2011-08-12

Measuring Mathematics Instruction in Elementary Classrooms: Comprehensive Mathematics Instruction (CMI) Observation Protocol Development and Validation

Sue Ann Womack
Brigham Young University - Provo

Measuring Mathematics Instruction in Elementary Classrooms:

Comprehensive Mathematics Instruction (CMI)

Observation Protocol Development

and Validation

Sue A. Womack

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Sterling C. Hilton, Chair
Douglas L. Corey
Ellen J. Williams
Damon L. Bahr
Pamela R. Hallam

Department of Educational Leadership and Foundations

Brigham Young University

August 2011

ABSTRACT

Measuring Mathematics Instruction in Elementary Classrooms:
Comprehensive Mathematics Instruction (CMI)
Observation Protocol Development
and Validation

Sue A. Womack
Department of Educational Leadership and Foundations
Doctor of Philosophy

Despite the availability of reform standards in mathematics since 1989 (National Council of Teachers of Mathematics), teachers have not yet aligned instruction with reform ideals on a widespread basis. (Cohen & Hill, 2000; Hiebert, et al., 2005; Spillane & Zeuli, 1999; J. W. Stigler & Hiebert, 1999). Consequently, mathematics education in elementary schools has not produced students with strong mathematical understanding (Hiebert & Grouws, 2007).

The Comprehensive Mathematics Instruction (CMI) Framework (Hendrickson, Hilton, & Bahr, 2007) was created to assist teachers in knowing how to teach for student understanding. As the CMI Framework has been implemented in schools, it is necessary to measure the impact the framework is having on instruction and on student learning through measuring the instruction that is occurring in classrooms. Prior to this dissertation research, there was not an instrument fully aligned to the CMI Framework for measuring classroom instruction.

The tool developed and validated in this research, the Comprehensive Mathematics Instruction (CMI) Observation Protocol, measures instruction through the lens of the CMI Framework. The results show three types of evidence of validity from the measurement perspective: content evidence, response processes evidence, and internal structure evidence.

Keywords: instructional measurement; elementary mathematics measure; observation protocol;

ACKNOWLEDGMENTS

This doctoral journey has been undertaken partly because I stand on the shoulders of my grandmother, Blanche Marie Lloyd Oakley, (1895–1986), who wanted to be a doctor when women were not supposed to have high aspirations. She fought against odds to make a path, and although she did not become a doctor, she became a fine nurse. She provided me with inspiration to continue striving when the going was rough. I also owe much to my father, Lloyd George Oakley (1929–1988) who wanted to be a doctor as well, but whose humble circumstances only allowed a high school education. He was a pioneer of sorts in the infancy of the computer programming field. His example of one who stayed on top of the rapidly changing information technology when others fell by the wayside has inspired me to continually learn, change, and improve. To my mom, Marion Francis Elliott Oakley, who unfailingly believes I am better than I am, I am also grateful. Although I know the truth about me, I appreciate her confidence in me and strive to become what she thinks I am. I acknowledge my husband, Steven Womack, who is always my head cheerleader. Thank you for the countless hours you spent alone while I was at class, in my office studying, or in my grad carrel writing. Without you as my grounding friend, I would have lost my way. Thank you too for catering my cohort’s dinners during our last year of coursework, without which we would’ve perished, I’m certain!

While my family has provided love, inspiration, and sustenance, my committee has provided thoughtful insight, rigor, more inspiration, great discussions, and several occasions for snacks. I am indebted to Dr. Damon Bahr and Dr. Ellen Williams for early conversations regarding instruction which influenced my thinking. Dr. Doug Corey provided significant input through discussions about measuring instruction. Thanks as well to Dr. Pam Hallam for her thoughtful reviews of my work and her ever-cheerful inquiries regarding my progress.

To Dr. Sterling Hilton, my chair and mentor, I owe more than I can articulate. His questioning pushed my thinking, and his example of scholarship and teaching gave me a vision of what I might aspire to.

Finally, I acknowledge God, my Heavenly Father, who has provided the "lamp unto my feet" (Psalms 119:105). It is He who set me on this path and sustained me through the journey. He placed wonderful people into my acquaintance who have enriched my spirit and my intellect. My purpose in pursuing this doctoral degree was to better fit me to serve Him and His children. I entered to learn; now I go forth to serve.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

The development and validation of a new measure of instruction is the focus of this dissertation. Instructional measures, though, are set in a particular context. In this dissertation Chapter 1 introduces briefly the context that gives rise to the need for a new measure. The introductory chapter begins with a discussion of the importance of facility with mathematics. Included in this discussion is the decades-long quest for student mathematical competence through instructional improvement in mathematics education. A recently developed framework for instruction is also introduced, with its conceptual framework of instruction. Although the instructional framework is discussed throughout this dissertation prospectus, justification of the framework is not the purpose of the dissertation work. The instructional framework creates the lens through which instruction will be judged, and is the basis for the new instructional measure. The introduction concludes with the need for a new measure for instruction.

The review of literature follows as Chapter 2. The literature review elaborates on the instructional framework. The view of instruction the framework offers will be discussed with examples of what students and teachers do during instruction. The review of literature also looks at how instruction has typically been measured, and the strengths and weaknesses of those methods. Issues in reliability and validity of measures of instruction will be addressed and a unitary concept of validity will be elaborated upon.

Chapter 3 explains the methods used in developing the protocol and in answering the research questions about the validity evidence for the new measure of instruction. As is typical in a dissertation, Chapter 4 will report the results of the development activities and the validation study, and Chapter 5 will conclude with a final discussion of the research.

National Background of the Problem

Student mathematical understanding is a component of the definition of mathematical competence and has been in the forefront of the United States' educational reform for the past three decades (National Council of Teachers of Mathematics [NCTM], 1980, 1989, 1991, 2000; National Mathematics Advisory Panel, 2008; National Research Council [NRC], 1989, 2001). It is agreed upon by the NCTM, the National Mathematics Advisory Panel, and the NRC that mathematics is a gatekeeper for educational attainment, career opportunities, and national productivity and security. For example, the NCTM states, "In this changing world, those who understand and can do mathematics will have significantly enhanced opportunities and options for shaping their futures. Mathematical competence opens doors to productive futures. A lack of mathematical competence keeps those doors closed" (2000, Executive Summary). The NCTM implicitly defines "mathematical competence" as understanding and doing mathematics. The National Mathematics Advisory Panel (2008) corroborates and expands the NCTM argument:

> In the contemporary world, an educated technical workforce undergirds national leadership. . . . There are consequences to a weakening of American independence and leadership in mathematics. . . . We risk our ability to adapt to change. We risk technological surprise to our economic viability and to the foundations of our country's security. . . . Sound education in mathematics across the population is a national interest. Success in mathematics education also is important for individual citizens, because it gives them college and career options, and it increases prospects for future income. (pp. xi–xii)

Clearly, from the statements cited, mathematical competence is not only valued, but understanding and doing mathematics is a critical component of 21^{st} century life.

In support of helping students attain mathematical competence through improved instruction, standards were issued by the National Council of Teachers of Mathematics in 1989. The standards were intended to "(1) to ensure quality, (2) to indicate goals, and (3) to promote change" (NCTM, 1989, Introduction, paragraph 6). The framers of the standards hoped the standards would serve as "facilitators of reform" (NCTM, 1989, Introduction, Paragraph 10). The NCTM standards also created a vision for what instruction for mathematical competence might look like and how students with mathematical competence might be expected to perform. The vision created in the 1989 NCTM document was more explicitly articulated in the updated Principles and Standards for School Mathematics in 2000. The vision created rests on six principles (NCTM, 2000, p 11):

1. Equity - Excellence in mathematics education requires equity—high expectations and strong support for all students.
2. Curriculum - A curriculum is more than a collection of activities: it must be coherent, focused on important mathematics, and well articulated across the grades.
3. Teaching - Effective mathematics teaching requires understanding what students know and need to learn and then challenging and supporting them to learn it well.
4. Learning - Students must learn mathematics with understanding, actively building new knowledge from experience and prior knowledge.
5. Assessment - Assessment should support the learning of important mathematics and furnish useful information to both teachers and students.
6. Technology - Technology is essential in teaching and learning mathematics; it influences the mathematics that is taught and enhances students' learning.

The initial NCTM standards were precipitated by evidence that traditional mathematics instruction was not producing students with mathematical competence. Likewise, neither the standards' promise of changing instruction, nor the promise of the principles in facilitating the vision of students exhibiting mathematical competence have come to fruition in the years following the NCTM issuance of standards and the periodic updates to those standards (Cohen & Hill, 2000; Hiebert, et al., 2005; Spillane & Zeuli, 1999; J. W. Stigler & Hiebert, 1999).

Teachers were having ongoing difficulties changing instructional practice substantively, and most instruction remained at the periphery of the NCTM recommendations (Cohen & Hill, 2000; Hiebert, et al., 2005; Spillane & Zeuli, 1999; J. W. Stigler & Hiebert, 1999). Much instruction centered on procedures without understanding. Even when teachers reported using "reform" instruction aligned with NCTM recommendations, many had simply re-labeled traditional strategies and hadn't changed practice. When confronted with the realization that instruction and student competence were not improving, researchers and policymakers responded in several ways. New curriculum programs, materials, and policy instruments were developed and researched in efforts to assist teachers in improving instruction for mathematical competence (Cohen & Spillane, 1992; Collopy, 2003; Desimone, Porter, Garet, Yoon, & Birman, 2002; Slaven, Lake, & Groff, 2009). There has been some success in these efforts, but a widespread change in mathematics instruction has not yet occurred (Slaven, et al., 2009). Without principled understandings of what the NCTM vision means for practice, teachers following programs may not attain an integrated and flexible understanding of how to plan for, implement, and assess instruction for mathematical competence (Ball & Cohen, 1996; Cohen & Ball, 1999; Cohen, et al., 1990).

A Local Mandate to Improve Mathematics

With the previously mentioned national backdrop, in 2003 the Brigham Young University Public School Partnership's Governing Board and the Center for Improving Teacher Education and Schooling (CITES) formed and gave an assignment to the Math Initiative Committee (MIC): identify best practices in mathematics and improve mathematics. The MIC was composed of researchers and educators from Brigham Young University, as well as public school math educators and administrators from the five surrounding partner districts. The MIC used research literature and collective experiences to develop common understandings and a common mission from which the Comprehensive Mathematics Instruction (CMI) Framework (Hendrickson, Hilton, & Bahr, 2008) grew. The CMI Framework is a framework of principles upon which practices may be based to create balanced mathematics experiences leading students to deepen their mathematical thinking and understanding. The CMI Framework helps teachers to translate the vision and theory of reform mathematics into practice.

Comprehensive Mathematics Instruction (CMI) Framework

The CMI Framework acknowledges and promotes the interactive nature of instruction. Instruction is conceptualized as interactions among students and between the teacher and students, with the environment affecting and being affected by the human interactions (Cohen, Raudenbush, & Ball, 2003). The interactions just noted are depicted as the foundational layer, or interaction level, in Figure 1. Additionally, instruction is conceived as a system (Hiebert & Grouws, 2007) with interplay among curriculum, content, materials, tasks, discourse, and assessment, shown in Figure 1 as the top layer, or content level. The two layers interact within and between each other, forming a system of interactions. Choices in one area affect options available in another. For example, a particular task will center on certain content at the exclusion of other content and will suggest certain materials, influence the direction and nature of discourse, and yield different assessment information than another task. Additionally, the same task might look different at another point in the curriculum, and the approaches of students may influence any of the instructional elements. Thus, instruction is somewhat of a package or system. What these interactions look like during a lesson will be illustrated concretely in the review of literature.

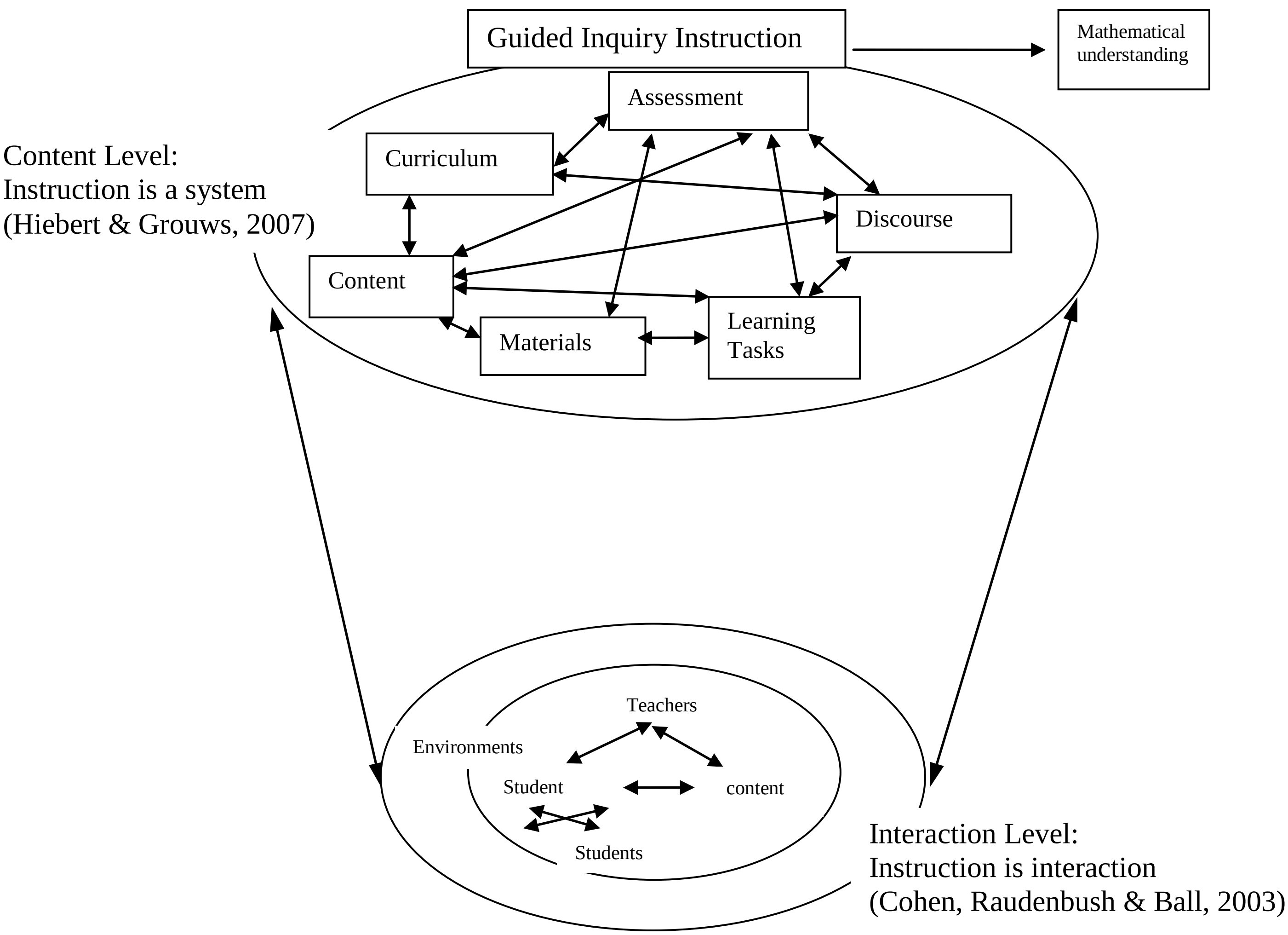

Figure 1. Conceptual model of instruction

The CMI Framework is not a program or prescription but a structure from which teachers can make the instructional decisions and plan for the interactions that will lead to effective teaching and student learning. The components of the CMI Framework represent a set, out of many different sets, that might be said to be sufficient to create the outcome of student mathematical understanding. The CMI Framework consists of teaching cycles embedded in a learning cycle (see Figure 2).

- Learning Cycle: Develop – Solidify – Practice Understanding
- Teaching Cycle within each phase of the Learning Cycle
- Stages of the Teaching Cycle, Launch – Explore – Discuss, are different across the phases of the Learning Cycle

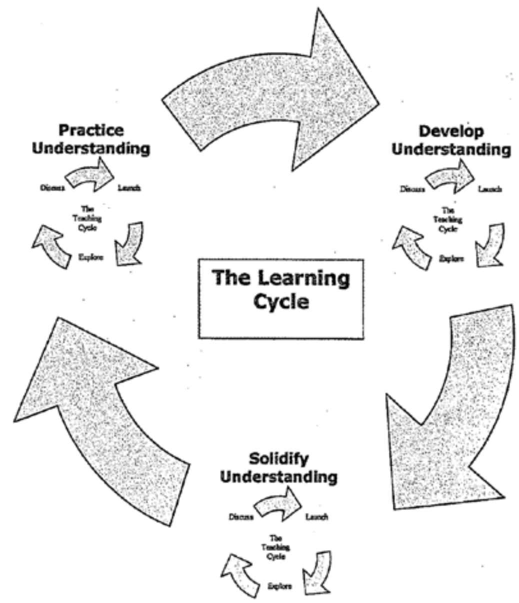

Figure 2. The Learning Cycle, showing the embedded Teaching Cycles

Note. From "The Comprehensive Mathematics Instruction (CMI) Framework: A new lens for examining teaching and learning in the mathematics classroom," by S. Hendrickson, S. C. Hilton, and D. Bahr, 2008, *Utah Mathematics Teacher*, Fall, p.44. Reprinted with permission

Each teaching cycle has a Launch stage, an Explore stage, and a Discuss stage. The Launch stage creates a context for the mathematics that will follow. The Explore stage provides a task or series of tasks within which students explore the mathematics individually or in groups. And the Discuss stage allows the entire class of students, collectively or individually, under the guidance of the teacher, to clarify, explain, justify, prove, and connect the mathematics just explored.

The learning cycle consists of three phases designed to increase students' mathematical understanding. When beginning a new concept, teachers plan lessons to develop understanding. Develop Understanding lessons, the first phase of the learning cycle, help situate the mathematical idea, surface student understanding or misunderstandings, and give a general overview of the important mathematics being developed. Solidify Understanding lessons, the second phase of the learning cycle, have a different purpose from Develop Understanding lessons. Solidify Understanding lessons aim to help students examine and extend their ideas to form a more solid concept of the mathematics being learned. The final phase of the learning cycle is Practice Understanding. Practice Understanding lessons allow students to refine and develop fluency with the mathematics, including the definitions, properties, procedures, and models that the mathematical community at large understand and use.

