at a higher level
of rigor and explained her reasoning, noting that she led students too much and that she
wanted to increase participation and students’ ability to explain and justify. She said she
really felt having a nonmath teacher perspective was a strength in the feedback.
“Veronica offered some really good strategies that I would not have thought of.”

**Lesson 4: A transformational approach to graphing.** Following this deep
reflection of her video lesson data, Ashley responded by planning and implementing her
strongest lesson in terms of Academic Rigor and Accountable Talk. In Lesson 4, students
were immersed in a transformational approach to graphing. Their task was to predict the
graph following the transformation of the parameters of a quadratic or exponential
equation. Since this was a similar task to the third lesson, I was curious to compare the
implementation trajectories. Ashley’s goals remained the same: increase participation,
 improve accountable talk moves specifically of student linking and teacher pressing
students to explain with evidence, and increase rigor of both student discussion and
teacher questioning. Quantitative results showed that this experienced teacher was rated
at a 4 level on a 0-4 scale in 7 out of 10 dimensions in the two domains of Accountable Talk and Academic Rigor.

**Engagement in the learning process.** More than 75% of students participated in the discussion, which was the equivalent to 4.0 on the scale. Participation showed an increase from 2.6 in Lesson 3 and was evident in the video footage of student explanations from each table of posters of their predicted graphs based on understanding of the role of each parameter of the equation. This differed from the teacher-led didactic discussion in the previous lesson where one table group dominated. Based on evidence in the video and quantitative results, Ashley used several key instructional strategies to increase equity of engagement. These came from the feedback in the survey reflections from the observers in addition to the IQA rubrics. Strategies began with think, pair and share, followed by sketching group graphs and recording visual sketches and explanations on large chart paper to compare predictions. Different groups did generate different graphs, which triggered novelty and cognitive dissonance. The cognitive level increased as students compared and contrasted different graphs. Furthermore, students were engaged in justifying by reasoning about their graphs and listening to other points of view to decide if their graph represented the equation. Observer reflections noted the changes:

Participation 13/14! Strong connections made by students between changes in equation and resulting graph; strong connections between representations. Hit big idea with meaning of coefficients in the equations. Connections between function types.
**Classroom discourse patterns.** Similar to Will’s fourth lesson, there was a buzz in the room of students talking in groups about math instead of the silence of the teacher talking to one student in a didactic model. All teacher talk moves were at a 4 level and student linking had improved to a 3.6. Quality of the student discussion was also a 4, because students consistently used evidence to support their claims. The quality of student discussion was set up by the rigor and directions of the task that asked them to create a new graph they had never seen from the changes in the first equation to the second and justify with evidence why it was correct. To do this, students had to explain how each change in the equation parameters affected the resulting graph sketch on their poster. They also compared how quadratic and exponential graphs shifted when the coefficients in their equations changed. Not only were students connecting between representations of table, graph and equation, but they were comparing shifts across function types, from quadratic to exponential.

**Strategies.** Ashley incorporated some of the strategies recommended by the observers in her first three lessons. Rather than posing the task discussion question to the large group, she gave them individual think time first followed by a table discussion in a small group where each table explained their thinking on chart paper. She urged groups to “have the discussion before you write it down.” She also had students use color to compare different graphs. All student predictions of graphs were made visible with group chart paper; students drew 4 graph predictions on the Smart Board with different colors for group predications in room. This set students up nicely for group discussion of next day to generalize how equation parameters affect graph.
**Linking.** This student group for the first time linked student ideas; some students linked by related their ideas to those of another student, which was new behavior for this group.

**Questioning.** Quantitative data showed Ashley’s rigor of questioning at a 4 on the cognitive level scale. Her driving questions by type were open rather than closed as were most of her discussion questions. Comparing the two exponential equations and predicting not just their graphs but how the second graph will shift offers deeper cognitive demand, which really gets at the essential understanding of the meaning of the equation parameters. Ashley asked lower level and closed questions in some teacher-student interactions to scaffold understanding. She continued up to higher level questions rather than stopping after the recall, comprehension, and application questions, or by resorting to “telling” to form the higher level connections. Even in the launch, Ashley began with a think-pair-share of “What is a parent function and why is it important?” This focused on the attributes of the reference point of the parent function as a starting point before connecting to transformations of the equation and resulting graphs.

A peer observer who is a special education teacher additionally noted that Ashley scaffolded her questioning to bring individual students to higher levels of complexity.

Ashley did a great job of scaffolding her questions to meet each student’s abilities and personal challenges. I know most of the students in the video and, therefore, know which students have IEPs. Ashley did a great job asking lower level questions first, and then taking the students with special needs higher as she asked more questions.
This description of scaffolding the cognitive level and complexity of questions was similar to Chin’s “cognitive ladder” that he described to bridge student learning. In this model, teachers begin with lower order recall and application questions and move to higher order questions asking students to justify, explain and generalize to new settings (Chin, 2006). Over the course of the four lessons, Ashley had changed her task implementation in terms of her own talk moves and rigor of questioning, likewise, students in her cotaught class also had improved in their talk moves and quality of discussion.

**Professional development.** As with all lessons, Ashley received feedback on her task implementation with suggestions for next steps. Some survey feedback to Ashley mentioned improving formative checking to collect individual evidence of student understanding of the daily learning target rather than just the questioning and group poster evidence. Monitoring individual understanding based on oral questioning remained problematic.

In addition to the group answers of the questions, collect individual responses for the last question slide comparing the exponential equations in an exit slip so that you know how each student is thinking to plan for next class discussion, already know(ing) the student misconceptions and understandings.

**Face-to-face collaboration.** Ashley attended the final face-to-face AFDA professional development day in June as the final professional learning event of the video lesson study. Teachers sat at a rectangular arrangement of tables facing one another and in an informal panel interview setting. They exchanged feedback on lessons and
discussed the model in response to questions from the Instructional Technology Supervisor and me. This administrator had helped with the logistics of the project and was interested in the potential for future division PD. As described in Will’s case study, Ashley discussed the merits of the video lesson study model and offered feedback to improve the model. She completed the final survey on performance tasks, viewed the coaching video of Barbara, and wrote her post-PD coaching reflection. Ashley was actively involved and completed all tasks. In the discussion, she emphasized three key components for her growth: peer feedback, self-analysis against set criteria with the IQA instrument and receiving feedback from a nonmath peer. “The most useful part of this to me was having time to analyze what I was doing in the classroom. I also like the fact that other teachers, both in my curriculum and outside of it, could give me suggestions on improvements as well as tell me where I did well and what I could improve on.” Ashley was consistent across three data sources in describing the benefits to her from the study.

**Summary of Ashley.** As part of her post-PD survey question asking what had changed in her use of performance tasks, Ashley replied, “I use the performance tasks to help the students to have to explain both in writing and verbally what they know, how they know what they know and why they know what they know and to explain how this knowledge will help.” She reported that she used rich tasks about once per week for a variety of purposes. “I use discussions in order for my students to understand that others need to know what they are thinking. Class discussions are to teach students how to communicate with each other, how to make them justify their beliefs and to help them clarify their ideas.”
Classroom discourse. Ashley reported that over 75% of her students typically participated in classroom discussions, which aligns with the data from her final lesson. She described how wait time had changed in her classroom discussions: “I am more willing to wait for the students to give their opinions, whether positive or negative, whether right or wrong. If the student gives a wrong answer, I believe that I need to push them toward the correct answer but they need to provide the vehicle to get there. I used to give a small amount of wait time and then I would just answer my own questions but now I am good waiting.”

Sarah’s Experience as a Special Education Teacher

Sarah came into the study with experience in using inquiry-based instruction and the modeling approach used in AFDA from coteaching for two years with Ashley. She also had taken part in several large PD projects: the AFDA collaboration at the division level and a NASA summer institute integrating math and science. She attended as many workshops as she could and was an active communicator in emailing me ideas and activities that she had researched. She had come to math only two years prior to the study from a special education English position. Before that, she had held a variety of special education placements over her 14 years of experience.

At the beginning of this study, Sarah asked to be videoed in her self-contained classroom with one Math Applications class for the four lesson cycles rather than in her coteaching role in the AFDA classroom with Ashley. She was eager to study her own teaching practice in the self-contained setting. Sarah felt that her role in the cotaught classroom was limited to one model of “one teach and one assist” with the general
education teacher teaching. Sarah had 12 observation reports from her four video lessons and could be studied in Ashley’s four videos as well in the coteaching role. All 12 of her observations of her self-contained classroom rated the tasks she selected as higher-level; eight were at a 3 level and four rated at a 4. Thus, Sarah’s improvement goals came from the task implementation dimensions.

**Beliefs about rich tasks.** At the beginning of the study, Sarah defined a mathematical task as, “A problem that needs to be solved using mathematic concepts. It may be procedural, or allow for more options in the solution.” Sarah described her role in rich tasks as, “Encouraging students to dig deeper and find different ways to a solution, or to make connections to prior knowledge. There may not be a ‘right’ answer.” She said she used tasks about once per week for a variety of purposes to “encourage deeper thinking, to check student understanding, to generate ideas.” She used tasks throughout instruction to, “Introduce new concepts/activate prior knowledge, apply a concept that has been taught, formative assessment, culminating unit activity, differentiation with remediation or extension, and VSEP collection.” Sarah added her use for data collection for a state alternative assessment portfolio for special education students.

**Goals.** In summary, Sarah already used tasks in her math classes and wanted to use the video lesson analysis study to improve her practice in several areas: “More student involvement from reluctant learners” and “better assessment of the success of the tasks.” She added that her students, “Show written work but not an explanation of why.” This video lesson study model allowed each teacher learner to set personal improvement goals and receive individualized feedback on those targets against set criteria from the
IQA instrument by multiple observers over time. In these ways, this model differentiated the professional learning and allowed for sustained practice over time to implement the desired behaviors, with embedded formative feedback along the way.

**Sarah’s baseline task.** Sarah’s first lesson’s task objective was to interpret and form inferences from a variety of graphs. This objective aligned with the eighth grade mathematics state standards as part of the Mathematics Applications I curricula. This course was part of a three-year sequence for a Modified Standard Diploma available to special education students. This diploma type was in the process of being phased out by the state as the diploma requirements for all students had been strengthened.

The driving question for the lesson was, “What could this (unlabeled) graph represent in real-life? Sarah projected different graphs on the electronic white board with the labels removed, asking students to reason using the scaled axes and data to form inferences about what the graph could mean. Then students digitally erased one at a time the labels of the axes, titles and other information, discussing how the axes labels and variable units could help them answer the question. Students then formed inferences from the graph after reading it. Graphs connected to many real-world topics including fuel efficiency, phases of matter, airplane flight, money, SAT scores and shoe sizes. Observers rated this task at a 3, or Procedures with Connections, which is considered a higher-level task with lower task implementation rating at 2.5 out of 4. To improve the cognitive demand during task implementation, Sarah listed her goals for this lesson as, “Student participation in discussion, student linking (connecting/relating), and student discussion after task.”
**Engagement in the learning process.** Sarah wanted to pull in the reluctant learners into the discussion. She succeeded; four observers rated her at 3.75 out of 4 on participation. With only seven students in the class, Sarah was able to employ equitable questioning, equitable participation on the white board at the front of the class and pull in every disengaged student.

**Classroom discourse patterns.** Sarah used a didactic teacher-to-student questioning pattern. However, once the first volunteer or called-upon student answered, she restated the student’s response and probed that student if necessary before asking the other students what they thought. She would probe, restate, link and press the other students to build upon the first response. With equitable distribution of questioning, every student was the “first’ student at least three times during the discussion. The class seemed comfortable with speaking and explaining their reasoning except for one reluctant student. Observers rated student responses at 3.25, fairly high for a baseline. Sarah would cold-call upon the reluctant student and give enough wait time for a response. When the other students took turns coming to the board to erase the “clues” on the graph information, Sarah had that student tell her what to erase on her turn since she was reluctant to go to the board. Sarah was rated at 3 and 3.25 for linking and connecting student responses and pressing for explanations. The growth opportunity for the teacher was to facilitate the transfer the initiation of the linking and connecting to the students. Students were rated at a 2 level for linking, so there was no real attempt for them to speak to each other to build upon ideas or agree or disagree. The entire conversation was held
between the teacher and individual students even though all but one participated readily in that process.