The three phases of the learning cycle, Develop Understanding, Solidify Understanding, and Practice Understanding are key in making the CMI Framework *comprehensive*. Under a "reform" label, mathematics instruction has sometimes incorporated inquiry or discovery, which is similar to the Develop Understanding phase's purpose of surfacing student thinking. However, the student exploration has frequently been an end, with teachers making no further use of what has been surfaced by students, leaving that discovery in an unexamined, unconnected state (Stein, Engle, Smith & Hughes, 2008). Traditional instruction most often has been concerned with learning and practicing procedures (Porter, 1989; Stigler et al., 1999), which has some similarities to the Practice Understanding phase. Traditional practice generally does not connect the procedures with their conceptual underpinnings. Neither discovery nor traditional instruction alone has been successful in building student mathematical understanding, as evidenced by the relatively flat mathematics achievement scores (Stigler et al., 1999). The Solidify Understanding phase of the CMI Framework forms a connecting bridge between the new understandings that have been developed in the Develop Understanding phase and the necessary fluency that is fostered in the Practice Understanding phase. The ideas, strategies, and representations that have been uncovered in the Develop Understanding phase are examined, extended, and connected with previous learning and become solidly understood. The more fragile ideas become concepts, the strategies solidify into algorithms, and the representations become more thoroughly understood, making them useful tools.

This framework of instruction explicitly acknowledges that instructional interactions differ depending on the type of understanding that is being built by students during the different learning/teaching cycle combinations. Within each phase of the learning cycle, the general idea of a Launch, Explore, or Discuss phase remains consistent. However, the purposes of the Launch, Explore, and Discuss stages of the teaching cycle, and the roles played by the teacher and students within those phases, change. A more complete explication of the learning cycle and the teaching cycles will be undertaken in the Review of Literature (see also Appendix A for the CMI Framework).

Within each phase combination, important mathematics are pursued through worthwhile tasks, classroom discourse, embedded assessment, and making use of student thinking within a coherent curriculum. “Important” mathematics is defined by the NCTM as “mathematics content and processes that are . . . worth the time and attention of students. Mathematics topics may be important for different reasons, such as their utility in developing other mathematical ideas, in linking different areas of mathematics and in preparing students for college, the workforce, and citizenship” (NCTM, 2009, paragraph 4). The CMI Framework provides teachers with a detailed way of thinking about, planning for, and engaging in instruction with students to facilitate mathematical understanding.

Frameworks such as the Comprehensive Mathematics Instruction Framework are important in enacting systems of effective instruction that will reliably improve student outcomes. Although the CMI Framework's principles and practices have been assembled based on scholarly evidence, there is much to be studied about the Framework's effects on teaching and learning. Foundational questions about whether the CMI Framework's system of instruction improves student mathematical understanding and about what affect the CMI Framework has on teachers' instructional practice need to be answered. A link between CMI Framework instruction and improved student understanding may be claimed only if there is evidence of the degree to which the instruction students are receiving aligns with the Framework. Without knowing what is actually occurring instructionally, there may be many other equally likely explanations for student improvement. The evidence that is necessary will come through measuring instruction.

Likewise, the evidence for the impact of the CMI Framework on instructional practice will come from measuring the instruction that takes place. Documenting the alignment of instructional practice with the CMI Framework as teachers learn to implement the type of instruction promoted by the Framework will provide evidence for the CMI Framework's influence on instruction. However, gathering the evidence to answer the questions raised by the CMI Framework necessitates the development of a measure that is a valid and reliable gauge of instruction that is aligned to the CMI Framework.

Instructional Measurement

Instructional measures focus on salient features of instruction, generally from a particular view of what constitutes good instruction (Ball & Rowan, 2004; Corey, 2007). The creation of a new instructional framework, such as the CMI Framework, is also the creation of a new lens through which to judge instruction. Therefore, existing measures of instruction almost certainly will not adequately align with a new framework. A fresh feature of the CMI Framework is the acknowledgment of instruction as an interactive system of teaching and learning. Currently instruction based on the CMI Framework has been measured by the Inside the Classroom Observation Protocol (Weiss, Pasley, Smith, Banilower, & Heck, 2003), which has been modified. However, existing measures, including the Inside the Classroom Observation Protocol, do not account adequately for the many aspects of the framework; thus they do not align completely with the CMI Framework, producing the need for a new measure, the Comprehensive Mathematics Instruction Observation Protocol.

Purpose of the Study

The purpose of this dissertation work was to develop a measure of instruction, the CMI Observation Protocol, using the CMI Framework as a lens through which the mathematics instruction of teachers will be viewed. An additional, equally important purpose was to gather validity evidence to support the CMI Observation Protocol's use as a tool that can be trusted to measure the system of instruction defined by the CMI Framework.

Chapter 2: Review of Literature

Instruction, generally, has been linked to student achievement (Brophy & Good, 1986; Corcoran, 2007; Darling-Hammond, 2000; Hiebert & Grouws, 2007). When evidence began to accumulate that student achievement in mathematics was insufficient for the current demands of economic and personal success, improving instruction was a logical solution (Stigler & Hiebert, 1999). However, the path to instructional improvement seemed to come in a circuitous route, via some redefinitions of school mathematics.

Reform-oriented Mathematics

The National Council of Teachers of Mathematics (NCTM) issued an agenda for action (1980) calling for a change in emphasis from procedures in isolation to problem solving in context. The mandate for instruction was to focus on effective and efficient techniques derived from research (1980, Recommendation 4, paragraph 3). In all, the eight recommendations of the NCTM addressed student learning, instruction, assessment, and policy, and were a call for reform in school mathematics. Nearly a decade later, another document from the NCTM, *Curriculum and Evaluation Standards for School Mathematics* (1989), outlined thirteen standards for each grade cluster (K–4; 5–8; 9–12). The standards suggested curriculum rather than instruction, but instructional assumptions were implicit in the examples of how a standard might be enacted. For example, in summarizing how basic subtraction facts might be presented in a problem-solving setting, the NCTM document stated, " mathematical ideas have originated with the children, rather than the teacher, in an *inquiry-oriented manner*" (Emphasis added, NCTM, 1989, K–4 Standard 1, Discussion, paragraph 4). In 1991, the NCTM published *Professional Standards for Teaching Mathematics* (1991) which had as its purpose,

> To make explicit and expand the images of teaching and learning implicit in the *Curriculum and Evaluation Standards for School Mathematics*, to elaborate a vision of instruction that can light the path toward such change. . . . Good teaching demands that teachers reason about pedagogy in professionally defensible ways within the particular contexts of their own work. The standards for teaching mathematics are designed to help guide the processes of such reasoning, highlighting issues that are crucial in creating the kind of teaching practice that supports the learning goals of the *Curriculum and Evaluation Standards for School Mathematics.* (1991, Introduction)

What the NCTM was recommending was mathematics learning that was conceptually-based. Conceptually-based learning has several components: reasoning, problem solving, and communication. Connections are central, and procedural mathematics are learned with understanding of when and how they were useful and why they work. Instructionally, it was clear that what was expected was qualitatively different than traditionally had been the case. One difference was the rejection of the two-track system of basic arithmetic for most and a full range of mathematics for an elite few, which had implications for instruction. If all students could and were expected to learn complex mathematics, then instruction would need to attend to student differences. Another difference impacting instruction was student engagement with mathematics. The standards spoke about instruction that engaged students with tasks and that expected students to conjecture, develop arguments, and validate solutions. Engaging students with tasks and expecting students to think have the potential to create instruction that is more student-centered than teacher-centered. Teachers needed to have a principle-based understanding of how to achieve student engagement, conjecture, argumentation, and proof in order to create instruction leading to the desired outcomes. The “professionally defensible” reasoning about pedagogy in context that the NCTM spoke of (NCTM, 1991, Introduction) is derived from the NCTM standards, which forms a foundation for principle-based understanding.

At about the same time as the NCTM's publication of *Curriculum and Evaluation Standards for School Mathematics*, the National Research Council (NRC) convened a panel to address the state of mathematics education. The document produced by the panel, *Everybody Counts* (NRC, 1989), recommended an active construction of mathematics that at once is "truth and beauty; utility and application" (p. 43), rather than transmittal of solely algorithmic work. The intent behind the language of truth and beauty was that, unlike a conception of mathematics as canon to be transmitted, mathematics is a discipline of sense-making and power, implying that it requires the learner and instructor to interact with each other and with the content. The NRC position was that all students should build mathematical power. Like the NCTM, the NRC report pointed out that a "significant common core of mathematics for all students" (1989, p. 81) was essential for success in the approaching 21st century.

A decade before the turn of the 21st century, then, the stage was set to significantly reform mathematics curricula and instruction. Mathematics education was re-conceptualized by mathematics educators from academe and public education. The reform conceptualization consisted of several components: (1) more, though not exclusively, student-centered instruction; (2) problem-solving in context; (3) a recognition of the necessity of conceptual understanding; (4) a decreased (but not absent) emphasis on procedures and rote skills; (5) the need for forging connections among concepts and across domains; and (6) an appreciation of mathematics as a useful pursuit, not something to be avoided (NCTM, 1989; NRC, 1989). Subsequent refinements by both the NCTM (2000) and the NRC (2001), along with a renewed emphasis by a new body, the National Mathematics Advisory Panel (2008), clarified that what has come to be called “mathematical proficiency” (National Research Council, 2001, p. 5) is a multi-dimensional prospect, as depicted in Table 1.

Table 1

A Synthesis and Comparison of Reform Mathematics: Conceptions of Student Mathematical Proficiency

NCTM (2000)	NRC (2001)	National Mathematics Advisory Panel (2008)
conceptual understanding	conceptual understanding	conceptual understanding
procedural facility	procedural fluency	procedural fluency
factual knowledge	----	automatic fact recall
knowledge flexibility	adaptive reasoning	----
----	strategic competence	problem-solving skills
perseverance	productive disposition	effort
autonomy	----	----

Although the NCTM, the NRC, and the National Mathematics Advisory Panel name the elements of mathematical proficiency slightly differently, the elements as a whole embody similar ideas. For example, the NCTM labels the ability to know addition, subtraction, multiplication, and division facts as *factual knowledge*, while the National Mathematics Advisory Panel labels the same element as *automatic fact recall*, and the NRC wraps that sort of knowledge into *procedural fluency*. The theoretical guidance begun in 1980 continued to be refined in order to assist educators in creating a new order of instruction.

The reform conceptualizations of mathematical proficiency created a chain of events leading to the necessity of improved instruction. A new conceptualization of school mathematics made new curricula necessary. The new view of what students should be able to know and do with the new curriculum, suggested that new experiences needed to occur for students. These new experiences suggested that a new paradigm of instruction was needed, which would be an improvement if the new paradigm resulted in student mathematical proficiency.

The shift to a new paradigm for instruction, which was suggested by the NCTM (2000), has been difficult. According to Cohen and Ball there was a major reason for the lack of improvement in instruction and the difficulty in making the paradigm shift:

> Even when interventions explicitly introduce new curricular materials or provide teacher "training," they rarely create adequate conditions for teachers to learn about or develop the knowledge, skills, and beliefs needed to enact these interventions successfully in classrooms. (Cohen & Ball, 1999, p. 1)

By design, the reform standards in mathematics were offered as a "vision, and not a recipe"(NCTM, 1991, Introduction, paragraph 5). Hendrickson, Hilton & Bahr (2008) asserted that "the conscientious lack of a prescriptive pedagogy often leaves teachers without a clear sense of direction" (p. 45). This lack of a clear sense of direction is one problem teachers have faced in knowing how to align their teaching with the reform vision of instructional improvement. The Comprehensive Mathematics Instruction (CMI) Framework was developed to help provide the adequate conditions spoken of by Cohen and Ball (1999) that will assist teachers in shifting to instruction which effectively leads students to mathematical understanding in today's conception of mathematics education.

CMI Framework Origins

Constructivism is a strong influence in the CMI Framework. From a constructivist perspective learners are active participants in constructing new knowledge (Ornstein & Hunkins, 2009). Learners make sense of new inputs by connecting the new information with existing information. Learners re-form either the input to conform to the existing information, or the existing information to conform to the new input (Fosnot, 1996 in Baylor, Samsonov & Smith, 2005; Piaget 1964). Some constructivists argue that students should be free to explore, discover, and make connections without structures imposed and that humans cannot know truth, but only have their individual interpretation of it. Von Glasersfeld (1992), who is considered the father of modern radical constructivism, stated: “Constructivist teachers can never justify what they teach by claiming that it is ‘true.’ . . . You activate students’ minds to construct knowledge by letting them struggle with problems of their own choice, helping them only when they ask for help. At best, the teacher can orient a students’ constructing in a fruitful direction, she or he can never force it” (p. 178). Radical constructivism, in Von Glasersfeld’s definition, is not extreme, but a departure from traditional ideas about knowledge.

Vygotsky (1978) took a different view of learning. Although active construction of knowledge by students is fundamental to Vygotsky’s theory, he promoted the social, interactive nature of knowing and learning. In contrast to radical constructivist views of helping only when students ask for help, Vygotsky’s social constructivism advanced the idea that if children collaborate with more skilled peers or adults, they will be able to do what they cannot yet do alone. By collaborating, they can reach higher levels of thinking or learning which then become part of their own knowing. Vygotsky argued for a zone of proximal development, the ZPD, which was just above the child’s present level of development. Within the ZPD, work with more knowledgeable others brings about growth and learning because it “awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers” (Vygotsky, 1978, p. 90). The actions of the knowledgeable others while working in a child’s ZPD has also been termed “scaffolding” from Bruner’s work (Wood, Bruner, & Ross, 1976). Once the developmental processes become internalized, the child is able to operate independently with them.

The CMI Framework, which will be fully explicated in the next section, is rooted in a moderated constructivist theory of instruction using an inquiry-based model of instruction which developers prefer to term *guided inquiry*. The CMI Framework developers' use of the term guided inquiry connotes that the teacher makes conscious decisions in shaping student inquiry, in contrast to an open inquiry process of radical constructivism in which students engage in free-form investigation. In CMI instruction, the teacher is still largely charged with selecting the important mathematics with which students will engage during inquiry. The teacher also has an important role in questioning and making use of student thinking to assist the students in sense-making and coming to correct thinking about the mathematics. Social-constructivist (Vygotsky, 1978) views of learning are clearly seen in the CMI Framework's promotion of interactions between teacher and student, among teacher and students, and between and among students. Student roles throughout the CMI Framework charge students with actively listening to the teacher and other students; questioning themselves, the teacher, and other students; connecting previous work and knowledge in discussions with others; and reflecting to themselves and with others on the current work. The CMI Framework entrusts teachers with the responsibility to select tasks essentially in the ZPD, and then facilitate student thinking and learning through listening to students, questioning appropriately, and adjusting the pace or content based on students' current understandings. Vygotsky's zone of proximal development concept of supporting students in performing tasks they would not have sufficient skill for alone, scaffolding learning, is also evident in the CMI Framework's emphasis on varying student grouping strategies for different purposes in different phase combinations. Scaffolding is also fostered by CMI's emphasis on using student thinking and in providing teacher guidance through appropriate questioning strategies.

While instruction using the CMI Framework is guided-inquiry based, it also retains effective elements of instruction from the traditional paradigm, in order to "bridge the gap between the good pedagogical strategies of traditional instruction and the recommendations of reform-based instruction" (Hendrickson, et al., 2008, p. 45). Although traditional mathematics instruction focuses on the transmission of information, rather than on sense-making, some of the strategies employed in pursuit of procedural knowledge are transferrable to the CMI Framework's goal of mathematical understanding. Accessing students' prior knowledge, attending to academic engaged time, engaging in active teaching, pacing of instruction, and using questioning strategies, which have all been deemed effective instructional practices (Brophy & Good, 1986), are present in the CMI Framework. In addition, a stated goal of instruction using the CMI Framework is for students to develop mathematical understanding in both the traditional procedural and reform-based conceptual domains "consistent with the broader mathematical community of practice" (Hendrickson, et al., 2008, p. 49). In short, the CMI Framework was developed as a package of cohesive principles to allow teachers to maintain an instructional focus on important mathematics, while capitalizing on student thinking to navigate a path to deepen mathematical understanding and proficiency.

The CMI Framework recognizes when students are first being introduced to a topic during the Develop Understanding phase that the Launch, Explore, and Discuss stages of the Teaching Cycle will be approached differently than during either the Solidify or Practice Understanding phases. The role of teachers promoted by the CMI Framework during all of these phases is consistent with the teaching paradigm promoted by the NCTM's Standards (2000) and what is known about effective instruction (Brophy & Good, 1986; Hiebert & Grouws, 2007). While others might have chosen a different set of principles on which to base practice, the CMI

Framework highlights worthwhile tasks, classroom discourse, embedded assessment, using student thinking, and the relationships between lessons, units, and curriculum as its core elements. The CMI Framework brings together reform-oriented and traditional instruction, and it capitalizes on the systemic and interactive nature of instruction.

CMI Framework Components

The CMI Framework, as previously introduced, embeds teaching cycles within a learning cycle. Each phase of the learning cycle accomplishes a distinctive purpose in extending student learning, while the teaching cycle embedded therein provides the framework for an instructional sequence appropriate to the learning cycle phase (see Appendix A for the CMI Framework document).

The learning cycle of the CMI Framework seeks to build students' mathematical understanding by beginning from students' current understanding, by providing guided experiences to expand and deepen that understanding, and by providing appropriate situations in which to become fluent and refined with thinking about and doing important mathematics. The descriptions of the learning cycle phases, how the teaching cycle looks at each phase of the learning cycle, and what measurable elements the phases contain follow in the next paragraphs.

Develop understanding. The first purpose of the Develop Understanding phase is to surface current student understanding of a selected mathematical purpose. Surfacing student thinking allows the teacher to identify correct understandings as well as misconceptions students already have about the topic. The second purpose is to further develop students' understandings. The teaching cycle for the Develop Understanding phase supports the development of understanding of ideas, strategies, and representations of the mathematical purpose. A given mathematical purpose may require only one lesson or a series of Develop Understanding lessons.

A Develop Understanding lesson begins with a Launch stage followed by an Explore stage, and culminated with a Discuss stage. The purpose of each of the teaching cycle stages supports the purposes of surfacing student thinking and developing understanding. The purpose of the launch in every phase of the learning cycle is to activate student background knowledge and to introduce and clarify the task. In a launch for developing understanding, the Framework specifies that a mathematical purpose aligned with a state or national standard or objective be present and be clear. Additionally, a task with multiple paths to solutions or multiple solutions should be posed. Because a Develop Understanding lesson seeks to surface numerous ideas, both correct and incorrect, an open task is used. The nature of the task requires a slightly different launch. The teacher must be thoughtful about giving just enough information to bring important background knowledge to the forefront and clarifying without degrading the opportunity for students to surface their own ideas as they work on the task. While the teacher carries out the purpose of the launch, the students also have an active role in listening, asking clarifying questions, and accessing their background knowledge.

After a task is launched, students engage in exploring the task posed. During the exploration, students may work individually or be grouped. The teacher makes that decision purposefully based on the task and what configuration she or he anticipates will be most helpful in surfacing student thinking. If the teacher feels that student ideas will be synergistic, a pair or a group will be appropriately selected. If it is anticipated that student thinking will be stunted by group work initially, then the task may begin to be explored by individuals. The teacher's major roles in the Explore stage are to monitor and record student thinking to be used in the discussion, and to maintain the task at an appropriate level through questions and comments. An "appropriate level" means providing optimal scaffolding. This allows students to grapple with the task, while the teacher provides a question or comment that allows students to understand their own thinking and push forward toward a new understanding. It is important that the teacher understand student thinking and know the trajectory through which students will proceed in coming to understand the mathematics. During the Explore stage, the teacher is looking for student thinking that will enable him or her to structure a logical, coherent discussion that will represent the range of student understandings. Monitoring and recording student thinking to use in the discussion enables the teacher to help all students build on their current level of understanding.

Meanwhile, students have several roles in the Explore stage as well. They are to pursue the task, of course. Additionally, they are to engage in sense-making through questioning their own thinking and the thinking of others if they are working together. Students should be actively making connections between the thinking on the task at hand and prior mathematics. Prompting thinking about the connections is part of the teacher's scaffolding strategies.

The Explore stage of the Develop Understanding learning cycle lays the foundation for the Discuss stage by raising numerous ideas, and discovering multiple strategies, and/or representations about the mathematics. The Discuss stage is orchestrated by the teacher so that the multiple ideas, strategies, and representations previously surfaced in the exploration are available to all students. Purposive selection of student thinking is ordered in such a way as to guide listening students through to the main points. Students talk about their discoveries and their thinking about the surfaced concepts or misconceptions. Listening students actively participate by questioning, confirming, or extending what is being presented. The discussion should end with students having a clear idea of what has been discussed.

Develop Understanding lessons look much like many lessons from a constructivist paradigm. However, in the CMI Framework, it is not enough to discover and surface ideas or to allow students to follow their own path until they ask for help, as in VonGlasersfeld's radical constructivism (1992). Surfacing ideas is not the end in itself. Developing understanding is a means toward becoming fluent and having a depth of understanding of mathematical concepts. The Develop Understanding phase is only the beginning step that prepares students to become more solid in the ideas, strategies, and representations they have developed.

Solidify understanding. Selecting from the ideas, strategies, and representations developed previously, the teacher orchestrates a lesson or series of lessons that are focused on solidifying one or two concepts. Solidifying means to examine and extend the idea, strategy, or representation so that it becomes more generalizable as a concept, algorithm, or tool.

Just as a Develop Understanding phase of the teaching cycle contains Launch, Explore, and Discuss stages, so does the Solidify Understanding phase of the learning cycle. The purposes of the teaching cycle stages align with the purpose of the Solidify Understanding phase. The task launched is focused on an idea, strategy, and/or representation and is designed to confirm, connect, generalize, and/or transfer mathematical understanding. The Launch is focused by teacher selection of a string of related problems, a problem with a string of related questions, or a string of related tasks. In the Solidify phase, the background knowledge that the teacher must activate comes from the Develop Understanding phase, but the teacher role remains to launch and clarify the task(s). Students also continue to actively listen, ask clarifying questions, and access background knowledge.

The Explore stage during a Solidify Understanding lesson looks much like that of a Develop Understanding lesson. However, rather than trying to raise numerous ideas, students focus on one or two, making and examining connections and eliminating misconceptions. Students still use appropriate tools and pursue the task in productive groupings that promote sense-making through self- and peer-talk. The teacher's questioning strategies help facilitate the making of connections and the elimination of misconceptions. Student thinking is monitored and recorded for use in a coherently ordered discussion.

The discussion that follows in a Solidify Understanding lesson differs from that of a Develop Understanding lesson by meeting the purposes of this phase of the learning cycle: to confirm, connect, generalize, and transfer mathematical understanding. This moves students to develop ideas, strategies, and representations into concepts, algorithms, and tools. Students are selected to explain and justify their thinking and do so in an order that builds coherently to student understanding. The teacher confirms thinking that coincides with what is accepted in the mathematical community. He or she also questions students in order to help students make explicit the connections they can use to make generalizations. Under teacher guidance, students make meaning of the discussion to solidify their understanding of the selected idea into a concept, a strategy into an algorithm, or a representation into a tool.

Practice understanding. One goal of the Practice Understanding phase is to allow students to refine the concepts, algorithms, and tools developed previously. A second goal is to acquire fluency with the mathematics. When both aims are attained, students have developed definitions, properties, procedures, and models consistent with the mathematics community.

Again, the Practice Understanding phase of the learning cycle contains the teaching cycle stages with appropriate modifications that align with the purpose of a practice understanding lesson. A Launch for Practice Understanding poses a task that re-engages students with concepts, algorithms, or tools that have become solid, but need to become fluent. Fluency is defined by the CMI Framework as accuracy, efficiency, flexibility, and/or automaticity. The Launch of a Practice Understanding lesson arguably may be the most succinct of all. If students are ready to practice it may be a matter of giving the directions to a game or worksheet. The Framework also suggests that practice may be embedded in a Develop Understanding or Solidify Understanding lesson. In this case, the activation of appropriate background knowledge for the Practice Understanding portion will be embedded with the general task launch. The teacher and student roles during a Launch of a Practice Understanding lesson remain to make the task clear and access background knowledge. The teacher is additionally charged with connecting the task with students' previous work.

The Explore and Discuss stages of the teaching cycle in a Practice Understanding lesson are fluid and more individualized. The task to explore may be a worksheet or a game, constrained to facilitate fluency and engaged in by individuals or small groups. The teacher in monitoring the exploration may simultaneously provide individual discussion, giving feedback and helping individuals recognize emerging generalizations, procedures, or models.

Understanding how the CMI Framework impacts teacher instructional practice and subsequently associating CMI-influenced teacher practice with student outcomes cannot be accomplished without measuring the instruction that takes place in the classroom. As has been previously explicated, the CMI Framework provides structure for instruction. It is this structure, with the stated purposes and student and teacher roles, which provides opportunity to create an instrument that will measure instruction through the lens of the CMI Framework.

Instructional Measurement

The following section discusses why measuring instruction is important and what methods have been used in measuring instruction. The strengths and limitations of the methods in measuring instruction will be outlined. The section concludes with a comparison between the CMI Framework's measurement requirements and the discussed methods' strengths.

Reasons for measuring instruction. Literature from the 1960–1990s of the effects of schooling on student achievement often concluded that teachers had less impact on student learning than socio-economic and family factors (i.e.Coleman, et al., 1966; Hanushek, 1981, 1997). However, most often in that research line, teacher characteristics—such as degrees and majors—were used as proxy measures of teacher inputs, (Hanushek, 1997; Hiebert & Grouws, 2007) rather than teaching—what actually happened in the classroom. But what teachers *do*, the instruction that occurs, rather than who teachers *are*, needs to be measured. According to the NCTM:

> Students learn mathematics through the experiences that teachers provide. Thus, students' understanding of mathematics, their ability to use it to solve problems, and their confidence in, and disposition toward, mathematics are all shaped by the teaching they encounter in school. The improvement of mathematics education for all students requires effective mathematics teaching in all classrooms. (p. 135)

Slavin, Lake, and Groff (2009) concurred with the NCTM that teachers' instruction had the greatest effect on mathematics learning. Slavin et al. synthesized 100 randomized or matched control group middle/high school studies and compared effect sizes for curriculum (primarily textbooks), computer assisted instruction, and instructional process programs, which they defined as professional development in effective instructional strategies. They concluded that:

> This review, in agreement with the review of elementary math programs, suggests that in terms of outcomes on traditional measures, such as standardized tests and state

> accountability assessments, *curriculum differences seem to be less consequential than instructional differences*. (emphasis added, p. 886)

In support of the consequential impact of instructional differences, Rowan, Correnti, and Miller (2002) assert that “the classrooms to which students are assigned in a given year can have nontrivial effects on students’ achievement growth in a calendar year” (p. 1532). “Classrooms,” as Rowan et al. term it, is synonymous with “teacher” in the context of their study. For example, Rowan et al. used a cross-classified random effects model to parse out the variance in student achievement into student effects and teacher effects. The researchers determined that a student assigned to a classroom with 1 standard deviation (SD) difference in instructional effectiveness from another classroom, as defined by the random effects model, would differ by 2.13 months in mathematics growth. The term ‘instructional effectiveness’ is used here to describe qualitative differences in teachers’ instruction. Students with better instruction learned more mathematics. Similarly, Rivkin, Hanushek, and Kain (2005) contend through their analysis of longitudinal data from the UTD Texas Schools Project that a 1 SD increase in teacher quality (quality of instruction) would increase student achievement by .11 SD.

Porter (1989) added a different perspective on how instruction impacts student learning when he observed that students receive mathematics instruction at differential rates, depending on the teacher. In the Porter study, teachers reported in interviews and in instructional logs the amount of time they spent teaching various mathematics topics. He asserts that the amount of instruction varies from one-half to twice as much instruction from classroom to classroom. Although quantity is not the same as quality, students’ opportunities to learn are associated with achievement (Hiebert & Grouws, 2007; Porter, 1989). Students whose teachers spent more time in mathematics instruction could arguably have more opportunity to learn and thus the potential for greater achievement.

Certainly these examples of variable instruction point to the need, generally, to determine what students experience in the classroom through instruction if improvements in the quality of instruction are to be enacted. In order to account for the experience students have during instruction, then, it is necessary to define and measure the interactions that occur among teacher, students, and materials in particular environments. Specifically, the CMI Framework suggests a system of instruction which promises to produce improved student mathematical understanding through improved instruction. In order to make the link between the CMI Framework and effective instruction that produces the desired student outcome, classroom instruction must be measured through the CMI Framework lens.

Methods of measuring instruction. Surveys and classroom observation have been two main methods for measuring instruction. There are also several other methods less frequently used or recently being tested: teacher logs, artifact analysis, and scenario response. Surveys have typically been used to measure content coverage and instructional methods that are used. When a look at classroom interactions was desired, classroom observations have been employed. Teacher logs answer questions about typical practice over a specified time. One recently developed measure of instruction is artifact analysis, where student work samples and teacher plans and materials are collected and examined as evidence of instructional practice. Scenario response asks teachers to respond with how they would instruct, given a scenario. Each of the measurement methods has purposes, advantages, and disadvantages.

While in some studies several methods of measuring instruction have been combined to strengthen the study (See for example Camburn & Barnes, 2004; Cohen & Hill, 2000;

Matsumura, et al., 2006; Spillane & Zeuli, 1999), each method will be treated separately for clarity of presentation. The following sections review examples of each measurement method, and summarize the advantages and disadvantages of each.

Survey. Surveys of teachers have perhaps been the most widely used measure of instruction. They are a relatively inexpensive way of obtaining information on a large scale. A broad view of instruction is obtained by asking about trends and frequencies of particular practices, as well as content coverage and time spent (Ball & Rowan, 2004; Desimone, 2009). Surveys of instruction typically ask questions about teachers' practice over the year just elapsed, and rely on teachers' memory and interpretation of terminology used in the survey. Some examples of how surveys have been used to measure instruction follow.

Spillane and Zeuli (1999) surveyed 283 math and science teachers across nine school districts about their awareness and use of reform instructional practices. The respondents included both elementary (3/4 grade) and middle school (7/8 grade) teachers The researchers' interest was in examining elements of mathematics instruction that showed alignment with state and national conceptions of reform mathematics. They chose an existing measure, *Teacher Questionnaire of the Third International Mathematics and Science Study (TIMSS)*, which included questions about both mathematics and science instructional practices. (See Appendix B, Table 11 for sample questions). Only the mathematics results were reported in the article cited. The questions were mapped onto the math reforms recommended by the NCTM to create "a scale of 'reformed' practice" (p. 6). Teachers were also surveyed about the degree to which they were aware of reform recommendations and standards.

Cohen and Hill (2000) surveyed with a focus similar to Spillane and Zeuli (1999). Teachers were asked questions about their familiarity with mathematics reform, their teaching practice, and about their opportunities to learn. The fourteen survey items about teaching practice (see Appendix B, Table 12) loaded onto two separate factors in a factor analysis: framework practice and conventional practice. This factor analysis gave researchers information about how teachers characterized their own instruction. The survey allowed researchers to quantitatively describe and summarize teaching practices as well as correlate practice with teachers' opportunities to learn about reform mathematics.

Blank (2002) undertook a survey study for the purpose of analyzing classroom instruction in an evaluation of systemic initiatives. The survey produced, the Surveys of Enacted Curriculum (SEC), measured instruction multi-dimensionally. The 150 question survey asked teachers to report on 9 areas. Table 13 in Appendix B lists the constructs which formed the scales on the SEC. One example of the multi-dimensional nature of the survey comes from the construct of mathematics and science content in classrooms. The survey listed topics taught and asked teachers to bubble in how much time was spent on the topic and also how much emphasis was given in different types of instructional activities. One of the claims of this study is that the survey afforded both a broad view and a fine grained look at teaching because the survey addressed coverage, time, and instructional strategies as well as student and teacher characteristics and teacher qualifications. This dual-grained view is a need expressed by Ball & Rowan (2004) regarding measuring instruction. However, the "grain" size in this study is arguably still rather coarse.

When appropriately constructed, surveys can be reliable and valid measures of instruction when asking about topics covered, time spent, and types of practices used (Desimone, 2009). Surveys also have the advantage of a low cost to administer and collect data. The three studies cited are representative of instructional aspects adequately measured in surveys. The methods used in the Blank (2002), Cohen and Hill (2000), and Spillane and Zeuli (1999) studies largely address the main criticism leveled at survey methods in the past, that of social desirability biasing the answers teachers give. All three studies addressed two key issues in avoiding bias by ensuring that the answers teachers gave would be confidential and not be used for job evaluation, (Desimone, 2009). Cohen & Hill (2000) also explicitly asked for data in order to form control variables for *affect* and *familiarity* to help detect answers reflecting social desirability more than reality. Spillane and Zeuli (1999), Cohen and Hill (2000), and Blank (2002) also sought to increase reliability of their measures by offering multiple indicators of reform instruction. Measuring one construct with multiple items was another key issue in using survey methodology to measure instruction (Mayer, 1999).

Spillane and Zeuli's (1999) added more validity evidence of to their survey by conducting observations in conjunction with the survey administration. Classroom observations were conducted to corroborate or refute the degree of reform practice reported on their survey. Of the observations that were conducted on twenty-five teachers who reported high levels of reform practice and high familiarity with reform ideals, only four showed practices that were found to match the reform conception of instruction. Eleven showed traditional practices recast in the form of reform practice and ten others had a mix. This mismatch between what teachers reported and what was observed points out a limitation to survey data that Ball and Rowan (2004) mention: "Validation is also lacking because key descriptors of practice used in survey instruments are seldom understood uniformly by respondents" (p. 5). The teachers surveyed thought their instruction reflected the items on the survey, but their idea of the construct did not match the NCTM-based idea of the researchers.

Although surveys are self-reported, the self-report nature of surveys is not entirely troubling, because although teachers do tend to inflate the amount of time spent on topics or certain practices, they do so uniformly so that the relative standing is accurate (Mayer, 1999). However, without carefully defined constructs, respondents may not understand the intent of the key descriptors and answer in ways that do not accurately reflect their practice. Additionally, although the Blank (2002) survey succeeded in creating a more detailed view of instruction than most surveys, it, like the others cited, still was unable to capture the interactions of instruction beyond descriptions of presence and frequency.

Teacher logs. Although not used as frequently as surveys or observation are, teacher logs have been seen as a way to obtain more detailed data for a relatively small investment (Rowan & Correnti, 2009). In many ways, teacher logs are like surveys administered more frequently. Like survey instruments, teachers report their instructional activities by checking types, marking frequencies, and recording time. While less frequently used as a measure of instruction, teacher logs have been seen by some as an improvement over surveys (Rowan & Correnti, 2009) because of the frequency of completion.

Two early studies using teacher logs were conducted by Porter (1989), who aggregated the results into one journal article. While the report is sketchy in detail of methods, it appears that in the first study of seven teachers from six schools in three school districts, the teachers were simply asked to write down what they did in mathematics instruction every day. Once a week, the teacher was interviewed about their log and any ambiguities cleared up for a shared understanding between teacher and researcher of what occurred.

For the second study, a slightly more structured log was provided. The larger sample of 34 teachers from 17 schools across 6 school districts recorded the topic that received primary focus each day. The topics were selected from a catalog of 288 topics and their descriptions provided by the researchers. Each description was written according to a researcher-developed taxonomy with three dimensions: intent of the lesson, nature of the mathematics, and operations required to solve problems. The descriptions enabled teachers to describe their teaching in common ways, largely circumventing the issue of mismatch between definitions of the researcher and respondent. From these data, Porter derived a picture of mathematics instruction regarding topic coverage within and across grade spans, time spent in mathematics instruction, and the amounts emphasis put on skills and conceptual understanding.

The large scale Study of Instructional Improvement (SII) produced several studies using teacher logs which examined both mathematics (Rowan, Harrison, & Hayes, 2004) and language arts instruction (Rowan, Camburn, & Correnti, 2004). Both of these studies improved upon Porter's teacher log by providing a structure for teachers to respond within. The structure provided items in the areas of interest, much as a survey would, calling for checking boxes and limited short answers, and allowing a branching structure. Both the Language Arts and Mathematics logs, for example, had "gateway" questions, which would send the completer to a different section or have them stop, depending on the answer given. The primary difference between a survey or questionnaire and the logs used in these studies was the frequency of administration. Using essentially the same methods, both SII studies examined content coverage and skill difficulty.

Teachers received training in using the log, were given a glossary of definitions and examples for the terms used in the log, and had a toll-free telephone number to call for assistance if they had questions as the logging began. Focal students were selected (eight per class) to create a representative sample in order to account for the differentiated instruction that might occur across students in a class. Each teacher completed 70 logs over the course of the year, approximately 9 per focal student. One example of this log structure, in the Number Concepts (Section A) of the Mathematics Log, is that teachers were asked, what were you using in your work on number concepts today? They were given three choices, whole numbers, decimals, or fractions, with the instruction to "mark all that apply". A sample of the Mathematics Log items appears as Table 14 in Appendix B.