**Questioning.** Sarah asked a total of 64 questions in a 24-minute discussion of the interpreting graphs task. Observers rated teacher questioning for this lesson at a 3 out of 4 on rigor. She showed a balance across all question types with 17% probing, 13% mathematical meaning and 25% generating discussion. Only 13% were procedural/factual and other-mathematical with 20% nonmathematical. Having about one-third of questions nonmathematical or other-mathematical may seem high, but these questions served the purpose of connecting the graphs to a real-world context. She also was activating prerequisite knowledge needed to understand the graphs. Her strategy for interpreting the meaning of the graphs was to understand the context of flight, fuel efficiency of cars, temperature, phase changes of water and the way shoe sizes are measured to apply to reading the graph. Even her procedural or factual questions were connected to the physical context. An example in the phase change graph was, “What is significant about 0 degrees Celsius and melting?” This statement linked to her next question that asked students to apply this meaning to the slope of the segment on the line graph labeled melting. In these ways, Sarah used sense making and reasonableness in the situation frequently in her questioning which resulted in connection to other mathematical or nonmathematical concepts.

Sarah’s questions were balanced between open questions like, “What could the graph be a story about?” and closed questions such as, “Temperature in degrees Fahrenheit is on which axis?” In some individual conversations, at times, she scaffolded
recall, comprehension and application questions first before moving to analysis and evaluative questions. However, for the most part, Sarah took an inquiry stance by asking students to analyze and synthesize the given information on the graph to evaluate the evidence to predict the scenario. Because key information was hidden, students had to ask questions and understand the value of the missing axes labels to make sense of the scale and data in the graph. Students had to use strategic thinking as described in Webb’s Depth of Knowledge to predict the meaning of the graph when pieces were missing. However, the students largely were operating at lower thinking levels of recall, comprehension and application as reflected by the 2.25 out of 4 rating for rigor of student discussion. One peer observer expressed low expectations for the group, citing their disability status. Fortunately, Sarah disagreed and added a new goal of rigor of teacher questioning to her next lesson cycle.

*Coaching/Professional development.* Sarah reached her participation and student discussion instructional goals for the lesson but not her goal for student linking where classmates would begin to connect and relate ideas to each other. In her reflection, she said she felt she told the students the big math ideas and did all of the linking and connecting, while students were participating, they did not get to big ideas of the mathematics in the lesson themselves. She also felt there were instances of missed press on her part. To continue to work on these goals, Sarah engaged in five professional learning activities surrounding the first lesson cycle: meet postobservation in face-to-face meetings with me in my coaching role, self-assess with the IQA rubrics and the reflection survey, receive survey feedback from peer observers and me, observe a peer teacher’s
lesson video and give feedback with the same process, and select research-based strategies from these sources or the division site to try in this class over the next month until the next study lesson.

**Coaching conversations.** Since I observed each of Sarah’s video lessons during the last block of the day, I was able to stay to debrief following each of the four lessons. We met for about 30 to 45 minutes to discuss what went well, what did not, and to plan strategies to pull in all of the students and to increase the rigor. These immediate face-to-face coaching conversations were considered valuable by Sarah as noted in her postsurvey reflection.

**Sarah’s Lesson 2: Angle relationships.** Sarah kept the original participation, student linking, and rigor of student discussion goals from Lesson 1 and added one: rigor of teacher questioning. In the first lesson, Sarah had rated herself at a 2 on questioning while her observers had given her higher, resulting in a mean score of 3. In this lesson, students were learning angle relationships: vertical, adjacent, complementary and supplementary. The standard asked students to recognize the relationships and apply to find missing angle measures. Rigor dimensions were rated by observers with the following mean scores: Task Potential (3.5), Task Implementation (3), Student Discussion (3), Teacher Questioning (3.5). Mathematical Residue was low at 1.5.

**Engagement in the learning process.** All of the students participated in the task discussions within this classroom’s small, safe environment. The climate was nurturing and questions were personalized and scaffolded where needed. Under the domain of Accountable Talk, student participation increased to 100%; Sarah was rated a 4 on the
dimension of Participation; students felt comfortable explaining their thinking. Students all participated in the open-ended warm-up problem on the white board, the cut-and-paste activity on matching angle relationships before the proof task that was the focus of the lesson discussion. Students were pressed to stick with the open-ended proof task at the end of class. These problems asked students to justify and prove; most had difficulty with this type of thinking but persisted.

**Classroom discourse patterns.** Ratings on Sarah’s teacher talk moves were high at 3.5 for Teacher Linking and Connecting and 3 for Teacher Press. Student talk moves remained lower at 2.5 for Student Responses and 2 for Student Linking. This trend was evident across the study participants; teacher Accountable Talk moves were rated higher than those of students.

**Questioning.** Sarah asked fewer questions in this discussion as compared to Lesson 1 (37 vs. 64) and students talked more frequently with increased elaboration. The video evidence showed a shift to more student-initiated discourse with students occasionally connecting ideas from each other to create a proof of the missing angle measures in a diagram. The questioning was still balanced with 38% probing as the most frequent question category followed by Generating Discussion and Procedural/Factual-type questions, each at 24% of total questions. Mathematical Meaning questions remained about the same at 8% and Other Mathematical and Nonmathematical questions went way down to only one each in the task due to the lack of a real-world context as in Lesson 1 that required nonmathematical connections to understand the graphs. The low incidence of mathematical-meaning and relationship-type questions aligned with the low
Mathematical Residue having a variable rating of 1.5 out of 4 and with Sarah’s self-reflection about the lesson.

**Coaching/Professional development.** In her reflection, Sarah said that she was happy with the participation in the lesson and felt that to the one thing that needed to change most was deeper student understanding. This aligned with her rubric feedback which showed four lesson observations rating Task Implementation as lower level thinking even from higher level tasks. I debriefed with her in a 45-minute meeting following the lesson. Sarah then self-reflected by completing the rubrics and reflection survey and decided which goals to set next and which resources to investigate for strategies used to meet this goal. Her goal was to increase rigor by planning rigorous questions, which would require students to think about the mathematical meaning and relationships among mathematical ideas.

**Feedback.** Observer feedback urged Sarah to not save deeper cognitive thinking for the end of class after the skill has been learned.

Don't save this type of rigorous thinking with justifying/proof for the application problems at the end of class. The main body of the lesson was a very lower-level task with cutting and pasting the type of angle relationships; earlier in lesson, you could have had students work in pairs to solve these, share out their strategies and write their proof on chart paper or white boards to share out.

Student participation in the discussion following the task had increased to a 4, or above 75%. Sarah was rated highest on her questioning rigor and her talk moves of connecting and linking student ideas and pressing students to explain. Student talk moves
still lagged behind the teacher’s quality, but rigor in student discussion had increased from 2.25 to 3.

**Sarah’s Lesson 3: Balancing equations.** Sarah kept the original participation, student linking, and rigor of student discussion goals from Lesson 1 and rigor of teacher questioning from Lesson 2. She reflected that she wanted to keep working on the teacher moves of connecting ideas and relationships and teacher press to push students to provide evidence to support their thinking with conceptual explanations (adapted from Boston, 2012). In Lesson 3, Sarah improved in the four Accountable Talk domain dimensions of both teacher and student talk moves. In the Academic Rigor domain, Task Potential and Task Implementation remained the same at 3.5 and 3; Participation remained high at 3.67, dipping only slightly. Sarah’s highest dimension was Rigor of Questioning while the lowest remained Student Linking. Quantitative data thus showed that Sarah had improved in every goal she had set for this lesson cycle; even Student Linking improved from 2 to 2.75.

Sarah’s Lesson 3 took place in the same self-contained classroom as in the other videos. The learning target was for students to solve equations from pictorial representations. The lesson consisted of a series of four tasks on the electronic white board of pictorial representations of objects where an equation could be translated and solved to answer the questions. The first two showed groups of oranges on an expressions work mat. The third task showed crayons on a balance. The final task projected a balance with students and a bus asked both a procedural closed question and an open-ended one. Students had to write an equation to represent this balanced equation and explain how
they determined how many students were in the bus. Sarah used the bus task as a formative exit ticket.

**Engagement in the learning process.** Over 75% of students participated in the discussion. Ashley was pleased with the improvements in accountable talk but still hoped for a deeper understanding from her students, “I am pleased with how far this group has come, both as a collective (group) and individually. I would still like to be able to help them gain a deeper understanding of basic math concepts and how they connect to their daily world.”

**Classroom discourse patterns.** This lesson showed evidence of a real shift in classroom discourse patterns from the two previous lessons. In this lesson, Sarah used more wait time and students moved from brief answers to elaborate more, explaining their thinking, often in several sentences. Twice students went to the board to sketch a different solution method. Sarah consistently pressed each student to justify, calling on each student to explain each of the three tasks used in the lesson. She restated the key ideas and connected them between students, restating student responses 22 times. Of interest, was that Sarah did not “tell” students how to translate the pictorial representation to a symbolic one. A student offered the idea of $3 + x = 12$ to represent the oranges in the picture. Another student said, “I have never seen that before with the $x$.” Sarah acknowledged that this was new, and the students began to build understanding of that symbolic representation when the student went to the board to circle the oranges and show how many were needed to balance. At this point, some connected the pictorial representation to the symbolic algebraic model. At this juncture, Sarah followed with six
procedural questions to have students solve and write the equation symbolically. Students tried out these new ideas in the crayon task where another student modeled the solution on the whiteboard. Each student explained how his or her thinking connected to the thinking of the student at the board. Student linking was limited to the teacher setting up the interaction in her connecting and linking questions. Thus, students were not yet linking ideas unless prompted by the teacher. Prompted, they were easily able to compare approaches and explain common ideas.

**Questioning.** Sarah asked 45 questions which were about the same number and rate as in Lesson 2, but they were distributed differently across the six question types. In the discussion of the first two pictorial representations, all questions fell into probing, mathematical meaning, and generating discussion-type questions. All of the procedural/factual questions were asked from the symbolic representation discussion of how to write and solve the equation from the pictorial representation of the oranges or crayons. All probing, mathematical meaning and generating discussion questions were open as were half of even the procedural questions. In the last lesson, Sarah’s questions had been fairly balanced across types, but in this lesson, they shifted to open ended: 64% were probing, 18% generating discussion, 4% mathematical meaning and relationships and only 13% procedural/factual. Sarah was very focused in her questioning; she had no nonmathematical questions or even other-mathematical connections outside of the lesson.

**Coaching/Professional development.** After the lesson, Sarah and I debriefed face-to-face, and she was pleased with the rigor of her questioning and the increased elaboration by students in their responses. She noticed that students were now more
engaged in sharing different methods and strategies. The energy in the room had shifted to a more student-generated discussion with three out of five students linking on more than two occasions. Four observers completed the IQA instrument and gave survey feedback on Sarah’s third lesson. Feedback indicated that 7 out of 10 dimensions had increased; Sarah had met her targeted goal for this lesson to increase the rigor of questioning with a mean rating of 4. Sarah shared her feelings about the growth shown by her students during task implementation, “I am pleased with how far this group has come, both as a collective [group] and individually. I would still like to be able to help them gain a deeper understanding of basic math concepts and how they connect to their daily world.” She wrote that the one thing she would like to change was, “More involvement from students (in forming) more connections with tasks and applications.” She felt that now students were beginning to link and connect their ideas at the recall and comprehension levels; now she wanted to press them to connect to the mathematical relationships and deeper cognitive thinking levels of application and beyond.

Sarah shared in her written reflection how valuable she felt this video lesson analysis study was for her own learning. “I am involved in some fantastic professional development. :)” She had been looking for more opportunities to grow as a mathematics teacher. In this face-to-face meeting, she asked me to write a letter of recommendation for a summer STEM leadership institute offered by a university in our region.

**Sarah’s Lesson 4: Solving equations with pictorial representations.** Sarah used feedback data to set and adjust her goals for Lesson 4. Sarah kept two of the original goals: student participation in discussion and student linking and connecting ideas
between one another. She chose two additional goals for this final lesson: Teacher Press for students to support their reasoning and Task Potential, to increase the cognitive demand of the selected task. She selected student participation, because it had dipped slightly in Lesson 3 and student linking because, although improving, it was still the lowest-ranked Accountable Talk dimension at 2.75. Task Potential had increased initially to 3.5 in Lesson 2 but remained unchanged at 3.5 in Lesson 3. Sarah tried to select a task that required increased rigor. Finally, she thought she could increase the rigor of task implementation, which was also flat at 3, by increasing Teacher Press. After receiving a 4 in Questioning, Sarah did not set her rigor of questioning as a goal for this lesson (adapted from Boston, 2012). She sought to increase rigor by pushing students to explain and connect ideas as a learning community.

In Lesson 4, Sarah’s five students worked together on a learning target to translate write, and solve equations from an equation string of four scaffolded pictorial representations. This intended learning outcome was aligned with the next step in the learning progression for solving equations.

**Engagement in the learning process.** Observers rated Student Participation at 3.75, which was an increase from Lesson 3. A peer described participation as, “Students were attentive and engaged while using and viewing the SMART board.”

During the entire task discussion, a student leader recorded on the electronic white board the different equations generated by the class from the discussion during the equation string task, while Sarah served as facilitator of the task discussion. For the first time in the video lessons, students in the class came up with multiple methods or solution
paths for solving the question in the task. For example, students thought of three different ways to write an equation from a scaffolded series of pictorial representation of three tennis ball cans containing a total of 24 balls. The student recorder thought of $8+8+8 = 24$; discussion led to repeated addition as multiplication. Sarah then asked if students agreed with his thinking and why. She facilitated Student 1 in translating the picture into an equation with $3c = 24$. Then she asked for another way; a second student suggested division while another connected to inverse operations with “24 balls divided by 3 cans give you 8 balls in each can, or, $24/c = 8$.” A third student then added, “Or, 24 [balls], divided by 8 [balls per can] equals 3 c [cans].” Sarah then highlighted the connections between all of the symbolic equation representations of the picture scenario, highlighting inverse operations.

**Classroom discourse patterns.** New to Sarah’s discussions was having a student serve as a recorder to record all of the different equations and solutions that students generated. A peer teacher described this change, “The energy in the room was conducive to learning as it appeared that you had the attention of all of your students. I also think you did a nice job of leading the students to results as opposed to giving the answers.” She used Teacher Press to push students to explain how they visualized the translation from pictorial to symbolic, “Think of another way to get to that answer?”, or “Tell me how you got 8+8+8?” By pressing and asking each student if they agreed and why, she was scaffolding her students to form connections between different student representations as equations. At least three times, her students formed those connections without her setting it up first, earning her a 3 for the Student Linking variable which
required students to link and relate to another’s idea at least twice (Boston, 2012).

Sarah reflected on the mathematical connections between concepts made by students as they translated symbolic equations from the pictorial representations, “For this group the discussion went well. They were able to connect some concepts from previous lessons. Considering the time of year they did very well.”

The one thing she wanted to change after viewing and coding her own video was, “Even more involvement and discussion would be great. Helping students make more connections and generalize information is a goal also.” Sarah was thinking of the one student who did not participate in sharing her method, although this student did participate in group formative checks.

**Questioning.** Sarah asked 31 questions or about 2.6 questions per minute during the task discussion, which was about the same rate as in previous lessons. About three-fourths of the questions were Probing, Mathematical Meaning, or Generating Discussion-type questions while the other fourth were Procedural/ Factual combined with one Nonmathematical. She had to restate student responses less often as students were talking more and linking and connecting themselves in at least three cases. Observers rated her question rigor at 3.5, which, although strong, was down from a high of a 4. Accountable Talk moves of Teacher Connect and Teacher Press during discussion were also down slightly to 3.25 from 3.75. Student Participation and Linking was up to a 3, contributing to increasing total Task Implementation to a high of 3.25. Mathematical Residue of the lesson’s high ideas was rated at a study high of 3.25. Because Lesson 3 had such high ratings, only the historically lower-scoring dimensions improved in Lesson 4. However,
these improvements were in key dimensions like Student Participation and Linking along with the rigor components of Task Potential, Task Implementation and Mathematical Residue.

**Coaching/Professional development.** Sarah met with me immediately following the observation to debrief. She was pleased with the way her students generated different equations to represent the situation and began to link between them without her setting it up. She felt most of her students were explaining their thinking with more descriptive conceptual explanations rather than brief one-word answers. Sarah had improved her two Accountable Talk domain goals of improving Student Participation and Student Linking. She had also met her rigor goal of increasing Task Potential. Following a study trend, Sarah had improved on what she focused on in goal setting.

In terms of improvement, she worried about the one student who remained quiet and shared her current functioning in terms of assessment. Peer feedback was positive but also generated ideas for future growth in use of practical instructional strategies for increasing rigor through classroom discourse. One teacher offered a strategy to increase participation for the one student who was quiet; “Perhaps assigning a number to each student and using some sort of tally approach would help ensure more equal participation among the students.” Like Will in his fourth lesson, Sarah had turned over some of the direction of the discussion to being student-led in terms of the different equations that students saw in each of the equation stings. In doing so, she relinquished the control of rotating each teacher-to-student question in order in a didactic pattern as she had done in previous lessons. She did, however, pull in any student who was not participating in a
given discussion point with, “Give us your thoughts, ___.” or, “What did you get, ___?”

Only one student remained quiet.

Sarah was very positive about the overall video lesson study analysis model of personalized professional learning. In her face-to-face meeting with me and in her written reflection, this special education teacher shared that, “Getting the direct feedback from the observers. It is helpful to get outside feedback. It was also enlightening to see what was going on in my classroom in reality, not what I thought was going on.”

**Summary of Sarah’s experience.** In Lesson 1, a peer first described Sarah’s students’ ability to participate in meaningful classroom discourse as very low. “I believe the teacher tries to get the students to make connections but with this specific group of students, I am not sure it is possible to make the connections necessary to earn the higher level of understanding. The teacher is trying to increase the rigor but with a self-contained group of students it is a difficult task. The students in this video did the best to their ability.” Peer descriptions changed by the final lesson video. One peer described the discourse after Lesson 4 as, “Special education students attempted to make connections in the classroom. They were engaged in the learning activity and focused on connecting the taught concepts to prior knowledge.” To achieve this change in student behavior, Sarah set goals over the lesson study to increase the participation by her students in classroom discourse and the rigor of task implementation in her classroom. Her strategy was to set intermediate goals from the 10 IQA dimensions. One trend noted was that Sarah improved on what she focused on as goals in a particular lesson, with Teacher Questioning Rigor positively associated with each of the student Accountable Talk and
Academic Rigor dimensions. A second trend was that task potential was high in all 12 observations as defined by a 3 or 4 score out of a 0-4 scale, however implementation means were lower than the potential in every case. Four out of 12 observation surveys rated lessons as lower-level in implementation. A third trend was that Sarah showed improvement in 7 out of 10 dimensions from Lesson 1 to Lesson 4, with the greatest gains in the task implementation and the student dimensions. The lowest means in the beginning of the study were in Task Implementation and student dimensions of Student Linking, Student Discussion, and Mathematical Residue following the task. All of these improved to at least a 3 level on a 0-4 scale. Sarah’s Task Potential and teacher moves of Teacher Connecting, Teacher Press and Rigor of Questioning were already higher level, or at least a 3 on the 0-4 scale.

Sarah’s students changed over the school year in the way they solved and discussed tasks together. At the beginning of the study, Sarah posed a task and asked students to solve it individually. Then she held a didactic teacher-to-student interactive discussion by questioning individuals and restating what they had said. She was equitable in her questioning and scaffolded questions for students to be able to access. Sarah did most of the work in the discussion while students gave brief answers at the recall or skills/concepts levels. Sarah asked one-third other-mathematical and nonmathematical type questions to draw students into the context of the problem. Students answered as individuals, never linking or connecting to peer ideas. Participation was high but rigor of implementation was at a lower level of cognitive demand. Sarah “led” students in her questioning to understand the learning target. She modeled linking by restating and
connecting student ideas, but students were not yet able to link to each other. In summary, the baseline lesson showed a higher thinking level task at a 3-level being reduced to a lower level task during the task trajectory of implementation by the lower-level thinking occurring during the classroom discourse.

In contrast, by the fourth lesson cycle near the end of the year, students were still participating at a high level, above 75%, but their behaviors had changed during classroom discourse, resulting in higher-level thinking during task implementation. The gap between Task Potential and Task Implementation had been reduced, with both falling within the higher-level thinking score ranges. Student leaders emerged during discussions to help Sarah to facilitate student discussions by recording different methods and class solutions on the electronic whiteboard. Students were now sharing different solution methods and connecting between them. Three of the five students were consistently showing strategic thinking beyond skills and concepts. Four out of five were consistently explaining their solutions with conceptual meaning. The energy in the room had shifted from all teacher-directed to more student-centered. Four of the students were actively engaged in solving the task and while they did not solve it in a group, they solved it together in the classroom discussion by putting together the pieces from individual work. Sarah had initially said that due to lack of comfort in social interactions, students were seated apart and did not work in groups. However, by the fourth lesson, students did come together as a group to solve tasks through the classroom discussion with Sarah facilitating. Sarah had improved the rigor of her questioning and consistently pressed students to explain and justify their thinking in every lesson as well as modeled linking. I
began to see a shift in the students as a classroom community in the third lesson with
great improvement in Student Responses, Student Linking and Student Discussion. These
gains were positively associated with Sarah’s Question Rigor, which was a 4 out of a 0-4
scale in that lesson.

In summary, Sarah changed her behaviors during task implementation to first get
participation up and secondly to increase the rigor of thinking during task
implementation. Over the course of the school year, student behaviors of explaining and
justifying mathematical thinking, connecting ideas and engaging in the classroom
discussion after a task followed. Next steps will be to work on becoming more consistent
with the level of cognitive demand. The third lesson had a higher quality student
discussion with students explaining and justifying at a higher level, while the fourth had
more active participation, student linking and mathematical residue as a result of the task.
Sarah wanted her lessons to consistently exhibit all of these dimensions at a higher level.

**Interpretive Summary**

Chapter 5 used three case studies to examine the teacher experience of a video
lesson analysis model used to study teaching and learning practices during mathematical
task implementation. Of interest were the Accountable Talk moves during classroom
discourse following the task and the Academic Rigor experienced by students during the
task trajectory. Ten dimensions were used to code these teacher and student experiences
during task implementation in order to answer two overarching research questions:
impact on task selection and implementation experienced by students and efficacy of the
professional learning model.
1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?

2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?

Data from all study sources were categorized and classified with five emerging themes: task potential vs. task implementation, student engagement, patterns in classroom discourse, rigor of teacher questioning, and coaching during professional learning. I was able to give an accurate description of the evolution of practices during the coaching experience by triangulating and layering the data from all sources: performance task pre- and postsurveys, pre- and postcoaching reflections of a common video lesson, mathematical tasks, classroom lesson videos, IQA instrument rubrics, and lesson survey reflections submitted by teachers, peer teachers and coaches. Each lesson was observed and coded by multiple observers, including the teacher. In these ways, I was able to layer my analysis by aligning multiple data sources to increase validity.

By design, this model offered a personalized coaching experience to teachers based on individual goal setting, yet a common analysis based on comparison against set criteria in the IQA rubrics. These common criteria gave peer teachers a standard way to give meaningful and specific formative feedback to each other. Coaching conversations were sustained during one summer and ensuing school year via digital means. Teachers received nonevaluative feedback via Google Forms on video links in YouTube in a blended model. Face-to-face collaboration was built-in during the initial two days of
summer training, during the year and through a substitute day for collaboration at the end of the year.

The case studies of new teacher Will, experienced teacher Ashley, and special-education teacher Sarah painted a picture by filling in the stories behind the themes that had emerged from the quantitative results. Five previously mentioned themes emerged from the analysis of six data sources. Ten variables were coded: Task Potential, Task Implementation, Participation, Teacher Press, Teacher Connecting to Concepts and Linking, Student Linking, Student Responses, Student Discussion, Questioning, and Mathematical Residue of big ideas.

Integration of Quantitative and Qualitative results across nine teachers fell into five themes: Rigor in Task Potential vs. Task Implementation, Student Engagement in Learning, Patterns in Classroom Discourse, Rigor of Teacher Questioning, and Coaching: Teacher Engagement in Professional Learning. The following summary describes commonalities and differences in Will, Ashley and Sarah’s experiences through these themes as they studied their teaching. I have used Maxwell’s concept of descriptive validity to give an accurate description by capturing the teachers’ experiences, actions, and perspectives (Maxwell, 2002, 2013). I have increased interpretive validity by layering the factual data from the rubrics and surveys with the actual words of the teachers and peer observers, taken from written reflections throughout the four lesson cycles during the school year. Together, I have layered teacher perspectives and their interpretation of the experience with quantitative ratings to increase credibility and
trustworthiness of findings. Readers can then decide how these results might transfer to their own need for personalized, digital professional development.

**Salient Themes**

Through analysis of the emerging themes, I had two overarching goals in my interpretation: to understand the learning process within each theme and compare and contrast that experience across teachers representative of three groups: new, experienced, and special education.

**Theme 1: Rigor: Task potential vs. task implementation.** Teachers came into the study already selecting tasks requiring higher-level thinking. The challenge was that cognitive demand during Task implementation was lower than that of the task itself across all baseline data sources in the study. Initially, Accountable Talk dimensions of Teacher Connecting and Linking and Teacher Press were lower than the rigor components. The lowest areas were the student dimensions of Student Responses in terms of explanation and justification of thinking along with Student Linking. However, these classroom discourse dimensions increased during the school year over the four lesson cycles for Will, Ashley, and Sarah. Will set his initial goals to increase rigor of Task Implementation along with Student Participation. Ashley and Sarah also chose Participation but chose the classroom discourse dimensions of Accountable Talk moves such as Teacher Connecting/Linking and Teacher Press as first steps. Improved teacher moves encouraged students to explain and made student thinking visible. As students elaborated by explaining and justifying, classroom discourse became richer, eventually associated with improved rigor of task implementation.
Experienced teacher Ashley was consistent in keeping tasks at the same cognitive demand during implementation. While she had slightly lower ratings for the task potential of her selected tasks with 87% of them offering higher-level thinking. However, she was more consistent than Will and Sarah in keeping the cognitive demand high during the implementation trajectory. One of her higher-level tasks became lower during implementation. Ashley’s 80% higher-level implementation rating was the highest of the three in the case study.