Porter (1989), Rowan, Camburn, and Correnti (2004), and Rowan, Harrison, and Hayes (2004) all used teacher logs in examining patterns of instruction over a school year. The frequency of teacher logs adds a dimension that surveys are unable to attain, that of patterns over time. Although a survey might ask, as the Spillane and Zeuli (1999), Cohen and Hill (2000), and Blank (2002) surveys did, about methods, materials, or topics, and even ask about frequency of use, one cannot get the day-to-day picture about instruction's ebbs and flows from surveys of instruction. Thus, teacher logs are able to capture some of the patterns of interactions that occur, for a slightly higher cost than surveys, but at a substantially lower cost than an equal number of observations.

More recently, Rowan, Jacob, and Correnti (2009) argue that because classroom instruction is multi-dimensional and highly variable across time, that teacher logs are the best way to measure curriculum content and coverage. For them, "instruction is conceptualized as a series of repeated (daily) exposures to instruction, and the key measurement problem is to obtain an estimate of the overall amount or rate of exposure to particular elements of instruction occurring over some fixed interval of time, such as a school year" (Rowan, et al., 2009). While Rowan et al. sought to study "the way in which a teacher interacts with his or her students" (2009, p. 13), the interactions measured were limited (See Rowan, Camburn, et al., 2004;

Rowan, Harrison, et al., 2004). Matsumura, Garnier, Slater, and Boston (2008) recognize that the interactions of teacher and content are well captured in teacher logs, but argue that "measures are needed that focus attention on the interactions between teachers and students" (p. 271).

Observation. Observation has long been considered the "gold standard" of instructional measurement strategies (Ball & Rowan, 2004; Rowan & Correnti, 2009; Stecher, et al., 2006) and can focus attention on the interactions mentioned by Matsumura (Matsumura, et al., 2008). It is also the most proximal way of gathering information about what happens in classrooms in instruction. Van Tassel-Baska describes observational measurement thusly:

> A performance-based assessment of the teacher within the context of the learning environment. . . . it is a relatively open-ended experience, with teachers exercising much control over the selection of the lesson to be taught. It allows for the demonstration of complex and higher-order behaviors, recognizing that good teaching derives from a sophisticated set of skills that unfold in an integrated way. . . . Most importantly, by using a structured form, it provides a benchmark against which the teaching process can be assessed. (2007, pp. 85–86)

A major advantage of observation is that well-trained observers are able to see first-hand what is enacted during instruction, as well as determine differences in deployment of practices of interest. Stecher et al. noted that "quality can be incorporated into observational ratings more easily than into any of the other methods" (2006, p. 121), where quality was identified as the effectiveness with which a practice is used. Defining the criteria, training observers on the criteria, and ensuring high inter-observer agreement avoids the misinterpretations that arise in self-reported measures of instruction where each individual teacher has their own definition of what certain practices mean. Well-trained observers can consistently record the complex and sometimes abstract constructs of instructional interactions.

As much as observation is considered the gold-standard for instructional measurement and has the advantages of proximity and detail, observation is quite costly. Although protocol development has similar costs associated with it as survey development, when a survey has been developed it is relatively inexpensive to deploy. Not so with an observation protocol. In addition to the costs of protocol development, costs are incurred during deployment. Observers must be obtained and paid for their service, which includes not only the time in the classrooms, but time in training and time in completing the protocol after the observation. Expenses are associated with the training materials as well. All of these costs must be considered when selecting a measure of instruction.

The following paragraphs review observational measures of instruction that have been recently used to measure aspects of reform-oriented instruction, in mathematics and in other areas. Both live and video observational measures are included here. While in the classroom, live observations are by far the modal practice for this method of instructional measurement, observation protocols which depend on video reproductions of classroom instruction have been employed more frequently as video technology has developed.

Video observation. The first cross-national study of a representative sample of classrooms using observation through video tapes was undertaken as part of the Third International Mathematics and Science Study (TIMSS: Hiebert, et al., 2005; J.W. Stigler, Gonzales, Kawanka, Knoll, & Serrano, 1999). The researchers were interested in documenting and analyzing typical practice in each of the three countries: Germany, Japan, and the United States. They particularly sought to associate patterns of practice with the achievement of students, which the larger

TIMSS study assessed. Two hundred thirty-one (231) eighth grade mathematics classrooms were video-taped and the videos then used to analyze the instruction students received. In addition to the video, each teacher completed a questionnaire which accomplished two things. One, the questionnaire gave the context of the lesson and demographic information about the teacher and the class. Two, the questionnaire indicated through the context whether the lesson was typical of instruction for the class. One camera was used in videotaping, which focused predominately on the teacher, largely missing student behaviors during instruction. However, both teacher and student voices were recorded with two different microphones, a lavaliere microphone for the teacher, and a boom microphone located on the camera. A coding scheme was developed using a sample of nine lessons from each country in order to allow cross-country comparisons. Guided by the NCTM Professional Standards for Teaching Mathematics and other reform documents, three broad categories were attended to: 1. *The nature of the work environment.* 2. *The nature of the work that students are engaged in.* and 3. *The methods teachers use for engaging students in work.* Within each category, descriptive codes were developed, about 75 in all. The coding protocol was followed for the main sample of taped lessons and the videotapes were analyzed in several passes. Each pass, or re-viewing, focused on a different feature or set of related features of the protocol.

More recently, a study was undertaken that utilized video observation analysis to determine the quality of instruction for the purpose of associating teachers' mathematical knowledge for teaching with their quality of instruction (Hill, et al., 2008). As in the TIMSS study a coding key was developed, this one with 33 items which reflected the 6 elements the researchers had determined from literature and their own prior work to be associated with instructional quality and teachers' mathematical knowledge, (see Appendix B, Table 15).

Lessons were videotaped with what was described as high-quality equipment that allowed a flexible view of the classroom and audio pick-up of both students and teacher. For analysis, the videos were chunked into five-minute segments, which served much the same purpose as the multiple passes employed in the TIMSS study. Each five-minute segment was coded according to the items corresponding to the elements of instruction which were used as a framework. The items were rated as to their presence or absence and as to appropriate or inappropriate demonstration of the item.

There are a few advantages to video observation. Video observation affords researchers or evaluators review of the lesson multiple times, thus increasing the ability to bring a particular aspect of the lesson into sharp focus. Even though the complexity of a live classroom is depicted in the video, much can be ignored during a pass (Stigler, et al., 1999). It is also convenient to replay a segment in order to more closely decide what of interest is occurring. Another advantage of video observation is the ability to have multiple viewers rating the lesson. In both of the video studies cited, groups of researchers viewed at least a subset of the videos, particularly in the code development stage. In the TIMSS study, a group of six mathematicians viewed a subset for analysis of the mathematics represented. Groups of "visitors" in a classroom would be quite disruptive, and likely would change the nature of what occurs, invalidating the findings of the observation.

Although video observation has some attractive advantages for measuring instruction, there are at least two disadvantages. One disadvantage is simply the set-up and equipment required. The equipment requires time prior to the observation to set up and adds visible reminders that something a bit out of the ordinary instructional period is occurring. Recording has the potential to trigger atypical behaviors or instruction. Stigler et al. (1999) put several measures in place to detect this type of threat to validity. Questionnaire items asked directly how typical the lesson was in structure and student behavior, and indirectly, by asking where the lesson fit in a sequence. Teachers were also asked to rate how nervous they felt being observed. The Hill et al. study does not mention any control for this threat.

A second limitation is that the observation is limited by the view a camera, or cameras, can capture. The limitation of camera view is very similar to the limitation of view that observers in a live observation might have. While the video observation view is limited to what the camera lens can "see", the live observation view is limited by how much the observer can attend to at once. Both studies used one camera, with two microphones, one worn by the teacher, and a camera-mounted boom microphone to capture student voices. The Stigler et al. study explicitly details the camera-use rules that were developed, and the videographer training that occurred to minimize bias created by limited view observers have through the camera lens.

Live observation. Notwithstanding the cost of observation, studies utilizing live classroom observation are numerous. Because of the complexity of instruction, nearly all select a perspective from which to view instruction and operationalize that perspective into a protocol. The two protocols selected for review here, the Reform Teaching Observation Protocol (RTOP) and the Inside the Classroom Observation Protocol, attend particularly to the interactive nature of instruction. Of note is that the RTOP arose from an organization with goals similar to the Mathematics Initiative Committee's (MIC) goals which led to the development of the CMI Framework. Because the goals in creating the system of instruction measured by the RTOP were similar to the goals of the MIC, looking at the RTOP gave insight into the type of measure that the CMI Framework required. The second protocol to be reviewed, the Inside the Classroom Observation Protocol, was the protocol that had recently been used by the CMI Framework developers, with some modifications, to measure instruction. It was selected by the CMI Framework developers because of its compatible view of instruction with the CMI Framework. While a useful measure, the Inside the Classroom Observation Protocol has not fully captured the instructional interactions embodied in the CMI Framework; it is reviewed here to briefly assess its strengths and shortcomings, although it will be more thoroughly evaluated as part of protocol development for the proposed CMI Observation Protocol.

The RTOP was designed to capture the Arizona Collaborative for Excellence in the Preparation of Teachers (ACEPT) researchers' idea of reform instruction in mathematics and science. The conception of reform articulated in the RTOP was arrived at after reviewing the standards and principles published by national science and mathematics organizations, and based on the ACEPT group's experience. According to the ACEPT view, reform is described as follows:

> A movement away from the traditional didactic practice toward constructivism, . . . students using data to justify opinions, experiencing ambiguity as a result of learning, and learning from one another. Additionally, reform presupposes that teachers do not emphasize lecture, but rather stress a problem-solving approach and foster active learning. (Sawada, et al., 2002, p. 246)

In the opinion of the ACEPT group, no existing observation protocol measures of teaching exclusively measured "the reformed nature of the classroom—all had other components reflective of "good" teaching more generally such as "lesson closure" or adequate "wait time" (Piburn & Sawada, 2000, p. 45). The researchers sought to measure the impact of reformed teaching on student achievement, particularly in college and high school classrooms, although the training materials states that the RTOP may be used in all levels, from early elementary through University. The RTOP measured three domains: Lesson design and implementation, Content, and Classroom culture. The RTOP domains and measured items are listed in Table 16 in Appendix B.

Teachers were rated on each item on a 0–4 scale, with anchor points at 0 (Never occurred) and 4 (Very descriptive). According to the training manual, "Intermediate ratings do not reflect the number of times an item occurred, but rather the degree to which that item was *characteristic* of the lesson observed. Possible scores range from 0 to 100 points, with higher scores reflecting a greater degree of reform" (Sawada, et al., 2000, p. 2).

The Inside the Classroom Observation Protocol (Weiss, et al., 2003) was developed in order to get a more detailed view and an outsider perspective on modal practice in U.S. mathematics and science classrooms. It had a more open view of instruction compared to the RTOP, although still tied to the student outcome of understanding. By "open" it is meant that the Inside the Classroom Observation Protocol developers allowed that many methods of instruction may lead to student understanding if they were centered on activities that were "purposeful, accessible, and engaging to students, with a clear and consistent focus on student learning of important mathematics and science concepts" (Weiss, et al., 2003, p. 25). This open view is in contrast to the RTOP which featured decidedly "reform" instructional strategies for measurable items. So although a "reform" view of science and mathematics outcomes guided the development of the Inside the Classroom Observation Protocol, no list of "reform practices" appears in the protocol.

The Inside the Classroom Observation Protocol contains four sub-scales, each with a number of *key indicators*, a synthesis rating, and space for recording supporting evidence. Each key indicator is rated on a scale from 1 (Not at all) to 5 (To a great extent), indicating the degree to which the indicator was descriptive of the lesson. Two additional rating categories were available, 6 (Don't know) or 7 (N/A), used to indicate either if the observer didn't feel there was enough evidence to judge or if the item was not applicable given the purpose and context of the lesson. The synthesis rating is not a simple average of all the key indicator ratings, but a placement on a scale of 1–5 indicating the degree to which the sub-scale indicators taken as a whole reflect best practice in mathematics or science education. A 1 indicates "Not at all reflective", while a 5 indicates "Extremely reflective" of best practice (see Appendix B, Table 17 for sub-scales and key indicators). In addition, there is an impact rating and a capsule rating included in the measure. The impact rating asks the observer to judge the likelihood of the lesson to move students toward understanding. All of the sub-scale synthesis ratings and the impact rating contribute to a capsule rating of quality (see Appendix B, Table 18). Again, the capsule rating is not an average, but the observer's judgment of how closely the overall lesson matches the quality descriptors.

Both the RTOP and the Inside the Classroom Observation Protocol are examples of evaluative measures of instruction. The items not only show whether the instructional element is present or absent, but also assigns a rating of how representative the item is throughout the lesson. The RTOP totals the item ratings to give an overall score of degree of reformed instruction. Similar to the RTOP in that an overall score is given, the Inside the Classroom Observation Protocol gives a 'capsule rating.' Differing in how the overall score is derived, the Inside the Classroom Observation Protocol assigns a holistic capsule rating, representing the degree to which the lesson reflects effective mathematics or science instruction.

Artifact collection. In measuring instruction via artifact collection, "researchers typically ask teachers to collect and annotate a set of materials, such as classroom exercises, homework, quizzes, projects, exams, and samples of student work" (Borko, et al., 2007, p. 9). The purpose of artifact collection is to provide evidence of the planning and the enactment of instruction. Two examples of measurement by artifact collection come from Borko, Kuffner, Arnold, Creighton, Stecher, Martinez, Barnes et al. (2007) and Silver, Mesa, Morris, Star, & Benken (2009).

The Borko et al. study used an extensive *scoop* method of artifact collection which included teacher reflection and photographs of the environment as well as the more typical annotated materials. The researchers defined scoop as a "one-week process in which teachers collect artifacts of instructional practice (e.g., lesson plans, instructional materials, student work), take photographs of the classroom set-up and learning materials, write responses to reflective questions, and assemble the results in a three-ring notebook" (Borko, et al.,2007, p. 2). The analysis of instruction focused on ten elements of reform instruction researchers felt were amenable to detection by way of artifact collection (see Appendix B, Table 19). A scoring guide was constructed that rated each element on a scale of 1–5, with classroom examples of low (1), medium (3), and high (5) performance on each element.

The analysis of artifacts from this study (Borko, et al., 2007) illustrates that in some respects artifact collection is an improvement over surveys and teacher logs, and in other ways suffers from similar shortcomings. The addition of artifacts allowed researchers to find evidence of both the intention and the enactment of instruction in lesson design, in student work, in photographs, and in teacher reflection. The scoop method had the advantage over other self-report methods in that raters had artifacts, including annotations and reflections, that could make clear the teacher's definition of particular constructs, and also determine through student work samples and assessments the teacher's expectation for mathematical or science understanding.

Although providing concrete evidence through the artifacts was helpful in constructing a picture of instruction, and the method was generally reliable and yielded valid interpretations, the researchers concede that even with extensive training, raters differed in interpretation of evidence when definitions of key constructs were multi-dimensional. As the rating procedure was refined, many of the definitions became simplified, allowing greater inter-rater agreement, but reducing the clarity of the construct. For example, when the “grouping” definition included the nature of work done in groups (i.e. collaborative, conceptual, substantive tasks) there was great variation in ratings. When the definition was simplified to rate the presence or absence of grouping, raters were able to easily come to agreement but at the cost of loss of detail. The artifacts and narratives did not always give enough information for raters to consistently discern the details of instruction and raters’ interpretations could introduce measurement error just as easily as the mismatch between researcher understanding and teacher reports on surveys or logs.

Silver et al. (2009) took advantage of artifacts collected in teacher portfolios submitted to the National Board for Professional Teaching Standards (NBPTS) from a random sample of thirty-two teachers seeking national certification. Silver et al. also compared their findings to the findings reported in a large-scale national survey of instruction in order to help interpret their results. The comparison between the artifacts collected and the national survey makes the Silver et al. study interesting for this dissertation.

Silver and colleagues created a framework for analysis from the National Assessment of Educational Progress (NAEP) topic categories (topics), a compilation of frameworks regarding tasks and cognitive demand (tasks), and literature (pedagogy; see Appendix B, Table 20). Three raters rated each lesson independently using the framework created. From the data collected through the portfolio artifacts, Silver and colleagues determined patterns of topic coverage, cognitive demand, and pedagogical features of lessons from the sample teachers' best work. As the portfolio data were compared to the findings of a national survey of a representative sample of teachers, some differences in modal practice were noted, several of which were attributed to differences in teachers' definitions. An example of the definitional difference is that the national survey data indicated a very large proportion of teachers required student explanations of work, while the portfolio sample showed much less use of this reform strategy. Given that the sample in the Silver et al. (2009) study were teachers seeking recognition as high quality teachers and submitted the best examples of their work, one would expect similar or higher proportions of this practice when compared to a national representative sample.

The authors' hypothesize that while their analysis of artifacts required mathematical justification to count as explanation, teachers answering the survey consider listing steps to a solution as explanation, a very different definition. This finding underscores the value of artifact collection as an instructional measure. Unlike a survey where researcher and respondent definitions may differ, but that difference is not known, an artifact can make the respondent's definition visible. However, as was found in the Borko et al. (2007) study, raters may still disagree on what the artifact represents when the construct definition is multi-dimensional.

One of the reasons artifact analysis is considered a promising instructional measurement strategy is that it is a hybrid of direct observation and survey (Silver, et al., 2009). Silver and colleagues claim that like observation, one may see the details of instruction without the cost, intrusiveness, or labor intensity. They also claim that, like survey measures, the teacher's own perspective is evident without the problems of misinterpretation of questions, questions of validity, and lack of detail. However, Borko et al. observe that "artifacts were more informative about structural features such as use of mathematical tools or scientific resources, and less informative about interactive aspects of instruction such as patterns of discourse and the nature of explanations" (2007, p. 67).

A drawback not noted directly in the literature is the respondent burden created by the collection of artifacts. The Borko et al. (2007) study asked for teachers to compile the artifacts and include annotations and journal entries, which goes substantially beyond what might be asked of teachers by a survey, teacher log, or observation. The Silver et al. (2009) artifacts made use of work required for teachers to submit as their application for board certification, perhaps somewhat lessening the perception of burden to the teachers. The teachers collected their portfolio artifacts in order to gain the certification they desired. Borko et al. noted that some of the teacher written annotations or explanations collected as artifacts in their study lacked sufficient quality to be useful, and suggested that artifact collection might be more useful "in situations where teachers are personally invested in providing the most complete and detailed information possible about their classroom practices" (p. 62). The Silver et al. study accommodated that need for personal investment.