The new teacher, Will, was in the middle; 93% of the tasks he selected had high task potential while 71% of the corresponding tasks remained higher-level thinking of students during implementation. Sarah, the special educator, had the highest ratings on the cognitive level of task selection for her self-contained students. 100% of her task ratings were higher-level, however, she had the lowest ratings on keeping the tasks high with 67% rated high level in implementation.

**Theme 2: Student engagement.** All teachers, including Will. Ashley and Sarah, set increasing student participation in classroom discussion as the first goal. As Will and Sarah increased participation by changing the discourse pattern from solely a didactic teacher questioning pattern. As teachers offered more opportunities for students to collaborate and discuss tasks, student cognitive engagement in thinking improved. Both quantity and quality of student responses increased. Students began to elaborate and in the final lesson began to link and connect between student ideas. In both Will and Sarah’s classrooms, this change came late in the fourth lessons where teachers gave up control of the questioning path and encouraged multiple solution paths and methods. Ashley had
from the start expected students to explain thinking and offer different methods or solution paths. However, her students were largely not able to explain their reasoning, link ideas, or give high-quality responses. Her personal challenge was to increase the rigor in student responses and discussion. All of these dimensions showed growth with Student Linking, Responses and rigor of Student Discussion the lowest-scoring dimensions throughout the study. Will and Sarah had improved these challenging student dimensions all in the higher levels of at least a 3 out of a 0-4 scale. Ashley was above 3.5 on all dimensions by the end of the study. This highlights how an identical dimension was personalized based on teacher and student needs.

**Theme 3: Patterns in classroom discourse.** Verbal patterns of classroom talk shape the learning environment by influencing the type of talk that students are engaged in (Gee, 2001, as cited in Smart & Marshall, 2013). Only the experienced teacher, Ashley, entered the study already using teacher moves of pressing students to explain reasoning, restating, and encouraging more than one idea, method or solution. Her students worked on tasks in groups while the students in the new and special education teachers’ classrooms did tasks individually in desks arranged in rows. Will and Sarah used a didactic teacher-to-student discourse pattern. In Will’s case, a correct answer stopped the conversation, and the teacher was the judge. In Sarah’s case, she was getting very brief answers from her students who focused just on the answer to the immediate question, with no linking or connecting either to each other or to the big mathematics learning target. Will did mark key mathematical ideas and Sarah often restated student answers and asked more than one student to respond. However, students did not discuss
or link with each other or connect mathematical ideas. Students gave brief, computational, or incomplete answers.

All three teachers improved in using accountable talk moves over the school year. One trend across the nine teachers was that in the first two lessons, mathematical discourse talk moves were lower than the Academic Rigor dimensions of Task Potential, Task Implementation and Questioning. Data also showed students lagged behind their teachers in learning to link and connect to each other and engage in higher-level thinking during classroom mathematical discussions. Students first needed teachers to be able to model these behaviors and the have practice over time between video lesson markers. Both improved by the end of the study but were still lower than indicators that could be preplanned like selecting rich tasks and writing key essential questions to use in class discussions. The “on-the-fly” accountable talk moves to orchestrate the discussion were mastered later, requiring more coaching and embedded practice. Also, teachers did not have much output from students to use to press, restate, link, connect and mark as important in the first two lessons. Once their questioning became more higher-level and students participated more due to various strategies including group tasks, then more student thinking became visible for them to act on. Then Will and Sarah could begin to practice talk moves. By the fourth lesson, all three classrooms had the increased energy of the “buzz” created by groups of students discussing a math task.

The benefit of the personalized coaching was that while all teachers were working on improving classroom discourse, each teacher was able to target personal goals based on timely feedback from their videos. Sarah described her progress in her
postperformance task survey: “Pulling more connections from the students. Getting students to build on each other's answers. Allowing more time for student response.” In the final meeting, Ashley wrote about her classroom discourse, “I use discussions in order for my students to understand that others need to know what they are thinking. Class discussions are to teach students how to communicate with each other, how to make them justify their beliefs and to help them clarify their ideas” (Postperformance task survey). Results had shown that Will had moved from discourse focused on the right answer to more interest in evidence of student reasoning. Will focused on the environment for discourse, “The most important thing I have done to change classroom discussions is firstly creating an environment that allows students to feel comfortable answering questions and taking a chance on giving their answer even if they feel they might be wrong. Allowing students to disagree and correct one another and creating an expectation that all students are expected to participate in the class will also help me generate better more in depth classroom discussions.” The dimension most highly correlated with discourse moves of teacher press, teacher and student linking and connecting, and quality of student responses to justify with evidence was rigor of teacher questioning.

**Theme 4: Rigor of teacher questioning.** Teacher questioning in inquiry settings differs from that in knowledge-level discussions focused on the correct answer (Smart & Marshall, 2012). Teacher questions in an inquiry task discussion are more flexible, adjusting questions based on student responses in order to encourage higher-order thinking (Chin, 2007). Students must explain and justify both “how?” and “why?”
beyond the correct answer (Ma, 1999). Teachers ask open rather than closed questions and are focused on finding out the perspective of student thinking. Teachers are neutral rather than evaluative in an inquiry classroom setting (Roth, 1996, as cited in Smart & Marshall, 2013). Lustick noted in 2010 the value of teachers writing key essential questions as part of planning before a lesson, “...the value of utilizing specific focus questions during inquiry-based instruction; these questions are developed by the teacher with the intent of supporting student understanding as they participate in the process of scientific inquiry.”

By planning their questions and studying the feedback on the six types of questions coded in the IQA instrument (Boaler & Humphreys, 2005), study teachers were all able to quickly improve the rigor of their questioning. The distribution of teacher questioning moved from more procedural/factual, other-mathematical and nonmathematical to a balance that included generating discussion, probing, and mathematical meaning about relationships-type questions.

Student responses to these more open, rigorous questions lagged behind in development. Teachers then had to develop more complex talk moves to press students to elaborate in order to give more conceptual responses. Teachers modeled what a discussion looked like in an inquiry environment by restating student responses and linking them to connect different student ideas. All three stopped short in getting the discussion to a deep understanding of the big mathematical ideas of the lesson. Will lamented, “the cognitive demand stopped short of doing mathematics even though my students are participating and engaged in classroom discourse.” Getting to the
mathematical residue would take more time to develop. Sarah agreed describing her special education students after the first lesson, “while students were participating, they did not get to the big ideas of the mathematics in the lesson themselves. I told students the big ideas.”

**Theme 5: Coaching/Professional learning model.** Results suggested two big ideas from the video coaching model: teachers improved on goals they focused on and teachers liked the self-analysis and peer feedback from the IQA rubrics. Goal setting worked; mean scores across 114 observations showed that all dimensions improved from the first to the fourth lessons. Even more interesting was that teachers reached individual dimension goals set for a particular lesson even if other dimensions remained the same or did not improve. Results showed that teachers and students improved in all 10 dimensions over the course of one school year by focusing on specific goals with clear descriptions of target criteria.

Will thought the most useful components were the peer observations, self-analysis of his teaching videos, talking about his lesson with other teachers/researcher and using video references about instruction. He elaborated on this value in his final postsurvey reflection, “Allowing myself to receive feedback from other teachers on what they thought of my lesson. Peer teachers helping me revise/modify and suggest what can be done to improve lessons is a valuable asset in helping me become a better educator in the classroom. I also liked being able to watch my own videos and self-assess how I got better from the first lesson to the final lesson.” Sarah and Ashley independently chose the same three activities as most useful, “Peer observations, self-analysis of your teaching
videos, The Instructional Quality Assessment Rubric” (Postperformance Task Survey). Sarah referenced our postobservation face-to-face conversations and peer surveys, “Getting the direct feedback from the observers. It is helpful to get outside feedback. It was also enlightening to see what was going on in my classroom in reality, not what I thought was going on.” Ashley wrote, “The most useful part of this to me, was having time to analyze what I was doing in the classroom. I also like the fact that other teachers, both in my curriculum and outside of it could give me suggestions on improvements as well as tell me where I did well and what I could approve on” (Postperformance Task Survey).

In the final meeting, teachers were asked to suggest improvements to the prototype that I would use when refining the division video lesson study model. Will suggested more face-to-face time, “Possibly sitting down with other teachers and having a face-to-face conversation about what else they liked and disliked may be beneficial. Often times creating a discussion about what was good and what could be modified will help create better instruction in the classroom” (Postperformance Task Survey). Another teacher thought having teacher observers from a different content area would offer a helpful perspective. Two peer teachers noted that the rubric for Participation during class discussion following the task did not include a place to record the classroom discourse occurring during group work during the exploration of the tasks. Having a way to monitor and code small group interactions over video would be an improvement in future models. The difficulty is that they occur simultaneously in groups and the camera can only focus on one group at a time or whole group where the individual conversations are
hard to hear. I was able to hear the small group conversations well of the groups videoed up close to get a qualitative sense of the conversations, but they were not coded.

Sarah and Ashley gave advice to future teachers studying how to implement tasks with a higher level of student thinking. Sarah counseled, “Don't be afraid. This should be a learning opportunity to see what you do well and what things could be done to help students be even more successful.” Ashley also talked about attitude toward the study, “Be honest with yourself and with your peers. If you aren't honest, this will be a waste of your time. Choose peers who are as vested in improving their [own] teaching as you are. Don't be afraid to go back and make changes on a task that you were not happy with.”

This summary summarized highlights of key themes that emerged from the comparative case studies of three representational teachers who participated in a year-long professional learning model to improve implementation of mathematical performance tasks. Through improving classroom discourse, teachers improved their capacity to keep the cognitive demand high during task implementation. Chapter 6 discusses and integrates findings from all data sources in both Chapters 4 and 5 by research question. In addition, I share implications of this study on the use of a personalized, digital lesson analysis model to coach teachers on improving instructional quality. Finally, I recommend ideas for future research on coaching teachers to select and implement rich tasks.
Chapter Six

The purpose of this study was to explore how to increase capacity of secondary mathematics teachers to deliver performance tasks with higher cognitive demand by increasing student reasoning through classroom discourse. Of interest was how to personalize teacher professional development given the typical constraints of a public school division in terms of resources. This study used one professional development model featuring formative assessment of 10 dimensions of academic rigor and classroom discourse talk moves occurring during mathematical task implementation using digital media to differentiate teacher professional development. Formative assessment of teacher and student coaching included self-assessment and peer feedback. This closing chapter provides a summary of the findings and conclusions of the study by research question, a discussion of implications for teacher practices and professional learning models, limitations of the study, and implications for future research.

Summary

In order to study the process and impact of personalized coaching using a blended model of video observations with digital feedback and face-to-face conversations, two overarching questions were studied:

1. In what ways did mathematical task selection and implementation change during a video lesson analysis professional learning model?
a. For what purposes did high school mathematics teachers select tasks for students?

b. What did high school mathematics teachers want to change about task implementation in their classrooms?

c. How did the potential of the task relate to the cognitive demand experienced by students during implementation?

d. In what ways did academic rigor change over the course of four lesson cycles for teachers and students?

e. In what ways did classroom discourse change over the course of four lesson cycles for teachers and students?

2. In what ways did a video lesson analysis professional learning model coach secondary teachers on mathematics performance task implementation?

a. How did teachers describe the change process in their own task implementation?

b. Which coaching activities did teachers perceive as leading to growth?

c. How did the learning process compare across three teacher groups: New, Special Education, and Experienced?

d. What suggestions did teachers offer to improve coaching activities in future models?

I used a mixed-methods research design with survey and qualitative case-study methods, interweaving five different data sources. The quantitative data source components consisted of selected questions on the pre- and postsurvey questions, the
ranking of the selected tasks in the lessons for cognitive demand, and the IQA observation instrument’s scaled responses across 10 dimensions. The qualitative data sources consisted of: (1) open-ended items on the pre- and postperformance task beliefs surveys; (2) the pre- and poststudy written video analysis of a common online math lesson taught by an Algebra I teacher; (3) classroom video footage; (4) teacher and observer written lesson reflections collected in Google Forms.

My conceptual framework for this study was created from an examination of literature on the use of rich mathematical tasks in a classroom utilizing the Process Standards and Mathematical Practices outlined by NCTM and the CCSS (CCSSO, 2010 and NCTM, 2000). These student practices of problem solving, reasoning, communication, connections, representation, precision, and strategic use of tools are evident in the coaching feedback presented in the study. I also used the literature and methodology on using a design process to develop a model, which in this case is a professional learning model for teachers to improve capacity to implement higher-level mathematical performance tasks. A key challenge for teachers was keeping the cognitive demand experienced by students high during the task trajectory. Literature findings on both teacher and student moves during classroom discourse (Accountable Talk) were used to select the IQA instrument that coded classroom discourse and cognitive demand (Academic Rigor) for these key criteria in the lesson cycles. Finally, I used the current work of John Hattie and Dylan Wiliam to justify the use of nonevaluative, formative feedback throughout both student and teacher learning.
From these key literature areas, my focus argument developed that in order to implement tasks with high cognitive demand, teachers first needed to select higher-level thinking tasks and then deploy them in a classroom environment where rich classroom discourse allows student thinking to become visible, both orally and in writing. Students self-assess their current understanding based on peer and teacher feedback to further construct their own understanding of mathematical concepts. Teachers can study to improve their own practices in much the same way: make their practices visible through video observations, set goals from video evidence, self-assess against set criteria, and consider coaching feedback from peers and coaches. Teachers then use this formative feedback to try out new practices and strategies to improve, self-assess each lesson as to effect on students, use more peer coaching feedback, and again, refine practice in an iterative cycle of design. I situated this model within the current realities of a typical public school high school setting where professional development funds, instructional coaching staffing, substitutes for observations, and teacher time are all finite and much less than necessary for ease of studying to improve teaching and learning practices. The hope of these study findings is that by harnessing free digital tools which make learning available 24/7, along with the power of collaborative peer formative feedback, teachers can improve their practices within existing conditions.

**Discussion of Findings: Research Question 1**

One purpose of the study was to describe the ways in which mathematical task selection and implementation changed during a video lesson analysis professional learning model over one summer and following school year. To do this, the study
collected and analyzed data on five components: (1) task selection purposes; (2) goals for improving task implementation; (3) the relationship between the cognitive demand required of students in the selected task compared to during implementation; (4) evidence of changes in the Academic Rigor dimensions over four lesson cycles by students and teachers; (5) evidence of changes in classroom discourse by students and teachers over the same four lesson cycles. In summary, teachers and students improved across all 10 dimensions of the Academic Rigor and Accountable Talk in this video lesson study. The teacher talk move of Teacher Connecting and Linking was the variable that improved the most along with all student dimensions, having started lower than other teacher behaviors. In the baseline task, Task Potential was the only dimension out of ten rated at a higher level. The other nine dimensions were rated at a lower level on the IQA scale. In the final lesson cycle, 9 out of 10 dimensions were rated as higher-level, meaning at least a 3 on the IQA rubrics. Only Student Linking was still rated below 3 on the IQA rubric, at 2.79.

**Research Question 1.a.: Task selection purposes.** Findings from teacher surveys and individual lesson goals indicated that before the video lesson analysis study, the most commonly cited use for performance tasks was for differentiation, followed equally by activating prior knowledge, application and formative assessment. Following the experience, all teachers said they used tasks to apply concepts, followed closely by formative assessment of student thinking. The majority of teachers came into the study using tasks instructionally during lessons as opposed to solely for summative assessment purposes. This remained so following the study. Not only were teachers comfortable
selecting tasks as the vehicle through which mathematics was learned, the quantitative results from the IQA instrument scoring of the Academic Potential variable showed the tasks to be typically higher-level. Teachers entered Lesson 1 already tending to choose high-level tasks at about the 3 level out of a 0-4 scale and improved slightly by Lesson 4. Experienced teachers as a group tended to select higher-level tasks than new or special education teachers. These results made sense given the division context of developing the AFDA course as an inquiry-based, task and problem-based learning course. Self-reported results and quantitative IQA rubric ratings indicated that it wasn’t the selection, but rather the task implementation that needed to be the target of improvement efforts.

**Research Question 1.b.: Teacher goal setting.** Survey results and individual lesson goal setting both indicated that teachers most wanted to improve task implementation, notably student and teacher Accountable Talk behaviors and rigor of the student discussion and of teacher questions. The top six lesson goals in order of frequency were: (1) student participation in classroom discussion; (2) rigor, or cognitive demand of student discussion; (3) teacher linking or connecting student ideas to the big mathematical ideas; (4) teacher pressing students to provide evidence to support their thinking with conceptual explanations; (5) student-to-student linking to others’ ideas; and (6) rigor of teacher questioning (IQA Instrument, Boston, 2012).

**Participation.** In all 38 lesson cycles, teachers selected increasing student participation in classroom discussions as a goal. Student participation across all classrooms tended to be in the 25-50% range in Lesson 1, moving to 50-75% by Lesson 4. No significant differences existed between teacher groups from pre to post in findings.
The means were identical, except that new teachers scored .01 higher than the experienced and special education teachers. While not significantly different, student participation was the only dimension in which the experienced teachers did not score higher than the less experienced or special education teachers. Engagement would be an interesting variable to study in future research: In what ways do new teachers may tend to do as well or even better than experienced teachers in engaging students in participation during classroom mathematical discussions?

**Cognitive demand of student discussion.** The second most frequent teacher goal was to improve the cognitive level of student discussion following tasks. This successfully targeted goal was aligned with need. Student Discussion means in Lesson 1 were rated as lower-level and improved to a higher-level 3 rating by Lesson 4. The IQA rubric descriptors coded oral and written responses as to whether students explained their solution using evidence of why their strategy or method worked to solve the task. Furthermore, students needed to connect to the big mathematical ideas and/or represent their work using different representations and show the connections between the ideas in the models. Teachers were looking for a thorough discussion across all strategies to show connections to the underlying mathematical ideas (Boston, 2012). Quantitative data from the IQA instrument indicated that Student Discussion was the second lowest dimension in Lesson 1. Students typically would describe their solution but not give mathematical evidence or explanation of why it worked; they also tended to represent their thinking in only one way and did not connect between and among representations. By Lesson 4, students across the study had begun to explain not only how their solutions worked
procedurally but also why their strategies worked mathematically. However, these discussions tended to be incomplete or not at a deep level. Students had begun to show more than one representation of their concept, but tended not to be able to connect the same ideas found in the different forms. Experienced teachers again were better able to facilitate and elicit student discussion, followed by new teachers. Special education students struggled most with this dimension.

The other most-cited teacher goals from the lesson cycles, in order of frequency, were teacher linking and connecting, teacher pressing for explanations, student linking, and rigor of teacher questioning. It is interesting that teachers targeted most frequently the variables that showed the most growth in the study. For example, Teacher Connecting and Linking was the most-improved dimension variable followed closely by Student Linking Participation, and Student Discussion. All study teachers improved on their focused goal areas. Discussion of findings from each of these dimensions will be discussed under Research Questions 1.d. and 1.e. regarding changes in academic rigor and classroom discourse.

**Research Question 1.c.: Task potential vs. task implementation.** To summarize results described in Chapters 4 and 5, the AFDA study teachers entered the study already selecting high-level tasks that had “the potential to engage in rigorous thinking about challenging content” (Boston, 2012), but tended to implement them at a lower level than written. Analysis of goal setting indicated that teachers were aware of this discrepancy from baseline videos and set this as the primary goal to improve. While falling in to the “higher-level” rating category, most selected tasks still had room for
growth, being at a Level Three out of four. Stein and Smith described this level as Procedures with Connections in their 1999 Task Analysis Guide. The IQA rubric for Task Potential describes these tasks as, “having the potential to engage students in complex thinking or in creating meaning for mathematical concepts, procedures, and/or relationships” (Boston, 2012). During this study, all teachers improved slightly in task selection, which was already rated as higher-level in the baseline, and more in task implementation where there was greater room for growth. However, these the accountable talk moves, especially by students, were the highest growth areas in the study, underscoring the need for improved classrooms discourse. Once the student reasoning became visible during classroom discourse, through these talk moves, then the cognitive demand of the discussions and mathematical residue of the big ideas became higher level.

In no case did tasks requiring lower-level thinking become higher-level during implementation. Study findings suggested that certain task implementation choices were associated with lower cognitive demand during implementation. The strongest association was with students not seeing the connections between different solution methods, representations or strategies for the same problem. Typically, students did not form connections between multiple representations of the same function, for example, between verbal, table, graph and equations unless explicitly prompted by the teacher. Where the teacher failed to model or prompt, those connections usually were not made. Secondly, students often were able to describe patterns in a table or behavior of a graph but not generalize specific cases to a rule without scaffolding. Video evidence showed
that where generalizations were made, student role models were selected in classroom discussions following a task by teachers to model the thought process in generalizing to an algebraic expression, equation or formula. Finally, in classrooms where there were many instances of Teacher Press, the quality of student responses was higher as students more frequently used evidence from their work and from mathematics to justify their solutions.

Generally, for the cognitive demand to be higher level in task implementation, the task needed to first offer the potential for students to think at higher cognitive levels. Then the teacher needed to take an inquiry stance by asking students to analyze and synthesize given information and to use reasoning to evaluate evidence to solve a problem. Lower-level rigor ratings were associated with teachers “telling” the generalization, rule or procedure before the students had constructed it. The focus in lower-level classroom discussions was a didactic teacher-to student recitation of the correct solutions.

Research Question 1.d.: Changes in academic rigor. The domain of Academic Rigor in this study measured five dimensions: Task Potential, Task Implementation, Student Discussion, Teacher Questioning, and Mathematical Residue. Three major findings stood out: (1) AFDA teachers came into the study already selecting higher-level tasks, but they implemented them at a lower cognitive demand. Experienced teachers offered higher academic rigor; (2) The rigor of teacher questioning mattered the most in improvement of rigor domain dimensions surrounding classroom discourse; (3) Teachers improved in their ability to select and implement rich tasks before their students had the
communication and reasoning skills to explain and justify with mathematical explanations and link or connect ideas of their teacher and peers. Once students learned these talk moves, then the cognitive level of student discussions improved so much that the means were rated in the higher-level category by Lesson 4.

While all five dimensions of the Academic Rigor improved over the yearlong study across all teachers, the variable measuring the rigor of teacher questioning was found to be most linked to all other teacher moves from analysis of all data sources.

**Task selection vs. implementation by teacher group.** While these two dimensions were discussed in detail in the previous section, there were other interesting associations. Experienced teachers as a group both selected and implemented tasks at a higher level than new or special education teachers who were, as a group, about the same. Experienced teachers were stronger in all five Academic Rigor domains. They had less variance in task selection and more consistently kept tasks at a high level. New and special education teachers showed more inconsistency in implementation, and more often than experienced teachers were not able to keep the cognitive demand up. They tended to talk more, “tell” more, ask procedural questions, and over-scaffold. They seemed, as a group, less willing to give up control or decision-making to students. However, they were about the same and even slightly higher in getting their students to participate equitably in the classroom discussion of the mathematics. A key area of improvement for experienced teachers was equitable participation as they tended to ask the same group of willing volunteer students most of the questions. By the end of the study, the majority of students were participating in the discussion when teachers moved to a more facilitating
role from total direct teaching. Many used total participation strategies in group discussions, turning over data collection and presentation of learning to students, and offering more choices in how students solved problems along with encouragement to use different representations.

**Rigor of teacher questioning.** Teacher questioning was the variable that mattered the most across all teacher dimensions of the Academic Rigor and Accountable Talk classroom discourse domains. This dimension was central to the data collection process, in that observers coded every teacher question in each lesson observation and categorized by type. It was part of the professional learning process that classroom teachers coded their own questions and those of peers. Teachers became very familiar with models of high and low cognitive level questioning, types of questions, and patterns in questioning. As teachers improved in their questioning over the lesson cycles, several trends stood out. Teachers changed the proportions of the types of questions that they asked. By the end of the study, teachers tended to ask a higher percentage of probing and generating discussion-type questions and fewer non-mathematical or procedural-factual questions. Increases in choice of solution paths, methods, and ways to organize and show thinking were associated with increased cognitive demand in task implementation. Question rigor was associated most significantly with increases in other Academic Rigor dimensions and all of the Accountable Talk dimensions. Finally, teachers improved more quickly than their students, but the students showed the greatest change by the end of the study.
**Student choice.** Over the course of the study, the highest scoring teachers gave up control and moved to more of a facilitator’s role. They offered students more choice by encouraging multiple solution paths using different methods to get to the solution. Some gave up asking only questions from a worksheet and having total control of the questioning path. When students initiated questions and began to connect and link to not only the teacher’s ideas but to that of their peers, the cognitive demand increased. My analyses suggested that the reason choice was so important was that the cognitive demand went up when students had to compare and contrast their method to different methods used in the same classroom. In explaining and justifying their method choices and thinking through another point of view, students were moving to higher levels of strategic and extended thinking. With choice, solutions did not all follow one set procedure so were more complex and less algorithmic. With different representations, students were often comparing data in tables to data in graphs and generalizing to algebraic representations of the same relationships between quantities. However, my data suggested that few students were able to connect the big mathematical ideas across representations without heavy scaffolding by the teachers. Such connections would be a next step in goal setting for teachers in future lesson analyses and for the division in planning professional development.