Scenario response. A final type of instructional measure, scenario response, also called vignette response, is briefly treated here in order to complete the discussion of tools that might be used to measure instruction. Scenario response has been least used of the measures being discussed, perhaps because it measures only in the hypothetical. An early example sets the purpose of scenario response as "designed to uncover both *what* they [teachers] think about and *how* they think" (McDiarmid & Ball, 1989, p. 13). In response to constructed scenarios or vignettes, teachers put themselves in the scenario and describe what they would do as a teacher. Scenario response may occur during interviews, as in the McDiarmid and Ball study, or as a written response to print, oral, or video vignettes. Answer methods may be open or closed.

Open methods simply pose the scenario and allow the respondent to freely address what they believe they would do, using follow-up questions if necessary. Closed methods ask respondents to select from possible responses. The McDiarmid and Ball study used open answers, presenting the teacher with the scenario, recording the answer, and using probing follow-up questions to more fully understand the response.

More recently Stecher, Le, Hamilton, Ryan, Robyn, and Lockwood (2006) undertook a study to measure reform-oriented instruction using vignettes with closed responses. Teachers read the vignette and eight responses. Teachers rated each response on a 4-point Likert-type scale indicating how likely it would be for them to respond to the teaching opportunity in that manner. Scenario response's main strength is that it can approximate teacher instructional behaviors for low-incidence events that may not be captured by other methods. Scenario response also gives insight into teachers' dispositions to respond to situations. A significant short-coming of scenario response is there is not yet evidence to corroborate what teachers really do in practice when confronted by the same scenario they responded to in the hypothetical.

The CMI Framework in relation to instructional measures. The CMI Framework is a system of instructional interactions based on principles (worthwhile tasks, classroom discourse, embedded assessment, using student thinking, and the relationships between lessons, units, and curriculum). The desired instructional interactions differ according to where in a learning cycle/ teaching cycle combination they occur. CMI Framework developers are interested in measuring whether teachers' classroom mathematics instruction matches the CMI Framework's conception of instruction, and to evaluate the degree to which individual teachers' instruction embodies effective and appropriate use of the core principles in the instructional interactions. Looking at the match between the CMI Framework and teacher instruction will give information about the

CMI Framework's effect on instruction, which is an important link in making the claim that the CMI Framework helps translate theory into practice. Evaluating the degree to which different teachers are able to effectively and appropriately use the CMI Framework's core principles in instructional interactions helps make the association between CMI Framework instruction and student understanding possible. A measure of instruction according to the CMI Framework needs to be able to record the interactions in ways that acknowledge the different purposes of lessons in different phase combinations of the learning/teaching cycle. Additionally, a CMI Framework-driven measure needs to include ways to describe levels of use of the core set of principles during the instructional interactions, as well as evaluate whether the use is appropriate.

Self-report methods. Self-reported measures such as surveys, teacher logs, and scenario response, could be constructed so as to record many of the interactions regarding the principles that the CMI Framework espouses. Although surveys tend to be used to give a broad retrospective look at instruction, one could be constructed to ask teachers to respond in regards to a particular lesson. Logs, by definition, record by considering individual lessons. Teachers could be asked to respond to scenarios that would offer opportunities to describe instruction that matches the CMI Framework. However, even with all of the safeguards that may be employed to encourage accurate reporting, self-reported measures do not seem appropriate for the CMI Framework measurement purposes for the reasons discussed in the next paragraph.

Although the aim of this dissertation is solely to develop and validate a measure of instruction, part of the purpose of measuring instruction through the CMI Framework lens is to be able to link levels of CMI Framework practice with student outcomes. Making a link between level of CMI Framework practice and student outcomes relies on consistent and accurate evaluations. Given that the literature shows little shared understanding of definitions and a social desirability bias in teacher self-reports (Ball & Rowan, 2004; Mayer, 1999), self-reported information likely would not be accurate. Different teachers reporting their practice answer self-report measures according to their personal interpretation of the definition of the practice.

Even with extensive support as was given in the SII studies (Camburn & Barnes, 2004; Rowan, Camburn, et al., 2004; Rowan, Harrison, et al., 2004), each teacher is still prone to personal bias. In fact, Camburn and Barnes (2004) found through triangulation between teacher logs kept by the teacher and two observers that when the teacher's entry differed from the observers' entries, it was because the observers used the glossary of terms to guide them, while the teacher used their own context and experience. Teachers who employ the same instructional practices may answer items differently due to their interpretations, introducing inconsistency into the level ratings. Additionally, although Mayer (1999) showed that the relative rankings of teachers regarding practices reported through survey data remained consistent, the inflation of practices and interactions measured for CMI purposes would lead to inaccuracy in the levels of practice. While self-report methods of instructional measurement would give insight into patterns of practice and into teachers' perceptions that would be valuable additions to learning about how the CMI Framework impacts instruction, they do not allow the consistency and accuracy needed to correlate the CMI Framework with student outcomes.

Outsider evaluation. Observation and artifact collection are the two methods that do not rest on self-report for the evaluation of instruction. While it is true that artifacts are offered by the teachers and that self-selection has the potential to bias what is evaluated, the Borko et al. (2007) study carefully prescribed what artifacts were to be collected and delineated a time frame in which the daily collection occurred. The study prescriptions made it difficult to "pick and choose," thus minimizing the bias potential. For this reason, artifact collection is included here.

Both methods, observation and artifact collection, rely upon outside evaluators or observers with a very high degree of shared understanding of the important constructs being measured, creating consistency and accuracy in the measurement. As already indicated, consistency across teachers in describing what CMI principles are evident in instructional interactions is an important factor in determining the CMI Framework's effect on instruction. The accuracy of evaluating the level or degree of CMI Framework practice is critical for examining the association between CMI Framework-driven instruction and the student outcome of mathematical understanding.

Artifact collections "have promise for providing accurate representations of selected aspects of classroom practice" (Borko, et al., 2007, p. 4). Artifact collection could offer much information about the CMI principles contained in a lesson, particularly about tasks, embedded assessments, and the relationships among lesson, unit, and curriculum. Evidence of both how tasks and assessments were planned for and how they were enacted could be available through artifact collection. However, according to Borko et al., (2007) informal assessments, the type most embodied in the "embedded assessment" principle of CMI, were difficult to record for their "scoop" notebook and were most easily captured by classroom observation. Narratives about how the particular lesson was related to other lessons, units, and curricula could be collected, adding insight into teacher perspectives on the CMI principle of how the part relates to the whole.

The principles of discourse and use of student thinking would be revealed in a more limited way, primarily restricted to intent rather than enactment. Collected lesson plans could show planned discussions and pre-planned questions to elicit student thinking. However, how the discussion unfolded and how questions changed to capitalize on student thinking would be less likely to be represented, even in retrospective notes that might be collected. The moments go by quickly and are difficult to recapture. Borko et al. note (2007), referring to teacher reflection questions about how the lesson unfolded, “The modal responses were, essentially, that the lesson unfolded as planned, that students learned what was expected, and that no changes were planned for the next day” (2007, p. 51). The teacher reflections added a self-reported artifact, and as outlined previously, self-reported information is biased by teacher definitions and may be biased by social desirability. Teachers and researchers from the Borko et al. study also commented that artifact collections can only show the outline of instruction, but not the interactions of classroom discourse. While artifact collection has the potential to add to an evaluation of teacher instruction, as the only measure for the CMI Framework it has limitations because of the difficulty of measuring discourse, use of student thinking, and embedded assessment.

Each method already discussed could measure aspects of CMI Framework instruction adequately. The SII studies (Camburn & Barnes, 2004; B. Rowan, Camburn, et al., 2004; B. Rowan, Harrison, et al., 2004) have also shown that a combination of methods may more completely represent instruction. Multiple measures eventually may be desirable for the CMI Framework. However, to meet the immediate purposes of measurement which are to determine the match between classroom instruction and the CMI Framework and to assign levels of practice based on the match, observation was selected as the measurement method. Observation was selected because of the potential for yielding consistent and accurate views of the interactions surrounding CMI Framework principles. Self-reported measures are unable to attain the consistency and accuracy required. Observers, guided by a well specified protocol and with adequate training, were able to record the interactions necessary with a high degree of consistency across teachers and with accuracy. Although video observation has the advantage of multiple viewings by observers, allowing observers the luxury of time to thoroughly consider their ratings, the disadvantages of cost, intrusion, and the limited camera view made live observation more feasible for this study. Live observations give observers the control over where to place their focus. While observers may miss some interactions as they make notes on others, they are still present and can select where and when to attend, increasing the likelihood of measuring CMI Framework instruction with consistency and accuracy.

Validity

Two other key issues in measuring instruction are the reliability of the scores obtained from the instrument and the valid interpretation of those scores. Any instrument has an intended purpose and the interpretation of the scores or ratings obtained will indicate something about the purpose. Evidence needs to be gathered to show that the instrument produces accurate interpretations of the purported constructs and that the scores mean the same thing each time the instrument is administered. The proposed CMI Observation Protocol, for example, would have as its purpose measuring teacher instructional enactment of the CMI Framework. The ratings obtained should be interpreted as the degree of Framework practice in evidence by a particular teacher, during that instructional episode.

Validity has been redefined for educational and psychological testing in the past decade as a unitary concept containing five categories of evidence (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The five categories of evidence are labeled content, response processes, internal structure, relations to other variables, and consequences evidence. Evidence from only one of the categories is considered weak evidence, while multiple sources of evidence strengthen the claim for reliable ratings and valid interpretations (American Educational Research Association, et al., 1999).

This dissertation subscribes to this unitary concept defined by the AERA et al., and the idea that evidence from multiple categories strengthens claims for reliability and validity. Three categories of evidence were selected for use in this study: content, response process, and internal structure evidence. Content evidence and response process evidence were selected because of their foundational nature. In beginning protocol development, knowing that the instrument represents the content domain well is fundamental. Response process evidence engenders trust that steps have been taken to minimize measurement error. Being able to trust the ratings is critical so that any inferences made from the data are meaningful. Unless there is content and response process evidence there is no assurance that there is any importance or validity to the measure. Although internal structure evidence is secondary, it is important to theory development. A domain can be measured without knowing the internal structure. However, finding an internal structure allows one to begin to describe the constructs being measured. Describing the constructs is part of developing a theory. The categories of evidence not represented in this study were thought by the researcher to be better left for future research. While they are valuable, they are not necessarily of equal importance when in the beginning stages of protocol and theory development.

Content evidence. Content evidence for validity refers to how well the content "represents a specified content domain" (Goodwin & Leech, 2003). In measures of instruction, the content domain would be the theoretical lens through which instruction is viewed, and the accompanying framework. Content evidence is concerned that the items, the wording of the items, and the format of the measure, are clear, are relevant, and describe the constructs to be measured. Evidence may be obtained by expert review. The evidence from expert review would consist of justification from theory and experience that the items actually represent the whole of the constructs.

Response processes evidence. Response processes evidence for validity entails gathering evidence about how items are answered or scored. In the case of observers, it "includes the extent to which the processes of observers . . . are consistent with the intended interpretation of scores" (American Educational Research Association, et al., 1999, p. 13). According to Goodwin and Leech, assurance needs to be given that observers "are applying the criteria as intended and not using irrelevant or extraneous factors" (p. 184). Reporting inter-rater reliability and observer training procedures are two ways of showing evidence of response processes validity. Inter-rater reliability is a measure of how well raters agree with each other as they score observations. It is an assurance that when one rater gives a score, it is consistent with the score another rater would give on that particular observation. Reporting observer training procedures gives further evidence about observers' ability to be consistent through exposing the opportunity observers had to understand the constructs and the scoring.

Internal structure evidence. Internal structure evidence concerns how the items of a measure interrelate and refers to how strongly the items of a measure are associated with the underlying constructs. Validity evidence of this type requires that there is a strong correlation between the items measuring a single construct, and a weak correlation between those same items and all other constructs. Factor analysis is a commonly used method for supporting a claim of internal structure evidence. D-studies (from generalizability theory) also provide evidence for the validity of the internal structure of an instrument.

In instructional measures, whether survey, teacher logs, scenarios, artifacts or observations, several sources of validity evidence have been used. Measures of instruction, when discussions of validity are reported, most often report content evidence using experts to validate items, response process evidence with inter-rater reliability percentages, or internal structure in the form of internal consistency reliability estimates. In fact, when Kilday and Kinzie (2009) reviewed the methods for determining validity of nine observation protocols measuring some aspect of teaching they recounted that five of the studies reporting on validity gave inter-rater reliability statistics, and three reported internal consistency indicators. It should be noted that Kilday and Kinzie may have underreported validity evidence, since in the case of the Inside the Classroom Observation Protocol, they missed content evidence of validity through expert development of items, as well as omitting information from a 2005 study by the developers of the Inside the Classroom Observation Protocol that provided evidence that the predictive value fell short of expectations (Banilower, 2005).

Research Questions

In addition to developing the CMI Observation Protocol, this dissertation gathered evidence for validity and reliability. The questions of interest were as follows:

1. To what extent can the ratings produced on the CMI Observation Protocol by observed classroom instruction be validly interpreted?
    a. What is the content evidence of validity?
    b. What is the response process evidence of validity? (i.e. inter-rater reliability)
    c. What is the internal structure evidence of validity? (i.e. construct validity)

2. Does the newly developed CMI Observation Protocol reliably measure classroom instruction along the dimensions of the CMI Framework revealed by the data? (i.e. internal consistency reliability)

Chapter 3: Method

This dissertation undertook the development and validation of a new observation protocol for measuring instruction according to the Comprehensive Mathematics Instruction (CMI) Framework. Protocol development and validation methods have aspects that are simultaneous and aspects that are recursive as well as those that are linear. While the researcher acknowledges the non-linear aspects of protocol development and validation, this method section is partitioned into development methods and validation methods as an organizing tool. Where simultaneous or recursive aspects of development and validation methods occur, these are explicitly noted.

Protocol Development

Protocol development entails making decisions about what to measure, what scale to use in measuring, and what structure to create that will facilitate measurement. In determining what to measure, Corey (2007) suggests that a framework is necessary to guide the development of items to be included on a measure of instruction. In this case, the CMI Framework is the lens through which instruction was viewed and therefore the guide to developing items that measure instruction. The CMI Framework embeds three-phase teaching cycles in a three-phase learning cycle to guide interactive, systemic instruction about important mathematics. The scale developed for quantifying the items depends on what information is desired from the measurement. As seen previously in the review of literature, scales may be dichotomous, simply measuring the presence or absence of the item (Hill, et al., 2008); scales may be constructed to measure amounts (Stigler, et al., 1999); or scales may indicate degree (Sawada, et al., 2000; Weiss, et al., 2003). When the questions about what to measure and on what scale are answered, the items need to be organized into a structure that allows for observer understanding and completion. In this dissertation no expectation existed of testing to see if the scale selected or the structure chosen were the “best” ones, as testing was beyond the scope of this dissertation work. The scale and structure were selected based on literature about other, similar protocols and the knowledge of an expert panel. The methods for development of this new protocol, the CMI Observation Protocol, are detailed in the following paragraphs.

Assemble CMI expert panel. The development of a new instrument is necessarily a collaborative effort. To that end, an expert panel was assembled to consult with during development. The CMI expert panel was composed of three university professors who were key in the development of the CMI Framework, two individuals who had observed and evaluated instruction in hundreds of classrooms, one principal, and one district math specialist. The observers, the principal, and the district math specialist all have also been teachers of mathematics. All of the members of the CMI expert panel are currently or have been members of the Math Initiative Committee (MIC). The composition of the MIC has purposively drawn upon multiple perspectives to guide CMI work. Current classroom teacher input was not sought for during this first protocol iteration because beginning development required intimate knowledge of the CMI Framework that had not yet developed among the teacher population. The nature of the CMI expert panel’s involvement will be specified in the appropriate sub-sections of this method section.

Conduct information gathering. Two information gathering methods were used in preparation for writing items for the new CMI Observation Protocol: 1) document examination and 2) interviews. Detailed examination of the CMI Framework and the modified Inside the Classroom Observation Protocol (Horizon Research Inc., 2000) played a critical role in the development of the CMI Observation Protocol. The purpose of examining the CMI Framework was to identify what instruction entailed if conducted according to the framework. It was also important to identify the key indicators on the modified *Inside the Classroom Observation Protocol* that mapped well onto the CMI Framework, and to clarify the aspects of the structure that had been problematic as well as useful. In their recent professional development work, the developers of the CMI Framework used a slightly modified version of the *Inside the Classroom Observation Protocol* to measure instruction. While the protocol has been useful, the *Inside the Classroom Observation Protocol* was developed based on a conception of reform instruction that is not identical to the CMI Framework. In developing the proposed CMI Observation Protocol it seemed prudent to build on the strengths and eliminate the weaknesses of the *Inside the Classroom Observation Protocol* so as not to "reinvent the wheel," but to develop a measure of instruction with valid and reliable results in measuring the CMI Framework.

As part of the *Inside the Classroom Observation Protocol* examination, opinions were sought from selected members of the CMI expert panel through interviews. An interview was conducted with each of the two CMI observers with experience regarding the strengths and weaknesses of the *Inside the Classroom Observation Protocol.* The observer interviews were digitally recorded for later reference.

Create the CMI Observation Protocol. Item writing, structure creation, and scale determination were an iterative and recursive process. Items were written by the researcher, organized into a structure, and the rating scales were affixed. A member of the CMI expert panel critiqued the developing protocol at each iteration. When a satisfactory draft was ready, a focus group was held with the CMI expert panel to solicit their expert opinions regarding how well the *CMI Observation Protocol*'s items represented the instruction promoted by the CMI Framework (see Appendix C for the focus group protocol). The focus group was digitally recorded. The protocol was improved and electronically transmitted to the expert panel members, who then responded to the changes by telephone or email. The "final" version of the *CMI Observation Protocol* was used in 12 classrooms as a pilot test regarding its ease of use in live observations. There were no modifications made as a result of the pilot.

Protocol Validation

The CMI Observation Protocol was developed to be used in classrooms by observers observing mathematics instruction in real time. In order to gather the multiple types of validity evidence required in a validation study, the newly developed CMI Observation Protocol was used to collect data in classrooms.