**Teachers changed more quickly than students.** Classroom discourse is a reciprocal and responsive process, so teachers first had to model the new talk moves and then students learned by interacting and replicating these moves with practice over many lessons. My observation through coaching conversations and teacher reflections was that
the IQA observation tool made it easy for teachers to change, because the criteria for the target norms and behaviors in the rubric descriptions were so detailed and clear. Crosson et al. (2006) argued that the IQA tool was especially useful for teachers to use to analyze their own and colleagues’ practice because the precise gradations of instructional quality in the rubrics make it as transparent as possible to offer a clear target. Through self-assessment and reflection, goal-setting, and coding peers on their questioning in their observations, teachers quickly became familiar with the different question types and the criteria for success. By knowing the target, they could replicate the behaviors. In contrast, students did not have set criteria for the classroom discourse behaviors as part of the study. Certainly, some teachers incorporated clear expectations for classroom discussion behaviors, but few had students practice accountable talk moves such as revoicing, repeating, adding on, linking, reasoning, revising, explaining, justifying, connecting, advancing and probing against set criteria. Without an established classroom learning community of practice with normed behaviors for safe classroom discussions, it made sense that students initially lagged behind their teachers. By Lesson 4, ratings with the same classes working in groups all year, showed the student dimensions of linking, participation, discussion, and responses as having the greatest gains along with the teacher move of connecting and linking which is the most correlated to all of the student variables.

By the end of the study, Student Discussion showed the highest improvement in means of the Academic Rigor variables. I argue that the initial gap would have been much less, especially in new and special education teachers’ classes, had the classroom
learning community norms been set up in the beginning of the school year with established norms for student discussions and use of rich tasks. Thus, many students did not know of the expectation to explain and justify until they were actually in the discussion and the teacher asked them to explain, “why?” Baseline ratings were very low for student cognitive demand. The changes to a student-centered learning community triggered improvement in dimension ratings for both Accountable Talk and Academic Rigor.

**Discussion of Findings: Research Question 2**

The second study question examined the video lesson analysis professional learning model itself from the aspects of teacher coaching on task implementation, the perception of the coaching activities by teachers as leading to growth, a comparison of the learning process across teacher groups, and improvement suggestions to enhance the model. The digital links of four lesson cycles created a portfolio of practice that showed changes in instructional practice over time. Data from this research question was used in my division to adjust the video lesson study model for the following year.

**Research Question 2.a.: Coaching.** This model allowed each teacher learner to set personal improvement goals by lesson and receive individual feedback on these targets based on common criteria. The evolution of teacher practices during the coaching experience indicated that teacher PD was not identical even for teachers who set the same goal in terms of the dimensions. Differences in teacher readiness and student needs required personalized feedback. For example, all teachers set increasing Student Participation as a first goal in the lesson cycles. Once they got increased participation in
terms of engagement in the student discussion, student thinking became visible so that teachers could press students to explain and justify. Teachers could employ other talk moves such as revoicing, reasoning and linking once students began sharing their mathematical ideas in the discussions. All teachers improved quickly in terms of rigor of task implementation from their moves as they had internalized the IQA rubric criteria by Lesson 2 by using it on themselves and peer teachers. The most common coaching theme emerging in reflections was advising teachers to shift to the role of facilitator. Students followed suit so that by Lesson 4, the rigor of student thinking made visible during classroom discussions had increased to being “higher-level” on the IQA rubric across study participants. Thus, the goal setting and comparison of teacher and student behaviors against the IQA rubrics were common coaching components in this model.

However, teachers’ implementation was not identical even with the same goals. Teachers, like students, differentiated the process and the product based on their individual readiness and the needs of their students. Lourcks (2010) said that effective PD should not be identical for each teacher. This model differentiated the PD by using emerging formative assessment from the lesson videos to give individualized feedback following each lesson so that teachers could use it to set a goal for the next lesson. Teachers researched and implemented improvements with the same class before being videoed for the next lesson cycle. By having multiple observers, teachers got multiple viewpoints and ideas.

**Feedback form.** The observation survey design allowed not only for rubric quantitative data collection but qualitative constructed response items to share written
coaching suggestions. Teachers had choice in selecting other strategy sources to try in the next cycle. I had provided a PD resource bank on the mathematics course site, face-to-face coaching, digital discussions and the option of teacher research of strategies on their own. As part of coaching, I had assigned journal articles to all in the first cycle and video clips from the resource bank based on video lesson data to some participants. The IQA survey itself was the main catalyst used for peer feedback, self-assessment and goal setting.

Some feedback was more content-focused while other coaching suggestions focused on instructional strategies to increase equity of engagement. Some peers discussed more strategic use of tools such as graphing calculators and manipulatives. For example, use of graphing calculators to confirm predicted shifts in graphs from changes to equation parameters was a content-focused strategy offered for a transformational approach to graphing. Feedback was unique to each teacher, even though the IQA rubric was common because the survey utilized mixed methods using a rubric scale along with open coaching and reflection questions for both teachers and coaches. Teacher observers emailed rubric feedback automatically generated via email from the Google form to each teacher.

**Changes.** Common themes from task implementation changes included teachers moving to a facilitator role by giving up some control to create a more student-centered experience. In every classroom, the largest change was in students solving the problem task themselves and discussing solution strategies with each other rather than the I-We-You model of direct teaching. Over time, teachers became more accurate and aligned in
self-reporting as compared to the reality of the common IQA instrument criteria as reported by peer assessments. One common remaining area of needed improvement was in selecting and sequencing student work shared in the classroom discussion following the task in order to get to the big mathematical ideas of the lesson objective (Stein & Smith, 2011).

**Research Question 2.b.: Teacher evaluation of professional development activities.** Teachers were consistent across three data sources as to the benefits of the video lesson analysis PD model. Self-analysis against the IQA instrument followed by multiple peer observational feedback and talking about one’s lesson with others, digitally or face-to-face, were the top activities noted by teachers as leading to growth. This aligns with John Hattie’s 2013 meta-analysis of effect sizes of instructional practices that lists teacher self-analysis as number one. Teachers described the value of video as compared to direct observation by an observer as a benefit for several reasons: self-assessment was from direct observation in video rather than from memory; observers and the observed teacher could discuss an observation from a common reference; the video could be reviewed multiple times; and the video could be shared across schools with multiple observers, all commenting via digital media. Video gave teachers more varied sources of feedback than was possible with direct observation due to constraints of time, funding and available personnel.

**Goal setting with feedback.** Teachers decided what they wanted to learn and improve based on individual data gleaned from a collaborative review of a baseline video lesson. This learner choice permeated the PD model. Quantitative data indicated that
teachers improved on written, focused instructional goals. Teachers set an instructional goal for each lesson and received individual feedback on that goal across 10 dimensions from multiple nonevaluative observers. Teachers liked having the ability to choose peers from which to get feedback, often expanding past their team to others in the school or across the division. Teachers noted strength in having nonmathematics teacher feedback to offer a different perspective and new instructional strategies that could be adopted from other content areas. In fact, this was suggested as a mandatory feature in future iterations of this PD model. In all, learner choice, which is so effective with students, is equally as important to adult learners. Teacher participants valued choice, paired with personalized embedded formative feedback on their instructional practices, in a safe, collaborative environment.

**Research Question 2.c.: Comparison across teacher groups.** The quantitative survey, IQA ratings, and qualitative case study together explored how the learning process compared across three teacher groups: new, experienced, and special education. Quantitative data indicated that experienced teachers were more consistent in executing task implementation trajectories having a higher cognitive demand. All three groups selected higher-level tasks, typically *Procedures with Connections*, but there were differences across groups in implementation. New teachers tended to be more teacher-centered, utilizing direct teaching, with students seated in rows. Classroom discussions were most often recitations between the teacher and student volunteers in the beginning of the study. Concerns of new teachers about group work included fear of students copying answers, poor social skills, off-task talking, and losing control. During the study
new teachers shifted to designing more student-centered lessons, building a mathematics classroom community where students had begun to explain and justify their thinking. Of interest was the data point that showed engaging students in participation as the one area where new teachers scored slightly above experienced teachers. Although not significantly higher, it would be of interest to study why students participated at a higher rate in classroom discussions with new teachers.

Special education teachers tended to offer a small safe environment whether in a self-contained or coteaching classroom. Task implementation was more scaffolded with a series of related smaller tasks replacing one larger task for the lesson. The shift that occurred with the special education teachers in this study was to move from students working alone on problems to collaborative problem solving in groups during the exploration stage. Following the task, teachers moved from reading the answers to the problems while students passively checked solutions to a group summary discussion where students shared solutions and strategies. The shift from individuals to a classroom community led to improvement in IQA dimensions of Student Responses, Student Linking, and Student Discussion tied to Rigor of Teacher Questioning. The key was improvement in teacher questioning during classroom discourse whether in small groups or whole group. Teachers like Sarah became more balanced in the use of question types so that students could build in thinking levels from where they started to more frequently experience higher-levels of cognitive complexity (Chin, 2006). While participation and questioning improved, data showed that special education teachers tended not to get to deeper understanding of the overarching mathematical concept of the lesson. I would
argue that more coaching over time with all teachers would improve the mathematical residue of their lessons. All three groups of teachers improved first in the more easily learned skills such as increasing participation and planning questioning. Accountable talk moves improved later in the study following increased student participation as trust in classroom communities developed. The complex moves of formative questioning based on emerging student thinking and facilitating the summary classroom discussion after the task to get the big mathematical idea with students would require more time and coaching than four lesson cycles.

**Research Question 2.d.: Improvements to model.** Study teachers met in June following the final lesson cycle to complete reflections and discuss the experience. A postsurvey reflection item also asked for suggestions to improve the learning model. Very few changes were suggested in either venue. Two teachers suggested more face-to-face coaching following their lessons; both of these teachers were case study teachers who had every lesson observed in person so had received more face-to-face coaching than most of the teachers. One teacher recommended release time to review videos. Another suggestion was to offer the opportunity to use this self-improvement study for teacher SMART goals as part of the teacher evaluation program.

**Implications for Teacher Practices and Professional Learning Models**

Results showed that teachers and students improved in all 10 dimensions of Academic Rigor and Accountable Talk over the course of one school year by focusing on specific goals with clear descriptions of target criteria. Formative feedback to the teacher by peer teachers and a nonevaluative specialist gave teachers an immediate reading with
common criteria of the IQA rubrics. Self-reflection survey questions on achievement of lesson goals along with self-rating on the IQA rubrics were the most important activities leading to growth.

This study suggests that teachers can actively work to improve their instruction by first setting specific goals based on a realistic view of their current practices through video. Key was the ongoing sustained comparison of current practice against common set criteria. Formative, embedded, and nonevaluative feedback on current specific teaching practices helped teachers and students continue the learning rather than shut down after a snapshot “grade” of their performance. Like students, a choice of strategies and activities to achieve learning goals was supported by the data in this study. Teachers must continually self-assess, reflect and adjust based on multiple inputs of data from peers and knowledgeable others such as specialists, coaches or administrators. Key was the comparison of teachers’ own perceptions of practice to feedback from others. Study results indicated that teachers became more accurate in honest self-assessment with practice as they internalized the rubric, used it on peers, and understood common look-fors. Improvement is a complex process requiring time and hard work. Growth occurs when teachers confront practices that do not work, based on data, and struggle to improve them. Such growth from errors aligns with brain research as made popular by Carol Dweck (2006) and in the mathematics community by Jo Boaler (2013). In effect, as humans, teacher learners improve in the same ways as student learners through adjustments following mistakes.
The delivery of this differentiated coaching model through video and digital media was offered as a solution to the more costly use of physical observations and face-to-face coaching. Professional development models requiring substitutes and manpower not available to scale for every teacher frequently enough to invoke change over time are not viable. Moving to use of free Google Apps for Education (GAFE) and other applications for sharing video and social communication allows teachers and coaches to collaborate 24/7 across time and space for no cost. Cost and convenience are secondary even to the increased number of ideas and viewpoints available to teachers when the audience opens up past the administrator and department teachers at one school. In a rural-suburban division like mine, teachers in rural areas who are the only teacher of their particular course content can collaborate not just on observation feedback but lesson planning, assessments, and other professional learning efforts as well. Teachers have been isolated long enough, receiving almost no immediate feedback on change efforts. This model offers a design framework that can be morphed into virtual PD for any content or purpose.

Finally, the use of a research-based observational instrument of classroom look-fors by teachers studying their practice in collaboration with peers brought research from higher education solidly into the everyday classroom through a form of action research. The Google curriculum site offering professional development resources including articles, case studies, and videos also brought research within easy reach for use to study teachers. In these ways, this professional learning model expected teachers to actively use
research to improve teaching and learning. Ultimately, improving teaching is the pathway to improving student performance (Stigler & Hiebert, 2009).

**Limitations of the Study**

I have designed multiple measures of validity to increase authenticity, so that I can increase trustworthiness of the implications (Denzin & Lincoln, 2008). This is particularly important due to the pragmatic paradigm in which this study operates. I needed to analyze the data from the stance of potential bias and design flaws in order to question findings, because I have been using the findings of this study to adjust this professional development model in my own division. I have been transparent about limitations in order to mitigate them with additional study addendums and to inform recommendations for future studies. As previously mentioned, I used five validity checks to improve trustworthiness of findings. However, design limitations and implementation realities limited the transferability of this design past this setting. The duration of this study over one school year with a sample of nine teachers as opposed to a longitudinal study over several years with multiple cohorts of teachers limits the impact past a local setting. Secondly, my role as both researcher and observer-participant biases the observational data reported by me. I tried to mitigate this effect by having multiple peer observers code and give feedback on each lesson, including the teacher of the lesson. I also used a research-based instrument rather than creating my own. One limitation of the video observation analysis process was that the IQA participation rubric was only applied to engagement in the whole-group classroom discussion following the task, not extended also to the small group conversations occurring during the exploration of the task.
Finally, I believe that teachers may have taught to the IQA rubrics when being observed and videoed. A truer measure would have been to video every class for a given time and randomly select classes to code from all taught lessons. To mitigate this limitation, I layered five different data sources over four lesson cycles in the data collection protocol. For observations, Matsumura et al., in 2008 stated that the IQA instrument is reliable, with fidelity of implementation, in as little as two lesson cycles per teacher to yield a stable estimate of classroom practice.

**Implications for Research**

Several interesting data points and themes emerged from the mixed methods used that would be interesting to explore in future studies. The first would be to replicate this study with a larger cohort over several years, looking at longitudinal growth. Secondly, it would be useful to scale up this professional learning model to discover design and implementation flaws that would surface in a larger study. It would be interesting to study if a factor in this success was that the teacher cohort was kept small. Although teachers did not know the entire peer observer pool, they knew most of them. Part of the design was collaborating with chosen peers in addition to getting feedback from assigned observers. Comparing feedback from peer teachers on a teacher’s team or department as compared to blind observers would also help to design this model to give more reliable feedback to teachers.

New teachers were able to engage students in participating in classroom discussions slightly more than experienced or special education teachers even though they were not as strong in the other nine dimensions of task selection and
implementation. It would be interesting to study student responses to new teachers under a variety of teacher moves as compared to experienced teachers, so we can share the engagement strategies that work.

This study compares the experience of special education teachers in this process to new and experienced general education teachers. There would be many questions to investigate in this arena. In the videos of Sarah coteaching in an AFDA inclusion classroom, I saw different teacher behaviors than in her self-contained classroom. It would be helpful to the profession to study how coteaching teams can collaborate to use video to study their teamwork and delivery of instruction. Effective coteaching teams are very much sought-after in the quest to improve teaching and learning for students with disabilities.

Finally, the addition of the analysis of the assessment of the student work of the mathematical tasks videoed is a missing data point that would better inform the impact of the cognitive demand of the task implementation. Student thinking captured orally and in writing together gives the ultimate indicator of student success. This study coded students’ oral responses in terms of reasoning and justification but did not consider their written solutions. Melissa Boston has recently adjusted her IQA instrument for use in assessing student work from tasks as one measure of instructional quality (Boston, 2014). In our division, our students have recently begun using a Google application to record explanations of mathematical task solutions alongside their written solutions, using the app pen to highlight such features as the connections between the representations as they talk. Student work samples can be conveniently filed in online folders for use in student
editing, sharing of strategies, self-assessment and to monitor growth. Future studies could extend the use of video to students for the same formative growth process as is used for teachers.

**Closing Ideas**

By improving classroom discourse, teachers in this study increased their capacity to keep the cognitive demand high during task implementation. The importance of this shift in cognitive complexity was that it moved students closer to the essential understanding of the deeper meaning of the mathematics. Once students have experienced extended thinking about a concept, they can begin to use the mathematics to create. This shifts our students from being passive receivers of knowledge to producers who can create solutions to move our culture forward.

The success of the teachers in this cohort to improve their own teaching began by setting goals, trying new strategies suggested by coaching feedback, and comparing performance on video against research-based criteria over time. This study offers an inexpensive model that all divisions can use to design their own professional learning models in any grade level or content. The hope of this study is that like students, all teachers can improve their instruction by first setting clear targets based on honest self-assessment of practice in a safe, collaborative learning community. Key is using formative feedback to reflect and adjust behaviors with enough time to sustain needed growth. Implementing rich mathematical tasks that engage students in high levels of cognitive demand through classroom discourse requires a complex skill set of teachers. It is imperative that we equip our teachers with these skills to be able to offer our students
the school and life experiences that will prepare them to be creative problem-solvers, communicators and producers in their futures.
Appendix A

Performance Task Surveys

<table>
<thead>
<tr>
<th>Question #</th>
<th>Preperformance Task Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How would you define a math task?</td>
</tr>
<tr>
<td>2</td>
<td>What makes a task &quot;rich&quot;?</td>
</tr>
<tr>
<td>3</td>
<td>How do you use math performance tasks in your classes?</td>
</tr>
<tr>
<td>4</td>
<td>About how often do you use a rich performance task in a unit?</td>
</tr>
<tr>
<td>5</td>
<td>For what purposes do you use class discussions?</td>
</tr>
<tr>
<td>6</td>
<td>What are the top two types of questions that you find yourself asking?</td>
</tr>
<tr>
<td>7</td>
<td>Describe typical student participation in your teacher-led discussions.</td>
</tr>
<tr>
<td>8</td>
<td>How often do your students support their ideas or solutions with evidence or reasoning?</td>
</tr>
<tr>
<td>9</td>
<td>To what extent do your students typically show their work?</td>
</tr>
<tr>
<td>10</td>
<td>What is the number one thing you would like to change in the way you use performance tasks?</td>
</tr>
<tr>
<td>11</td>
<td>What is one thing that you would like to change about your classroom discussions?</td>
</tr>
<tr>
<td>12</td>
<td>Pseudonym</td>
</tr>
<tr>
<td>13</td>
<td>ID Number</td>
</tr>
<tr>
<td>Question #</td>
<td>Postperformance Task Survey</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>1</td>
<td>How would you define a math task?</td>
</tr>
<tr>
<td>2</td>
<td>What makes a task &quot;rich&quot;?</td>
</tr>
<tr>
<td>3</td>
<td>How do you use math performance tasks in your classes?</td>
</tr>
<tr>
<td>4</td>
<td>About how often do you use a rich performance task in a unit?</td>
</tr>
<tr>
<td>5</td>
<td>For what purposes do you use class discussions?</td>
</tr>
<tr>
<td>6</td>
<td>What are the top two types of questions that you find yourself asking?</td>
</tr>
<tr>
<td>7</td>
<td>Describe typical student participation in your teacher-led discussions.</td>
</tr>
<tr>
<td>8</td>
<td>How often do your students support their ideas or solutions with evidence or reasoning?</td>
</tr>
<tr>
<td>9</td>
<td>To what extent do your students typically show their work?</td>
</tr>
<tr>
<td>10</td>
<td>How have you have changed in the way you use performance tasks?</td>
</tr>
<tr>
<td>11</td>
<td>What is one thing that you have changed about your classroom discussions?</td>
</tr>
<tr>
<td>12</td>
<td>Which PD activities/tools did you feel were useful to you to improve something in your teaching?</td>
</tr>
<tr>
<td>13</td>
<td>What was the most useful part of this PD?</td>
</tr>
<tr>
<td>14</td>
<td>Please suggest improvements for other teachers studying their classroom tasks / discussions</td>
</tr>
<tr>
<td>15</td>
<td>Please list your pseudonym.</td>
</tr>
<tr>
<td>16</td>
<td>ID Number</td>
</tr>
</tbody>
</table>
Appendix B

IRB Exemption Letter and Informed Consent Form

Office of Research Integrity and Assurance
Research Hall, 4400 University Drive, MS 605, Fairfax, Virginia 22030
Phone: 703-993-5445, Fax: 703-993-9580

DATE: August 2, 2013
TO: Margret Hjalmarson, PhD
FROM: George Mason University IRB

Project Title: [488182-1] Differentiated Coaching of High School Mathematics Teachers: The Instructional Trajectory of Performance Tasks

SUBMISSION TYPE: New Project
ACTION: DETERMINATION OF EXEMPT STATUS
DECISION DATE: August 2, 2013
REVIEW CATEGORY: Exemption category # 1

Thank you for your submission of New Project materials for this project. The Office of Research Integrity & Assurance (ORIA) has determined this project is EXEMPT FROM IRB REVIEW according to federal regulations.

Please remember that all research must be conducted as described in the submitted materials.

Please note that any revision to previously approved materials must be submitted to the ORIA prior to initiation. Please use the appropriate revision forms for this procedure.

If you have any questions, please contact Bess Dieffenbach at 703-993-4121 or edieffen@gmu.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within George Mason University IRB’s records.
Differentiated Coaching of High School Mathematics Teachers: The Instructional Trajectory of Performance Tasks

INFORMED CONSENT FORM

RESEARCH PROCEDURES
This research is being conducted to create effective coaching models for differentiated professional development of secondary mathematics teachers. If you agree to participate, the data collected from your individualized lesson cycles will be analyzed for its usefulness for teacher professional development. Specifically, data from the pre and post surveys, observation rubrics from videos and use of professional development resources will be studied. The research will not take any additional time, since all activities will be completed as part of the professional development whether or not participants decide to participate in the research.

RISKS
There are no foreseeable risks for participating in this research.

BENEFITS
There are no direct benefits to participation.

CONFIDENTIALITY
All personal data in this study will be confidential. For coded identifiable data, (1) your name will not be included on the surveys and other collected data, (2) a code will be placed on the survey and other collected data, (3) through the use of an identification key, the researcher will be able to link your survey to your identity; and (4) only the researcher will have access to the identification key. Videos will be stored on an external hard drive kept in a secure location. Teachers will be given flash drives containing their videos and those for their review. No videos will be shared on the web. Videos will be removed from the devices used for filming once they are uploaded to the hard drives. All data will be password-protected with external drives kept securely at GMU for three years. All data will be destroyed 3 years after the completion of the study.

PARTICIPATION
Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you decide not to participate or if you withdraw from the study, there is no penalty or loss of benefits to which you are otherwise entitled. There are no costs to you or any other party for participating in the research.

ALTERNATIVES TO PARTICIPATION
Teachers who decide not to participate in the research will still complete all of the professional development activities, but the researchers will not use their data as part of the research.
CONTACT
This research is being conducted Dr. Margret Hjalmarson, Professor of Mathematics Educational Leadership, at George Mason University. She may be reached at (571) 276-1603 for questions or to report a research-related problem. Deborah Crawford is conducting research supervised by Dr. Hjalmarson as part of a doctoral program. You may contact the George Mason University Office of Research Integrity & Assurance at 703-576-4121 if you have questions or comments regarding your rights as a participant in the research.

This research has been reviewed according to George Mason University procedures governing your participation in this research.

CONSENT
I have read this form and agree to participate in this study.

__________________________
Name

__________________________
Date of Signature

Version date: 7/26/13
# Appendix C

Video Observation Feedback Survey

<table>
<thead>
<tr>
<th>Question</th>
<th>Video Observation Feedback Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adapted with permission from Instructional Quality Assessment Instrument, Melissa Boston, 2012</td>
</tr>
<tr>
<td>1</td>
<td>Username</td>
</tr>
<tr>
<td>2</td>
<td>Please describe your role!</td>
</tr>
<tr>
<td></td>
<td>Teacher in video</td>
</tr>
<tr>
<td></td>
<td>Peer teacher in same school</td>
</tr>
<tr>
<td></td>
<td>Peer teacher in another school</td>
</tr>
<tr>
<td></td>
<td>Coach/Specialist/Supervisor</td>
</tr>
<tr>
<td></td>
<td>Online Peer</td>
</tr>
<tr>
<td></td>
<td>Other:</td>
</tr>
<tr>
<td>3</td>
<td>Math Topic of Lesson</td>
</tr>
<tr>
<td>4</td>
<td>Course</td>
</tr>
<tr>
<td>5</td>
<td>Tally of the Exploring Mathematical Meaning and Relationships Questions: Ex. “How would your expression work for any function?”</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1-5</td>
</tr>
<tr>
<td></td>
<td>6-10</td>
</tr>
<tr>
<td></td>
<td>11-15</td>
</tr>
<tr>
<td></td>
<td>Greater than 15</td>
</tr>
<tr>
<td>6</td>
<td>Teacher in Videos (Initials)</td>
</tr>
</tbody>
</table>
Tally of the Probing Questions ("How did you get that answer?")