Sample. The classrooms selected for this study came from the pool of elementary schools which have been associated with the BYU-Public School Partnership. Principals were contacted and asked for permission to conduct observations in classrooms. Those principals who agreed to classroom observations had various approaches for enlisting their teachers. Some principals volunteered their entire faculty, while the majority of principals allowed teachers to opt in if they wished. Eleven schools provided classrooms for observations. The participating schools came from five school districts along the Wasatch Front and may be considered suburban schools. Five of the 11 schools were classified as Title I schools. Each of the teachers whose instruction was observed signed an informed consent to be observed for this study. In total, 144 classroom observations were made at the 11 schools. Figure 3 shows the composition of grade levels represented in the sample.

Observers. Observers were recruited and hired who had a mathematics background and experience with school mathematics. The observers for this study were graduate students from the mathematics education department of the College of Mathematics at Brigham Young University. Recruitment consisted of emailing a flyer to all graduate students from the department, personally recruiting in a class attended by most eligible candidates, and interviewing the six who responded to both appeals. Three observers were selected from this process. The successful observer candidates were selected primarily because they demonstrated knowledge of reform mathematics principles consistent with CMI instruction. A secondary consideration in selection was the candidate's availability for fitting an observation schedule. The researcher also served as an observer to maximize the number of observations that could be accomplished, given the schedules of availability for the graduate student observers.

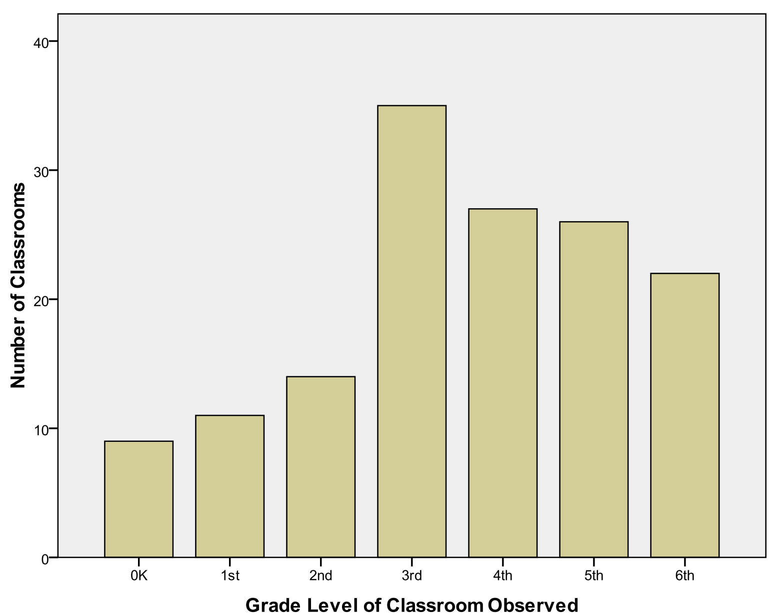

Figure 3. Grade level distribution of CMI Observations

Observer training. Observer training consisted of 3 hours of theoretical training and 9 hours of field training. The theoretical training took place in a university classroom. Observers became familiar with the CMI Framework through discussion and activities. They also worked through the CMI Observation Protocol, talking about and writing down the definitions of items, the item scale, the meaning of synthesis and capsule ratings, and discussing scenarios that could arise. A brief training pamphlet was given to observers, consisting of the procedures observers should follow, space to create the glossary of definitions, and contact information for the study personnel.

The field training was conducted at a nearby elementary school over a two day period. All four observers viewed four lessons in the same classrooms. After each observation the observers met and discussed the lesson ratings for each item, synthesis rating, and capsule rating.

Analysis methods. Several quantitative methods were used to establish the validity and reliability of the CMI Observation Protocol. Inter-observer reliability, a component of validity evidence, was established by computing the intra-class correlation coefficient (ICC). In order to determine a reasonable internal structure for the protocol, an exploratory factor analysis (EFA) was completed using SPSS software with the data collected during the classroom observations. EFA is a statistical method that uses the covariance of the data to describe similarities in variability among the observed variables in terms of an underlying factor. EFA "groups" together observed variables that seem to be related to the same underlying factor. The EFA was a necessary preliminary step prior to testing the internal structure using confirmatory factor analysis (CFA). CFA is a statistical method whereby a model of relationships is proposed and tested among the observed variables and the factors. After determining a factor structure with the EFA, each construct that was revealed was tested for unidimensionality as a single factor model using AMOS software (Arbuckle, 2008). A confirmatory factor analysis (CFA) was also conducted for the all of the constructs together in a first-order model. Another piece of internal structure validity evidence is internal consistency reliability. Do the items purported to measure the same construct do so in a consistent way? Cronbach's alpha is the typical measure for reliability and was used in this study. Raykov's Rho was also used since Cronbach's alpha may under- or overestimate internal consistency reliability when errors are correlated, which was expected in this study.

Method Summary

In summary, the development and validation of the CMI Observation Protocol was an iterative process. The process used experts in the CMI Framework and former observers, which together comprised the CMI Expert panel, to collaborate with the researcher to examine the CMI Framework and the currently used modified *Inside the Classroom Observation Protocol*. Structure, measurement scale, and items were developed and modified during development, based on the inputs of the CMI expert panel and literature. A small pilot test was held in twelve classrooms by members of the expert panel. Observers were trained in using the CMI Observation Protocol through discussion and classroom experiences. The intraclass correlation coefficient (ICC) was calculated to ensure inter-rater reliability was sufficient before field testing in 144 classrooms. Field-testing occurred in live classroom observations with trained observers using the CMI Observation Protocol. Exploratory factor analysis (EFA) was used to determine a reasonable factor structure of the protocol, and reliability was calculated using Cronbach's Alpha and Raykov's Rho. These methods sought to guide the development of a new measure of instruction, the Comprehensive Mathematics Instruction Observation Protocol, to gather validity evidence, and to estimate reliability of the ratings.

Chapter 4: Results

This study developed an observation protocol aligned to the Comprehensive Mathematics Instruction (CMI) Framework. The developed protocol, the Comprehensive Mathematics Instruction (CMI) Observation Protocol, was then used in 144 elementary mathematics classrooms to collect data to provide validity evidence in three areas: content, response process, and internal structure. Additionally, internal consistency reliability was calculated. Although the protocol development was not associated with a research question, the results of development activities will be reported in this section along with the results which will answer the questions regarding validity of the protocol.

Development Results

This section details the results of implementing the development methods outlined in Chapter 3 and describes the completed protocol. As described in the methods section, protocol development entailed document examinations, expert panel member interviews, focus groups, and consultations. All of these activities resulted in information that was used to create the CMI Observation Protocol.

Results of development activities. Document examination found that every learning cycle contained a) a task, b) students engaging in the task for the purpose of making meaning and connections, c) discourse between and among students and teacher, and d) teacher anticipation of and use of student thinking. The differences among learning cycles came in the openness or constraint of tasks, and in the type of scaffolding provided (see Appendix C, Framework Examination Matrix). The CMI Observation Protocol needed to measure all of the dimensions through all of the learning cycle phases.

The results from examining the Inside the Classroom protocol determined that a majority of the 33 items of the Inside the Classroom protocol had at least a weak connection to the CMI Framework principles. However, more than half of those items were not aligned sufficiently with CMI Instruction to retain. In all, some version of 15 items from the Inside the Classroom protocol were descriptive enough of CMI instruction and were retained for the new CMI Observation Protocol. The 15 items predominately were those that described classroom climate and math content. Because of the Inside the Classroom protocol's focus on reform mathematics and the particular type of classroom climate needed to effectively enact reform-oriented mathematics, the retained items were also descriptive of CMI instruction. Similarly, the math content items that were retained described the expectation of rigorous, important mathematics. Rigorous, important mathematics describes the reform mathematics measured by the Inside the Classroom protocol and applies to CMI instruction as well.

The interviews with the observers from the expert panel regarding the Inside the Classroom protocol resulted in several insights about how to approach the structure and rating scales for the new protocol. Observers reported that the arrangement of the items into sub-sections helped the observer to think about the lesson in related chunks, which they found helpful. This structure of sub-sections was retained in the CMI Observation Protocol, although the sub-sections were not the same as contained in the Inside the Classroom protocol. Second, the observers pointed out that the scale for rating each item in the new protocol would be easier to rate with more descriptors than just the end point anchors of "not at all" and "to a great extent," such as were used on the Inside the Classroom protocol. The observers also felt the overall subsection ratings and the overall lesson ratings were helpful. The overall subsection ratings, termed "synthesis" ratings on the Inside the Classroom protocol were incorporated into the CMI Observation Protocol as a holistic rating of each sub-section. The overall lesson ratings, termed “capsule” ratings, as used for the Inside the Classroom protocol were intended to be holistic ratings of the entire instructional episode observed. The observer interviews revealed that these capsule ratings would be easier to use if the descriptors for each level had a parallel structure.

Items for the new protocol were suggested by the document examinations as well as the interviews with experienced CMI observers from the expert panel. Frequent consultation with one of the expert panel members led to many iterations before a well-developed protocol was presented to the entire expert panel for critique and ultimate approval. The completed protocol created as a result of the development methods is described in detail in the following section.

CMI Observation Protocol description. The completed protocol contains six sub-sections. The first three sub-sections of the CMI Observation Protocol are based on the time-order around which the CMI Framework structures lessons. The teaching cycle of a CMI lesson unfolds in a predictable manner: launch, explore, discuss. Although it is appropriate for some lessons to have multiple teaching cycles, each cycle contains the pattern and the elements of each phase. An observer would expect to see the teaching cycle phases and would logically rate the elements of each phase as they unfolded. The remaining three subsections, Mathematics Content, Classroom Climate, and Lesson Coherence, measure elements that thread through all phases of the teaching cycle (see appendix D for the complete CMI Observation Protocol).

Each sub-section contains a number of individual items that are each attached to a five point rating scale. Each point on the scale is labeled to facilitate consistency in rating by observers. A sixth point was included, labeled ‘not applicable’ to accommodate items of a branching nature. Some branching items were included that helped situate the lesson in the learning cycle. For example, different types of tasks are appropriate for each of the three learning cycle phases, Develop, Solidify, or Practice. In items designed to identify the type of task, a 1–5 rating would be given for the item describing the type of task observed, perhaps a 'Develop' task, while a 6 would be marked for the items describing a Solidify task and a Practice task. The assignment of a rating for one of the items and a 6 for the other two items situates the lesson as a Develop Understanding lesson, which shapes expectations for the lesson. Branching items are noted by placing a D, S, or P in a box located before text of the item.

In addition to the individual items that describe elements measured by each sub-section, a holistic rating scale, the *synthesis rating*, gives an overall rating of each sub-section. The synthesis ratings use the information from the item ratings in a global sense, but are not an average of the item ratings. The synthesis ratings measure how well, overall, the instruction uses CMI principles expected to be demonstrated in each particular sub-section.

A final rating known as the *capsule rating* is included in the CMI Observation Protocol. The capsule rating is also a holistic rating, using the information gathered from focusing on the items throughout the entire protocol. The capsule rating represents a level of implementation of CMI principles and practices, measured on a seven point scale. Each scale point is accompanied by a detailed descriptor of what CMI instruction would embody at that level.

The final protocol contains 76 individual items because of the branching nature of 5 clusters of items (a total of 11 individual items). If a lesson becomes identified as a Develop or Practice lesson, there are 70 individual item ratings, while a Solidify lesson has 72. The result of the development activities was the creation of the CMI Observation Protocol, for which validity evidence needed to be gathered.

Validity Results

The research questions driving this dissertation work were about validity. The three types of evidence came from the domains of content evidence, response processes evidence, and internal structure evidence for validity.

Content evidence. The CMI Observation Protocol has strong content evidence for validity. The ultimate content evidence is opinion of experts that the domains to be measured are adequately represented in the measuring instrument. This research employed both a seven-member CMI-expert panel and an instructional expert to give their opinions as content evidence of validity.

The final expert panel opinion was positive toward the ability of the CMI Observation Protocol to measure all of the important elements of the CMI Framework. Adding to the strength of the expert panel's final opinion, the development process was inclusive and iterative. By that it is meant that members of the expert panel were consulted frequently during development and the entire panel weighed in for the final fine tuning before data collection.

Subsequent to the CMI expert panel's approval of the CMI Observation Protocol, the protocol was also given to an outside expert on instruction and observation. Her response was positive:

> Your protocol is precisely and comprehensively articulated. I liked that you covered each phase of the CMI framework individually, yet had some data collected on issues that cut across all three phases of the framework. I can see your concern over the length of the protocol, but this is a complex instructional practice you are researching." (E. Williams, personal communication, January 17, 2011)

Dr. Williams later suggested some minor word changes within a few items, as well as possible attention to the descriptors of the synthesis rating scales.

The CMI Observation Protocol was then given to two of the CMI expert panel members to pilot in the course of their daily professional responsibilities. One of the panel members observed student teachers in his responsibility as a preparer of mathematics educators. His observed instruction took place in secondary classrooms. However, the expectation for his student teachers is that their instruction will closely match CMI instruction, which made those observations good settings for trying the CMI Observation Protocol. He was able to determine if the Protocol measured what he felt he was looking for in CMI instruction, and to give feedback on its use in live classroom observation. His words articulate that the CMI Observation Protocol has content validity:

> I found the protocol very easy to use—even though there are a lot of items, it didn't take much time to rate them. The items also raised my awareness of what was problematic about the lesson, so I found it very useful in preparing for my debriefing sessions with my student teachers. I found that items that seemed most difficult to rate (in that I spent more time trying to decide what score to give that item) were the ones that gave me the most insight into the lesson. These "difficult items" differed from lesson to lesson, which was also insightful and valuable, since it highlighted different issues in each lesson. That suggests to me that you have captured a varied and complex set of issues in the items themselves—which is certainly a strength of the instrument. (S. Hendrickson, personal communication, February 25, 2011)

Response process evidence. The data from the observations obtained for this research can be trusted to represent accurate and consistent ratings. Response process validity evidence captures the idea that steps have been taken so that the sources of measurement error are minimized as much as possible to ensure unbiased data. In this dissertation work, the major potential source of measurement error lies with the observers. However, the selection, training, and calibration of the observers and the quantification of inter-rater reliability serve as strong evidence that measurement error has been minimized.

An asset of the observers selected to observe in classrooms, which was uncovered during the interview in the hiring process, was that they began with a shared understanding of mathematics instruction due to their shared graduate work. Their shared understanding was largely consistent with the principles and practices of the CMI Framework prior to training. The three hours of theoretical training regarding the CMI Framework and the CMI Observation Protocol helped synchronize their prior knowledge with the knowledge of CMI instruction they needed to successfully rate CMI instruction. The live classroom observation work engaged in as part of training further ensured that each observer understood each item in a similar way. After each of the four shared training observations, the four observers met to debrief every item on the CMI Observation Protocol. Differences in understanding were uncovered, talked through, and resolved. If the training is adequate, there is a greater chance of attaining consistency across observers.

A goal of observer training is to produce observers who are essentially interchangeable. With a tool as complex as the CMI Observation Protocol, inadequate training would reveal itself in poor inter-rater reliability measures. The intraclass correlation coefficient (ICC) is a measure of inter-rater reliability that calculates the ratio of between subjects variance to the total variance. In other words, the ICC is a measure of how big a part of the variance is due to differences among the observed teachers' instruction compared to variance due to the combination of differences in instruction and measurement error introduced by the observer. The closer the ratio is to one, the more it can be assured that the ratings reflect differences in instruction rather than differences in observers. After the four training observations, two additional shared classroom observations were made for the purpose of calculating inter-rater reliability using an intraclass correlation coefficient (ICC). The result of the ICC was .71. A rule of thumb is that ICC = 0.40 to 0.59 is moderate inter-rater reliability, 0.60 to 0.79 substantial, and 0.80 outstanding (Landis & Koch, 1977). Thus, the observers attained substantial reliability after training and prior to the main data collection using the protocol. The ICC of .71 coupled with the training methods offer solid evidence of response process validity for the CMI Observation Protocol ratings.

Internal structure evidence. The internal structure of the CMI Observation Protocol showed that the protocol items measure important CMI instruction constructs. Internal structure evidence rests on the coherence between the constructs that are being measured and the items that measure those constructs. Items that align with a construct should be strongly correlated to each other and to the construct, and should not have strong correlations to other constructs.

While the CMI Framework was developed in accordance to literature, accepted “best practices,” and the collective practical expertise of the Math Initiative Committee members, there had not yet been an articulated theory of what CMI instruction was comprised. The item development was guided by the CMI Framework’s view of instruction, the requirements of guided inquiry, and the temporal order of a CMI lesson, not around theoretical constructs. The Launch, Explore, and Discuss sections of the CMI Observation Protocol could be considered descriptions of “things” rather than constructs. The Launch section, for example, has items that describe a launch, the activities that occur to productively get a lesson under way.

The other three subsections, Classroom Climate, Mathematics Content, and Lesson Coherence, sound like constructs, but their use in the CMI Observation Protocol was practical instead of theoretical. These subsections describe the underlying substance that CMI instruction requires. With no explicit theory to guide the analysis of internal structure of the CMI Observation Protocol, the constructs represented by the data needed to be uncovered to make any claims about internal structure. An exploratory factor analysis (EFA) was a first step toward uncovering a reasonable construct structure that could subsequently be tested in confirmatory factor analysis (CFA).

Exploratory factor analysis. The purpose of the EFA was to determine which items from the protocol grouped together and to decide what the relationships were among items thus grouped. A factor extracted in an EFA groups items from the protocol together based on their correlations; however, factors must be examined in order to identify potential constructs. This examination consists of grouping conceptually related items into tentative constructs and naming these constructs.

Employing SPSS software, exploratory factor analysis using a principal component method was performed. The classroom observation data from 65 of the 76 items were used in the analysis. The branching structure of 11 of the items made them unsuitable for use in the EFA. A Kaiser-Meyer-Olkin measure of sampling adequacy (.91—anything over .6 is adequate; Arbuckle, 2006) and a Bartlett's test of sphericity ($p = .000$) both showed the data suitable for the EFA (Arbuckle, 2006). A scree-plot of the data suggested a three factor solution (Figure 4). A 3-factor EFA with varimax rotation provided a solution that accounted for 57.7% of variance with few crossloadings. Each of the three factors from the EFA was interpreted to describe multiple constructs. The interpretations were based on the researcher's knowledge of CMI instruction viewed through the CMI Framework lens.

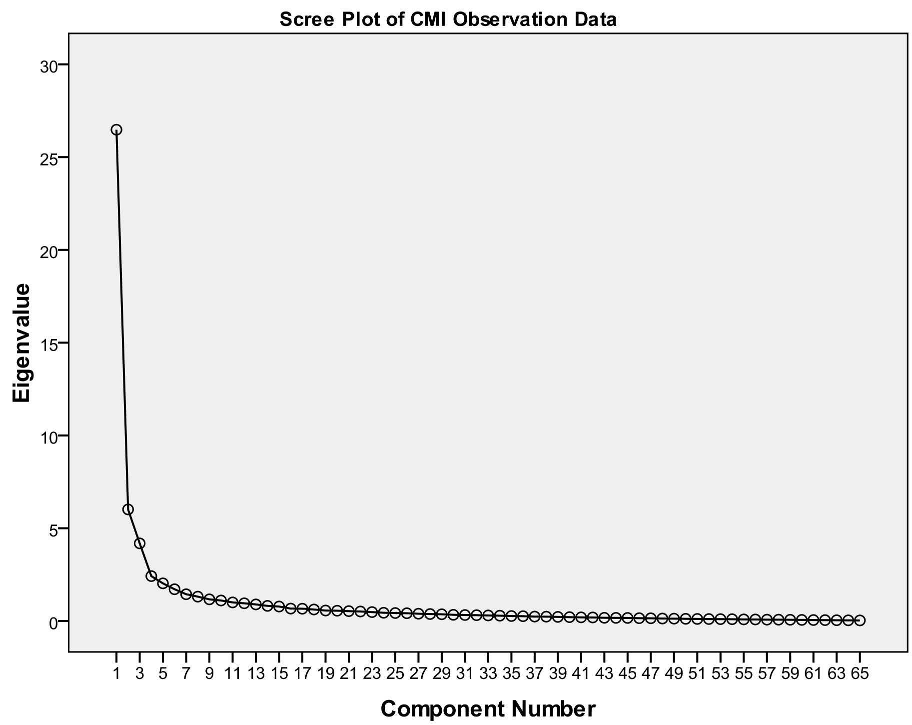

Figure 4. Scree Plot of CMI Observation Data

Factors described. The first factor contained 30 items from the CMI Observation Protocol. This factor was divided into three constructs: flexible teaching for understanding, maximizing student thinking, and organization for supporting student engagement.

The second factor was clearly about discourse surrounding mathematics. It contained 14 items. This factor was split into two constructs: student discourse and teacher discourse. Two constructs were proposed because of the CMI Framework's emphasis on interactive instruction with teacher and student roles.

The final factor contained 21 items. This factor described the milieu in which the instruction occurs. There seemed to be two constructs described by the items loading on the factor: instructional climate and math content. The items that describe the named constructs are listed in Tables 2--8.

Constructs defined. The first three constructs describe central aspects of guided-inquiry instruction. These three constructs are clearly related to one another, but they are different in important ways. For example, *flexible teaching for understanding* focuses on how teachers plan instruction in anticipation of student needs and adjust the plan in response to students in order to support them in constructing meaning. The 10 items in this construct are about the differentiation within a lesson. Teacher actions in this construct create favorable conditions for student thinking. The items describing flexible teaching are found in Table 2.

While *maximizing student thinking* is also an important teacher consideration, the items in this construct focus on what both teachers and students do to become aware of and deepen student thinking. In this construct the nine items emphasize questioning but also include a time factor and a content factor. Teachers use questions to move student thinking and students question themselves to examine their thinking. The relation of *maximizing student thinking* to *flexible teaching for understanding* seems a matter of focus. Flexible teaching helps students begin to think by creating favorable conditions by adjusting to student needs, while maximizing that thinking capitalizes on the conditions created but presses for breadth and depth. Table 3 shows the items describing maximizing student thinking.

The third construct, *organizing to support engagement*, describes structures, materials, and tools that help students become active participants in the lesson. The 11 items shown in Table 4 include student groupings, teacher actions in getting students involved in the task, as well as the use and accessibility of materials and tools. The emphasis is on organization and material support that help students connect with the mathematics.

Table 2
Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: Flexible teaching for understanding

Construct	**Protocol section**
Flexible teaching for understanding	
The task selected had multiple entry points, accommodating a variety of student needs	Launch
The teacher differentiated within the task to meet student needs	Explore
The teacher maintained the task at an appropriate level to meet the purpose of the phase (e.g. supplying only necessary scaffolding, allowing purposeful struggle, avoiding "telling")	Explore
The teacher was able to "read" the students' level of understanding and responded appropriately (includes moving to discussion at appropriate times)	Explore
The teacher strategically monitored student exploration (e.g. not needing to get to every individual/group)	Explore
The teacher displayed an understanding of student trajectory through the mathematics (e.g. by connections activated/built, selection of student work, discussion)	Mathematical Content
The task(s) aligned with the mathematical purpose	Lesson Coherence
The task's level of constraint was appropriate for the purpose of the lesson (from very open to very constrained)	Lesson Coherence
The pace of the lesson was appropriate for the needs of the students (the right amount of time was provided for exploring the task, holding the discussion, etc.)	Lesson Coherence
The instructional strategies and activities used in the lesson reflected attention to students' experience, preparedness, prior knowledge, and/or learning styles	Lesson Coherence

Table 3
Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: Maximizing student thinking

Construct	**Protocol section**
Maximizing Student Thinking	
The teacher allowed adequate exploration prior to engaging with students	Explore
The teacher monitored student thinking during exploration by asking brief questions, eavesdropping on conversations and/or visually scanning student work	Explore
The teacher's questions facilitated exploration (e.g. by engaging students in the task, prompting or guiding exploration, clarifying student thinking, deepening student thinking)	Explore
The teacher's questions facilitated and /or directed student understanding & ownership (e.g. by prompting, clarifying, guiding, scaffolding, probing, and/or connecting mathematical thinking)	Explore
The teacher's questions helped students to become more aware of what they were thinking and/or doing	Explore
The teacher's questions helped students to refine their thinking/understanding	Explore
Students engaged in sensemaking by asking themselves, "Is this accurate;do I understand this; can I explain this; where would I use this."	Explore
The mathematics content was significant and worthwhile	Mathematical Content
An appropriate balance of teacher-talk and student talk was achieved	Lesson Coherence

Table 4
Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: organizing to support engagement

Construct	**Protocol section**
Organizing to support engagement	
The teacher activated or built students' contextual background knowledge necessary for the task	Launch
The teacher facilitated student engagement in exploration (e.g. through task structure, reminders or norms, pre-alerting)	Explore
The teacher made appropriate tools available to assist students with mathematical understanding	Explore
Student groupings (individual, pairs, small group) facilitated the exploration of the task(s)	Explore
The students used appropriate tools to assist with mathematical understanding (e.g. drawings, manipulatives, etc.)	Explore
The task was neither too easy nor too difficult for the majority of students	Lesson Coherence
The resources available in this lesson contributed to accomplishing the purposes of the instruction	Lesson Coherence
The student groupings allowed for the mathematical purpose to occur (e.g. proximity to others, size of group)	Lesson Coherence
The structure of the task encouraged appropriate collaboration among students	Lesson Coherence
Previous student thinking is evident in the classroom (e.g. charts of previous discussions)	Classroom Climate
Materials were organized and accessible to students	Classroom Climate

The two constructs derived from the second factor were *teacher discourse* and *student discourse*. Discourse has been dichotomized in some literature as *univocal* or *dialogic* (Truxaw & DeFranco, 2008). As the term implies, univocal discourse is that speech which is rendered by one to transmit information. In contrast, dialogic discourse has a two-way nature with the purpose of constructing meaning. The constructs in this study, *teacher discourse* and *student discourse*, connote the latter meaning with its give and take implications. The eight items that describe *teacher discourse* describe those things a teacher does to enable the dialogic discourse to occur (see Table 5). The six items grouped under the construct *student discourse,* shown in Table 6, describe the students' roles in engaging in the dialogic discourse.

Table 5

Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: teacher discourse

Construct	**Protocol section**
Teacher Discourse	
The content of the discussion reflected purposeful selection of student work from the exploration that focused on the lesson's purpose	Discuss
The discussion had a logical, coherent structure that built on the exploration (e.g. ordered by the level of complexity)	Discuss
The discussion was ordered such that the development of connections was facilitated (sequencing builds to the 'aha')	Discuss
Student thinking/work about the mathematics was represented in the whole class discussion	Discuss
The teacher asked probing/directed questions to draw out explicit and specific connections (the teacher doesn't leave the 'aha' to chance)	Discuss
The teacher used appropriate "talk moves" (revoiced, used wait time, asked for students to restate, dis/agree, add to)	Discuss
Explicit/direct instruction was used when appropriate (e.g. for giving information, conventions, norms)	Discuss
The discussion closed with a clear sense of what had been discussed	Discuss

Table 6

Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: student discourse

Construct	**Protocol section**
Student Discourse	
Students communicated, explained, and/or supported their own thinking during the whole class discussion	Discuss
Students question the thinking of peers during the whole class discussion	Discuss
Students elaborated on the thinking of peers during the whole class discussion	Discuss
Non-sharing students actively listened during the whole-class discussion (showed readiness to restate, explain, add to, dis/agree)	Discuss
Students probed other students to articulate their reasoning behind ideas, strategies, representations, etc. for the purpose of advancing understanding	Discuss
Students connected discussion points with an important mathematical purpose	Discuss

The last two constructs that emerged from the researcher's interpretation of the third factor from the EFA are *instructional climate* and *mathematics content*. Instructional climate (Table7) contains 12 items that describe the norms that form the environment which prepare students to be able to freely think and that minimize distractions diverting attention away from the mathematics. Mathematics content (see Table 8), with nine items, is described by the mathematical accuracy and understanding that the teacher displays, as well as the appropriateness of the mathematics for students. Additionally, mathematics content includes whether the lessons include a clear mathematical purpose.

Table 7

Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: instructional climate

Construct	**Protocol section**
Instructional climate	
The teacher clarified the task for students (to a degree matched to the purpose of the phase) before they began the exploration	Launch
Students responded appropriately during the launch: they listened actively (eyes on speaker, etc.) and/or they asked clarifying questions before beginning, if needed.	Launch
Students appeared to know what the task entailed, (i.e. affect, body language) and were prepared to immediately engage in the task (participated/supported participation)	Launch
Students persist and maintain effort on the exploration task	Explore
Students display positive affect about the task (e.g. enthusiasm, curiosity, etc.)	Explore
The teacher's classroom management style/strategies enhanced the quality of the lesson (e.g. clear procedures, effective transitions)	Classroom Climate
There was a climate of respect for students' ideas, questions, and contributions	Classroom Climate
Interactions reflected collaborative working relationships between teacher and students	Classroom Climate
The climate of the classroom encouraged students to generate ideas, questions, and/or conjectures	Classroom Climate
Interactions reflected collaborative working relationships among students	Classroom Climate
Active participation of all was encouraged and valued	Classroom Climate
The teacher displayed confidence in his/her ability to teach mathematics	Mathematical Content

The results of the researcher's interpretation of the EFA using the data from 144 classrooms suggested that the Comprehensive Mathematics Instruction (CMI) Observation Protocol measures seven constructs. Using the results from the EFA, each construct with its descriptive items was individually tested for sound internal structure through CFA. An entire model built with the seven constructs revealed by EFA was also tested using CFA.

Table 8

Constructs uncovered in the EFA and their associated items from the CMI Observation Protocol: mathematics content

Construct	**Protocol section**
Mathematics Content	
The teacher activated or built students' mathematical background knowledge necessary for the task	Launch
The teacher activated or built students' connections between the current lesson and previous work	Launch
The teacher noted/recorded student thinking pertinent to use in the discussion (e.g. examples, common misconceptions)	Explore
The teacher used mathematical language, conventions, and symbols accurately	Discuss
Students used mathematical language, conventions, and symbols accurately (appropriate for grade level)	Discuss
The content information provided by the teacher was accurate	Mathematical Content
The mathematics content was appropriate for the developmental levels of the students in the class	Mathematical Content
The teacher displayed an understanding of the mathematics being taught (e.g. in his/her dialogue with students)	Mathematical Content
There was a clear mathematical purpose intended for the lesson (adjustments may be made as the lesson unfolds)	Lesson Coherence

Confirmatory factor analysis. The confirmatory factor analyses of single factor models show the constructs proposed as underlying CMI Observation Protocol to have satisfactory internal structure. Single-factor models were built for each proposed construct in order to confirm that the items loaded appropriately onto their respective construct, and that the construct was unidimensional. Unidimensionality describes the condition where every item that represents a facet of the construct is only associated with that construct and no other. An example of a single factor model is shown in Figure 5. Figure 5 depicts the single-factor model of the construct *flexible teaching for understanding*. Ten items from the protocol are represented by the squares on the model diagram, while the circles represent the error associated with the items. As can be seen by the connecting arcs, some errors are correlated. That is to say that measurement error is likely to be influenced by the same circumstances in items that have correlated errors. All of the single factor models are located in Appendix E.

An overall model built from the 7 constructs containing all 65 of the CMI Observation Protocol non-branching items showed some areas of concern and was modified, weakening the claim for internal structure on the entire protocol. The first area of concern was that two constructs, student discourse and teacher discourse, were nearly perfectly correlated. The second area of concern was items with low (<.60) factor loadings. Both the perfect correlation and low factor loading conditions directed the modification of the model.

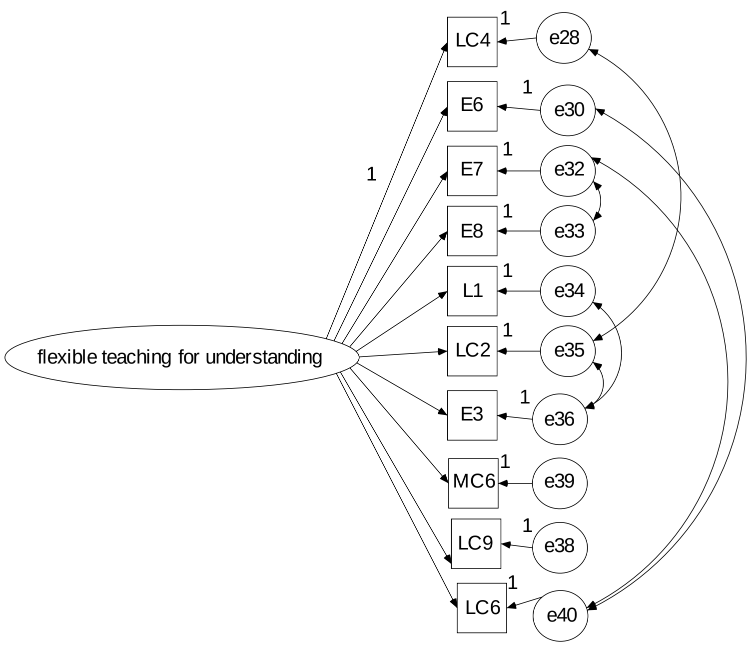

Figure 5. Single factor model for flexible teaching for understanding.

First, in order to get unbiased model fits, the two discourse constructs needed to be collapsed into a single construct, *teacher & student discourse*. Without the modification of taking the two separate discourse constructs and merging them in to one, the analysis of model fit would be misleading. Second, eight items with the lowest factor loadings were strategically removed from the model. Each was removed individually to test the impact on the factor loadings and on the model. Three items came from the *math content* construct, three from *organizing to support engagement*, and one each from *flexible teaching for understanding* and *maximizing student thinking*. A modified model was created that used the remaining 57 protocol items and showed adequate internal structure (see Figure 6). A good model fit for the modified model would suggest that the CMI Observation Protocol measures the constructs in the model, and thus has evidence that it can be trusted to measure CMI instruction.

The CFA single factor models of each proposed construct, the overall 65-item model, as well as the modified 57-item model were checked for fit by three measures: Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). Multiple indices selected from the different types are recommended (Brown, 2006). These three measures, the TLI, the CFI, and the RMSEA are suggested by Brown (2006) because of their "overall satisfactory performance" (p. 86). Both the TLI and CFI are comparative measures, although the TLI adjusts for model complexity, and the RMSEA is also a parsimony adjusted measure. The TLI, because it is a non-normed measure, may have values less than 0 or greater than 1, but the CFI and RMSEA range is between 0 and 1. The interpretation of the TLI and CFI is as follows: < .85 indicates poor fit; .85–.90 is a mediocre fit; .90–.95 is an acceptable fit; and .95–.99 indicates a close fit. RMSEA values between .10 and .08 are considered evidence of mediocre fit, values below .08 are considered an adequate fit, and values below .05 suggest a good fitting model.

The fit indices as shown in Table 9 suggest that data from this particular sample provide evidence that the constructs the CMI Observation Protocol measures are *flexible teaching for understanding*, *maximizing student thinking*, *organizing to support engagement*, *discourse*, *instructional climate*, and *math content*.