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td></td>
</tr>
<tr>
<td>Greater than 15</td>
<td></td>
</tr>
</tbody>
</table>

For Teacher in Video: What are you working on now to improve student reasoning and communication of their thinking?
Please check all that apply.

- Student participation in discussion
- Teacher Linking-connecting student ideas and relationships between
- Student Linking-connecting/relating
- Press students to provide evidence of their thinking/conceptual explanations
- Potential of the task re cognitive demand (rigor)
- Task Implementation
- Student discussion after task
- Rigor of teacher questioning (Cognitive Demand)
- Mathematical Residue (big ideas stick)
- Other

Number of Students (if known)

Tally of Generating Discussion Questions: Ex. "Explain what John was saying."

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td></td>
</tr>
<tr>
<td>Greater than 15</td>
<td></td>
</tr>
</tbody>
</table>

Tally of Procedural/Factual Questions: Ex. "What is the coefficient?"

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td></td>
</tr>
<tr>
<td>Greater than 15</td>
<td></td>
</tr>
</tbody>
</table>
|   | Tally of Other Mathematical Questions: Ex. "How could you use this in the real world?"
|---|---
| 0 |
| 1-5 |
| 6-10 |
| 11-15 |
| Greater than 15 |

|   | Tally of Nonmathematical Questions Ex. "Where is your calculator?"
|---|---
| 0 |
| 1-5 |
| 6-10 |
| 11-15 |
| Greater than 15 |

<table>
<thead>
<tr>
<th></th>
<th>Consider whole-group discussion only. How effectively did the lesson-talk build Accountability to the Learning?</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Over 75% of students participated in the discussion.</td>
</tr>
<tr>
<td>3</td>
<td>50-75%</td>
</tr>
<tr>
<td>2</td>
<td>25-50%</td>
</tr>
<tr>
<td>3</td>
<td>Less than 25%</td>
</tr>
<tr>
<td>4</td>
<td>No discussion was held.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Does the teacher support students in connecting ideas and positions to build coherence in the discussion?</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Teacher consistently connects ideas or has students connect AND show how ideas relate to one another?</td>
</tr>
<tr>
<td>3</td>
<td>At least twice during lesson, teacher connects and relates...</td>
</tr>
<tr>
<td>2</td>
<td>At one or more points teacher links speakers' connections to each other, but does not show how ideas relate to one another OR only does this once (1 strong link).</td>
</tr>
<tr>
<td>1</td>
<td>No effort to link or revoice speakers' contributions.</td>
</tr>
<tr>
<td>0</td>
<td>No class discussion OR Class discussion not related to math</td>
</tr>
<tr>
<td></td>
<td>Were students PRESSED to support their contributions with evidence and/or reasoning?</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>16</td>
<td>4-Teacher almost always asks students to provide evidence and press for conceptual explanations.</td>
</tr>
<tr>
<td></td>
<td>3-Teacher sometimes presses for evidence and conceptual explanations, but there are instances of missed press.</td>
</tr>
<tr>
<td></td>
<td>2-Most of the press is for computational or procedural explanations or memorized knowledge.</td>
</tr>
<tr>
<td></td>
<td>1-No efforts to ask students for evidence for their contributions AND to ask for students to explain their thinking.</td>
</tr>
<tr>
<td></td>
<td>0-Class discussion was not mathematical OR no class discussion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Did students support their contributions with evidence and/or reasoning?</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>4-Consistently</td>
</tr>
<tr>
<td></td>
<td>3-Once or twice students provide evidence for their claims OR students explain their thinking, using conceptual explanations.</td>
</tr>
<tr>
<td></td>
<td>2-Students provide computational, procedural or of memorized knowledge, OR what little evidence/reasoning given is inaccurate, incomplete, or vague.</td>
</tr>
<tr>
<td></td>
<td>1-Speakers do not back up their claims or explain reasoning.</td>
</tr>
<tr>
<td></td>
<td>0-Class discussion not related to math or no class discussion.</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Students' Linking Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>4-Students CONSISTENTLY connect their contributions to each other and show how they relate.</td>
</tr>
<tr>
<td></td>
<td>3-Students link at least twice (connect AND relate).</td>
</tr>
<tr>
<td></td>
<td>2-At one or more times, students link ideas, but did not show HOW ideas relate to each other OR only one strong effort is made to connect.</td>
</tr>
<tr>
<td></td>
<td>1-Students do not link or revoice students' contributions.</td>
</tr>
<tr>
<td></td>
<td>0-No whole class discussion OR discussion not related to math.</td>
</tr>
<tr>
<td>Other:</td>
<td></td>
</tr>
</tbody>
</table>
|   | Did the task have potential to engage students in rigorous thinking about challenging content?  
|   | Cognitive Demand: Potential of the Task  
|   | 4-Doing mathematics; using complex and nonalgorithmic thinking (not a predictable, well-rehearsed approach explicitly suggested by the task, task instructions, or a worked-out ex.); OR Procedures with Connections: applying a broad general procedure that remains connected to underlying mathematical concepts.  
|   | 3-Task has potential for cognitive demand, but either there is no prompt for evidence of students' reasoning and understanding, it is not appropriate for the specific group of students (too easy/hard), students identify patterns but do not generalize/justify, may have multiple representations but are not explicitly prompted to form connections between them, or students may be asked to make conjectures but not asked to provide mathematical evidence/explanations to support conclusions.  
|   | 2-Limited potential for cognitive demand; students are using a procedure that is either specifically called for or its use is evident based on prior instruction, experience or placement of the task. There is little ambiguity in what needs to be done and how to do it. Task does not require connections to concepts or meaning underlying procedure. Focus on correct answers rather than developing mathematical understanding. OR, Task is at least 2 grade levels below.  
|   | 1-Task potential limited to engaging students in memorizing or reproducing facts, rules, formulae, or definitions. The task does not require students to make connections to the concepts/meaning that underlie the facts, rules, formulae, or definitions.  

|   | Task Implementation: At which level did the teacher guide students to engage with the task in implementation?  
|   | During the task  
|   | 4- Explicit evidence of of students' reasoning/understanding. Task is “Doing Mathematics” or “Procedures with Connections.” (Ex. solved challenging problem, showing reasoning in work, developed explanation for why a formula works, generalized from patterns to rule and justified why it always works, explicit connections between representations, followed a procedure to explain/illustrate a math concept, process or relationship)  
|   | 3-Task offers high cognitive demand but no explicit evidence of student reasoning, conjectures not backed up by math evidence or explanations, multiple reps without connections between them, or identified patterns but did not generalize, or underlying math not at appropriate level for students to sustain high cognitive demand.  
|   | 2- Students engaged in using a procedure with little ambiguity about what to do, no connections to concepts or meaning, focus on correct answers rather on developing mathematical understanding OR activity was at least 2-grade levels below the students' level.  
|   | 1-Low Cognitive Demand-Students are engaged in memorizing a procedure or reproducing facts, rules, formula, or definitions without connections or concepts or meaning.  
|   | 0-Students did not engage in mathematical activity.  

271
### Student Discussion Following Task

To what extent did students show their work and explain their thinking about the important mathematical content?

- **4** - Students show written work for solving and/or engage in a discussion of the important mathematical ideas in the task. During discussion, students provide complete and thorough explanations of why their strategy, idea, or procedure is valid. Students explain why their strategy works and is appropriate to the problem and make connections to underlying mathematical idea. OR students discuss more than one strategy/representation for solving the task, provide explanations of why both work and/or connect them.

- **3** - Students show/describe work and/or engage in a discussion of the important math ideas in the task. During discussion, students provide explanations of why their strategy, idea, procedure, is valid and/or students begin to make connections BUT the explanations and explanations are not complete, lack precision, or fall short in making explicit connections.

- **2** - Students show or describe written work but do not engage in the discussion of why their strategies, procedures, or mathematical ideas work (do not connect to concepts). OR Students only show/discuss one strategy or representation OR Students make presentations of their work with no questioning or prompting by teacher to explain the mathematical work, make connections, etc. (Presentations with no discussion)

- **1** - Students provide brief answers OR nonmathematical responses.

- **0** - No discussion of the task.

### Rigor of Teacher's Questions

During the video, tally question types: Probing (P), exploring mathematical meaning and relationships (MM), generating discussion (GD), procedural/factual (PF), other mathematical (OM), nonmathematical (N).

- **4** - Teacher consistently asked students to explain, elaborate thinking (P/GD), identify important mathematical ideas in the lesson, form connections between ideas, representations, or strategies (MM).

- **3** - At least 2 times, teacher asks P, GD, and MM questions.

- **2** - There are one or more superficial, trivial, or formulaic efforts to ask deeper cognitive demand questions or to ask students to explain their reasoning.

- **1** - Teacher asks procedural/factual (PF) questions that elicit facts or procedures or require brief, one-word responses.

- **0** - Teacher did not ask questions during task or questions not relevant to mathematical discussion of big ideas.
| 23 | Mathematical Residue: What is the extent to which the whole-group discussion builds new, important mathematical ideas?  
Big Ideas/Essential Questions  
4-Discussion following student work surfaces the important mathematical ideas, concepts, or connections embedded in the task and serves to extend students' understanding of the main mathematical goals/ideas/concepts of the lesson. Discussion leaves behind important math residue.  
3-Mathematics during the discussion is only partially developed, perhaps due to time or students' readiness, but they wrestle with the big ideas. Main concepts not pursued in great depth or have not solidified by lesson close.  
2-During class discussion, either 1. teacher tells students the important mathematical ideas, concepts, connections while students take notes or provide brief answers but do not make meaningful contributions to discussion/superficial contributions taken over by teacher OR 2. Discussion is mathematical but does not hit the concepts, big ideas, or connections underlying the task OR 3. The discussion is about mathematics that is not relevant/important to this group.  
1-Important big ideas/essential questions do not surface during the discussion following student work on the task. Discussion may not have a clear focus on developing the big math concept, or discussion was about nonmathematical aspects of the task. No mathematical residue left behind.  
0-No discussion following the task. |
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>FOR PEER REVIEWER COACHES ONLY: Describe what went well with student reasoning and discussion.</td>
</tr>
<tr>
<td>25</td>
<td>FOR PEER REVIEWERS/COACHES ONLY: Please list several possible strategies for increasing the cognitive demand of the task, student reasoning, or communication.</td>
</tr>
<tr>
<td>26</td>
<td>FOR TEACHER IN VIDEO: Please list what you are happy with in terms of the cognitive demand, student reasoning and communication during your task/discussion.</td>
</tr>
<tr>
<td>27</td>
<td>FOR TEACHER IN VIDEO: Please list what you would like to change?</td>
</tr>
<tr>
<td>28</td>
<td>FOR TEACHER IN VIDEO: What question(s) will you research to find strategies to implement now before next video?</td>
</tr>
</tbody>
</table>
References


Y. Li, E. Silver, & S. Li (Eds.), *Transforming Mathematics Instruction* (pp. 259-281). Cham, Switzerland: Springer International Publishing.


Boston, M. D. (2014). Assessing instructional quality in mathematics classrooms through collections of students' work. In Y. Li, E. Silver, & S. Li (Eds.), *Transforming mathematics instruction* (pp. 501-523). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-04993-9_27


275


280


technology, engineering, and mathematics learning and teaching (pp. 21-46). London, UK: Routledge.


Schoenfeld, A. H. (2014, November). What makes for powerful classrooms, and how can we support teachers in creating them? *Educational Researcher, 43*(8), 404-412. DOI: 10.3102/0013189X1455


Biography

Deb Crawford currently serves as Mathematics Instructional Supervisor with Frederick County Public Schools and has been involved in a range of projects across the K-12 spectrum in a cross-section of schools in Pennsylvania, Texas, and Virginia. She has served as a math and science teacher, math department chair, math specialist, Texas Instruments T³ Instructor, speaker, and author. Her research foci include curriculum design research, professional development of math teachers, teacher action research, personalized problem-based learning, and math modeling with data collection technologies.

Debbie has coauthored 10 books for math and science teachers, including *Hands-On Teaching: H.O.T. Strategies for Using Math Manipulatives*, and *Activities for Middle Grades Science With the CBL 2 and TI-73*. She has enjoyed sharing with teachers by presenting at conferences including NCTM, VCTM, Teachers Teaching with Technology (T³) and the SOL Content Academies at JMU. Most importantly, she enjoys being the mother of seven children.