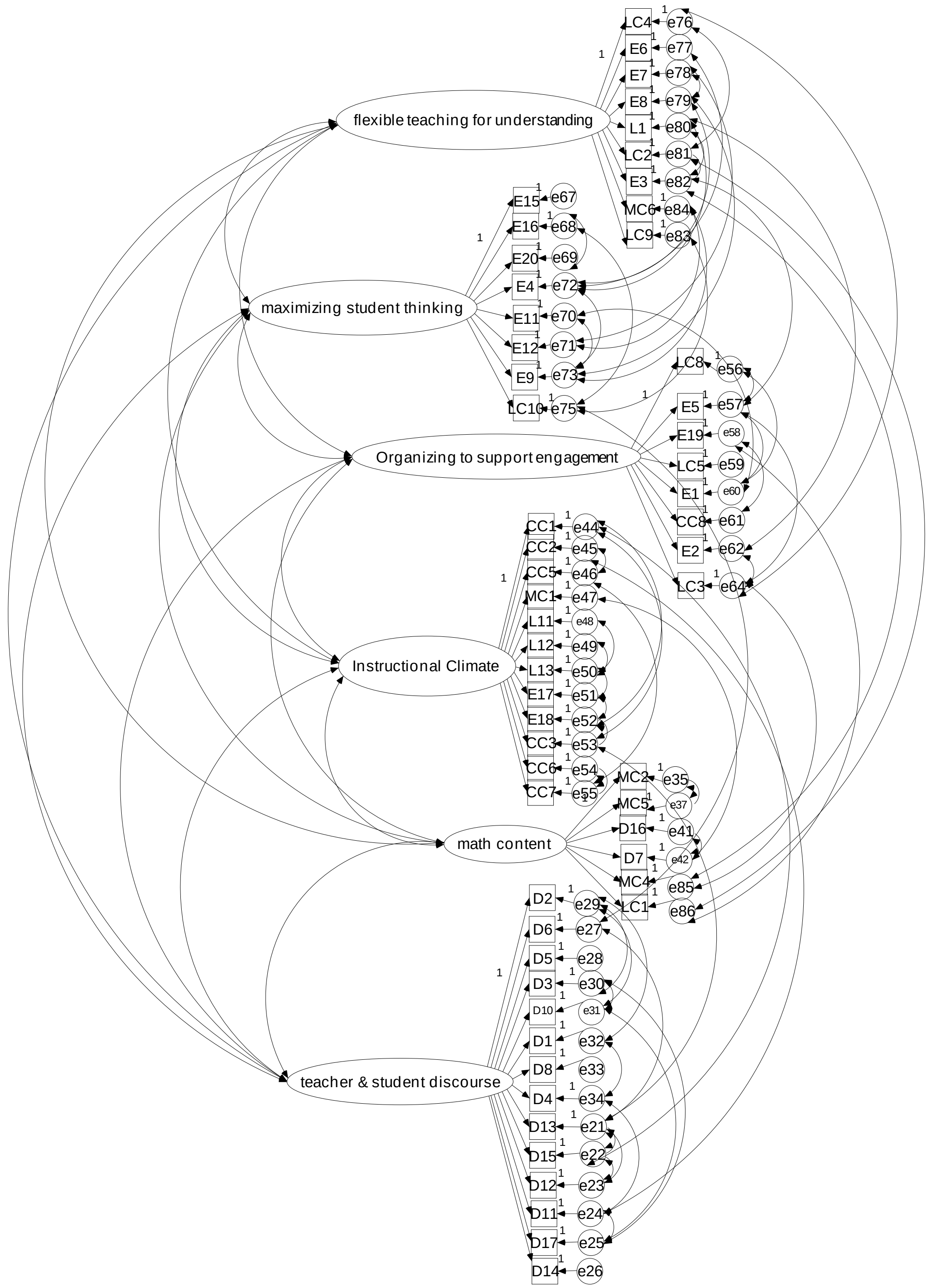

Figure 6. The modified 57-item model of constructs of the CMI Observation Protocol

The modified model fit indicates that the 57 items from CMI Observation Protocol produce an adequate internal structure. The mediocre-to-adequate fit for the modified 57-item model suggests that the CMI Observation Protocol measures much fairly well, but could be improved.

Table 9
Proposed Constructs and their CFA Single Factor Model Fit Indices

Construct	Tucker-Lewis Index (TLI)	Comparative Fit Index (CFI)	Root Mean Square Error of Approximation (RMSEA)
flexible teaching for understanding	.968	.980	.066
maximizing student thinking	.987	.992	.049
organizing to support engagement	.973	.983	.055
teacher discourse	.986	.992	.060
student discourse	1.000	1.000	.000
teacher & student discourse	.944	.961	.095
instructional climate	.940	.961	.083
math content	. 961	.975	.057
overall 65-item model	.798	.813	.076
modified 57-item model	.848	.860	.071

Internal consistency reliability. Overall, the CMI Observation Protocol has a high degree of reliability. Reliability, the consistency of the ratings, is related to validity in that validity cannot exist unless the measure can be counted on to measure the same way every time it is used. The internal consistency reliability of the protocol was measured using Cronbach's alpha and Raykov's Rho. A general 'rule of thumb' for the interpretation of Raykov's rho and Cronbach's alpha is > .9 excellent, > .8 good, > .7 acceptable, > .6 questionable, > .5 poor, < .5 unacceptable (George & Mallery, 2003). While Cronbach's alpha is widely reported for internal consistency reliability, it is known to over- or underestimate reliability (Brown, 2006). Raykov's Rho is considered more robust against the violations of the Cronbach's alpha assumptions of tau-equivalence and un-correlated error (Brown, 2006). Since the models tested in this research had correlated errors, both Cronbach's alpha and Raykov's rho were reported. Table 10 shows the internal consistency reliability estimates for the single factor models of each construct and for the modified model of the protocol. These results indicate that the items that describe each construct show good to excellent reliability in measuring the constructs.

Table 10
Cronbach's Alpha and Raykov's Rho Estimates of Internal Consistency Reliability of CFA Models

Construct	Cronbach's α (alpha)	Raykov's ρ (rho)
flexible teaching for understanding	.923	.916
maximizing student thinking	.932	.925
organizing to support engagement	.900	.934
teacher discourse	.946	.925
student discourse	.924	.887
teacher & student discourse	.964	.943
instructional climate	.928	.937
math content	.802	.867
overall 65-item model	.976	.986
modified 57- item model	.975	.987

Results Summary

The development activities resulted in an observation tool, the Comprehensive Mathematics Instruction Observation Protocol. A panel of eight experts gave strong content evidence for validity. The observer training methods together with the substantial inter-rater agreement evidenced by the intraclass correlation coefficient (ICC) of .71 offer strong response process evidence for validity as well. Internal structure evidence for validity for the individual constructs proposed as underlying the CMI Observation Protocol is also strong. The TLI and CFI of the overall (65 item) model are interpreted as a poor fit to the data. However, in comparison to the null model of no correlations (which is what the TLI and CFI do), the overall model is 80% better at reproducing the correlations among the data. The modified model shows evidence that the protocol measures most aspects of CMI Instruction adequately. The items that were included in the modified model represented about 88% of all the items on the protocol, 57 of 65. However, the claim of internal structure for the total protocol is somewhat weakened because all of the items are not represented. The proposed constructs from the EFA interpretation seem to describe important aspects of the CMI Framework. Internal consistency reliability of the items in measuring these constructs is high, and the reliability of the 57 items from the modified model is high, as well. The CMI Observation Protocol items represent CMI instruction quite well, and, with some limitation, validly measure important constructs with high reliability.

Chapter 5: Discussion

Mathematics achievement for students has been the focus of reform over the past several decades (NCTM, 1980, 1989, 1991, 2000, 2009) and continues to be a concern to American educators and political leaders today (National Mathematics Advisory Panel, 2008). Improved instruction has been identified as a large part of the remedy for lagging student mathematics achievement (Stiegler & Hiebert, 1999; Hiebert & Grouws, 2007). The developers of the Comprehensive Mathematics Instruction (CMI) Framework recognized that often educators did not have an instructional framework to guide them in implementing the kinds of practices that would produce powerful mathematics instruction. The CMI Framework is potentially a solution to the problem of how to improve classroom instruction so student mathematics understanding and achievement are strengthened.

The word “potentially” is used carefully here. The CMI Framework and its effect on instruction and on subsequent student achievement have been studied since 2006. The preliminary results have shown improving instruction and improving student achievement (Hilton, Hendrickson & Bahr, 2010). The tool used to measure CMI instruction in these studies, the *Inside the Classroom Observation Protocol*, measured many aspects of reform instruction, but it did not exactly capture CMI instruction. However, the tool that has been used to measure CMI instruction is something akin to using standard wrenches on metric bolts. One may find a close match that is workable, but there will likely be some slippage. Without a specific measure, the effect of the CMI Framework on instruction and that instruction’s effect on student achievement remain somewhat ill-defined. This dissertation, then, was undertaken to solve the problem of accurately and specifically measuring instruction according to the CMI Framework.

Review of Purpose

The purpose of this research was to develop and validate a tool that would measure instruction reliably and produce data that could be trusted as a valid snapshot of CMI instruction. The development of a measurement tool requires many decision points. Through what lens will the measurement take place? What are the salient features of instruction that are important to be captured? What method of measurement best aligns with the purpose of measurement? How will the important issues of reliability and validity be managed? Thus, measuring instruction is a complex undertaking (Ball & Rowan, 2004).

The complexity of instructional measurement was approached incrementally. The CMI Framework provided the lens through which to view instruction. Experience with the CMI Framework, important conversations with CMI and instructional experts, and the research literature all contributed to answering the questions about salient features and method alignment with purpose posed by Ball and Rowan (2004). The resulting protocol drew upon existing instructional protocols, but uniquely captures CMI instruction.

Content Evidence

The evidence is strong from the expert panel about the complete representation of CMI instruction in the CMI Observation Protocol. Although each member of the panel brought a different perspective (researcher, observer, math specialist, or administrator), as a group the consensus was that using the CMI Observation Protocol to record a lesson observation would give one a complete and accurate snapshot of CMI instruction. A teacher who aligns instruction with the CMI Framework and demonstrates skill in executing that instruction would be marked accordingly and would differ from a teacher who teaches in some other way.

Additional content evidence comes from the alignment between the CMI Protocol at both the item level and the construct level and published literature. The CMI Framework was developed to answer the call on a local level to improve mathematics, which echoed the decades-long national call embodied in documents from the NCTM and NRC, and more recently from the National Advisory Panel. The Math Initiative Committee (MIC) brought their collective experience, which included awareness of and agreement with reform mathematics, to bear when developing the CMI Framework. It is not surprising then that at both the item level and at the construct level the CMI Observation Protocol contains elements that are consistent with the literature from the reform mathematics domain.

For example, in the Professional Standards for Teaching Mathematics (NCTM, 1991), six standards are promoted in four domains: 1) tasks, 2) discourse, 3) environment, and 4) analysis. Similar to Professional Standard 1, the CMI Observation protocol contains items measuring the content and characteristics of tasks and how they are implemented. Task content and characteristics are represented in the maximizing student thinking and organizing for student engagement constructs of the protocol. Discourse has a prominent presence in the CMI Observation Protocol, and includes student roles and teacher roles, just as Standards 2 and 3 of the Professional Standards do. Items regarding each role were found describing the construct teacher and student discourse. Similar to the Professional Standards Standard 4 are items about tools supporting discourse. These items are part of the description of the organizing to support engagement construct. The NCTM Environment domain, Standard 5, is mirrored by items in the Classroom Climate sub-section, as well as items in the Explore sub-section. Standard 5 finds a parallel in the instructional climate construct underlying the CMI Observation Protocol. Finally, the NCTM Analysis Standard 6 is captured by the CMI Protocol items measuring flexible teaching. This connection between the Professional Standards and the CMI Observation Protocol adds to the evidence for content validity by showing the alignment between the Standards, the CMI Framework, and the CMI Observation Protocol items and constructs.

Additional concurrent evidence for content validity comes from the more recent NCTM publication, Principles and Standards for School Mathematics (NCTM, 2000). Five of the six principles noted in chapter one of this dissertation, equity, curriculum, teaching, learning, and assessment, are embodied in the constructs of the CMI Observation Protocol. One would find elements of the principles located in items describing the six protocol constructs, and scattered throughout the constructs. For example, the equity principle is explained as providing high expectations and strong support for all students. The construct *organizing for student support* contains items aligning to the equity principle. The teaching principle, understanding what students know and need to learn and then challenging them and supporting them to learn it well, would find counterpart items in the math content, organizing for support, maximizing student thinking, and flexible teaching constructs in the protocol. For each CMI Observation Protocol construct, there are corresponding ideas in the Principles. Knowing that the items and constructs of the CMI Observation Protocol find alignment and connection with current conceptions of mathematics instruction promoted by the mathematics education community at large adds to the assurance that important aspects of mathematics instruction are being measured.

Internal Structure Evidence

The CMI Framework was created based on the practical needs of teachers. Although the Framework drew from understandings created by research literature, and the theoretical and practical knowledge of the creators, there was no articulated theory to describe the CMI instruction. In order to test the internal structure of the CMI Observation Protocol, a structure needed to be posited. The initial work with EFA and then the CFA seemed to indicate that the protocol measured at least seven constructs: 1) flexible teaching for understanding, 2) maximizing student thinking, 3) organizing to support engagement, 4) teacher discourse, 5) student discourse, 6) instructional climate, and 7) mathematics content.

When these seven constructs and their corresponding items were included in a CFA model, the teacher and student discourse constructs were almost exactly correlated, indicating that in these data the items tap into a single construct. Thinking about the dialogic nature of discourse in CMI instruction and the several roles promoted, there is an argument to be made about separate constructs that would retain the distinct roles of teacher and student. In CMI instruction, the teacher's role is not only to have discourse interactions between the teacher and a group of students, but also between the teacher and individuals. Likewise, a student's role is to discourse with the teacher as an individual and as part of a group, as well as interact in mathematical discourse with individual students and groups of students.

Although there is some overlap in who is engaging with whom in discourse, there are distinct expectations for teacher and student. In guided inquiry, a teacher prompts, presses, and provides support to move student thinking forward. A teacher also has a larger vision of what each student is thinking and has students bring forth their thinking to contribute to the classroom discourse. The teacher is the mentor, an experienced other, who scaffolds the learning for students by shaping the discourse of the classroom. The teacher selects which student ideas to pursue, based on the contributions the ideas are likely to make in mathematical understanding.

Although the teacher shapes the discourse of the classroom through their expanded view, in CMI instruction students offer the material that is shaped and by so doing, also determine what discourse takes place. The students' role in discourse is to explain and justify their thinking and to question the thinking of themselves and others, including the teacher. The different roles of student and teacher in creating and engaging in discourse would seem to describe separate constructs. In future iterations of the protocol, it would seem wise to examine the items pertaining to the discourse roles and consider whether they accurately capture CMI's ideas about discourse.

While it seems more descriptive of CMI instruction to have the teacher and student discourse roles represented as separate constructs, the CFA results strongly indicated that they are different aspects of a single construct. Although it is not this researcher's view, perhaps the roles of students and teachers in discourse are too closely intertwined to disentangle. For this analysis with these data all of the items from the separate teacher discourse and student discourse were combined into one *teacher and student discourse* construct. Because the two constructs were as highly correlated as the two discourse constructs were, the software warns that inaccurate fit estimates may be produced. When the two discourse-oriented constructs were collapsed into teacher and student discourse, a 6-construct model was produced with which to test the internal structure. This 6-construct model using all of the 65 protocol items (excluding the 11 branching items) produced an adequate RMSEA value (.076) and inadequate TLI and CFI values (<.850). Although the overall model was 80% better at representing the relationships among the items than a null model assuming no relationships, a better fitting model was sought.

The three fit indices (TLI, CFI, and RMSEA) indicate how well the model fits the data and are considered together. In the case of the overall model, the composite information that the indices gave indicated that the model was inadequate in describing the current data from the CMI Observation Protocol. A better-fitting model was produced by removing eight items from the model. All of the eight items removed had factor loadings less than .6, indicating that only a small portion of the variability in the items was accounted for by the construct they were associated with. Those eight items were not contributing much to the measurement of the constructs they had been associated with. Since a CFA model attempts to reproduce the relationships seen in the data, removing some items from the model changes the relationships among the others that remain. With the 8 items removed from the model, the TLI and CFI values were ≥.850, and the RMSEA was .071. The modified model containing 57 CMI Observation Protocol items showed adequate internal structure. Those 57 items measure 6 constructs well.

The items removed from the overall 65-item model testing the internal structure of the CMI Observation Protocol included items on activating contextual, mathematical, and connecting background knowledge, and the degree to which previous work was visible. Other items that were removed measured pacing, the degree to which the mathematics was significant and worthwhile, the degree to which the task structure encouraged student collaboration, and the degree to which student thinking or work were noted or recorded for use in the discussion. All of these items seemed important to measuring implementation of CMI instruction.

Of note in the modified model is that three of the lowest-loading items that were removed came from the *math content* construct. The math content construct ended with the fewest items and the lowest factor-loadings. The low factor-loadings in the overall 65-item model and the scarcity of items measuring math content in the modified model suggest that the construct could be better measured. A reason this may be the weakest construct is that content may be difficult to observe, even with good descriptors. For example, one of the items retained in the model rated teacher understanding of the mathematics. The evidence for the rating was to be taken from the teacher’s dialogue with students. It is possible that circumstances surrounding the dialogue prevent the teacher from demonstrating understanding. Perhaps the discourse got short-circuited by management issues. Perhaps there is not sufficient opportunity in the course of the lesson to demonstrate understanding. Although it was felt that teacher math understanding could be derived from hearing what teachers had to say and how they questioned or otherwise moved students along, much of their understanding may remain inaccessible to an observer. Those items that tap into the math content construct, as shown in the CFA, might need to be rewritten to better describe how math content contributes to CMI instruction.

The remaining 5 items that were removed from the overall 65-item model should also be examined in light of their performance in the CFA. Perhaps rewording would produce a better descriptor that would strengthen their correlation with the construct on which they loaded. This may be the case for item three from the Math Content sub-section of the protocol which loaded weakly on the maximizing student thinking construct: The mathematics content was significant and worthwhile. Perhaps the removed items are important aspects of CMI instruction but are not really associated with a construct. One item on pacing of the lesson, which weakly loaded on the flexible teaching construct, and another item about previous work evident in the classroom from the organizing for student support construct may fit this category. A final consideration in thinking about removed items is that they may actually represent other, as yet unidentified, constructs. If that is the case, further revision of the protocol and further data collection will help clarify the relationship of the removed items to CMI instruction constructs.

Additional support for internal structure validity was provided by internal consistency reliability. Although reliability is often discussed separately from validity, validity cannot exist without reliability. The internal consistency reliability of each single-factor model was quite good, ranging from $\rho = .867$ (good) to $\rho = .943$ (excellent). The modified model of 57 items also showed excellent reliability ($\rho = .987$). Knowing that each construct showed internal consistency revealed that the protocol was coherent and gave assurance that the ratings can be trusted.

In the EFA and CFA that were conducted, analyzed, and interpreted using the data from the 144 observations of instruction, 6 constructs emerged: 1) flexible teaching for understanding, 2) maximizing student thinking, 3) organization for supporting student engagement, 4) teacher and student discourse, 5) instructional climate, and 6) math content. Referring back to the conceptual model of instruction upon which this dissertation is based (see Figure 1 in the introduction), each construct uncovered is consistent with the multilayered, interactive nature of CMI instruction. This section justifies the constructs that have been proposed as making up the internal structure of the CMI Observation Protocol.

Flexible teaching for understanding. This construct is consistent with CMI instruction's teaching and learning cycles. The Framework emphasizes that instruction changes depending on the purpose of the lesson (Develop, Solidify, or Practice Understanding). The Framework also proposes that within a teaching cycle (Launch, Explore, Discuss) student and teacher roles interact. These interactions also change instruction. In the context of both the learning cycle and the teaching cycle, teaching must be flexible for those utilizing the CMI Framework to guide their instruction. That is, teachers must be able to anticipate how the lesson might unfold but also respond "on the fly" to capitalize on student understandings as they become apparent. The type of flexible teaching proposed by the framework maintains conditions whereby students are constructing meaning. For example some of the teacher roles in the explore phase of a Solidify Understanding lesson are to "anticipate student thinking and misconceptions and plan responses to guide focused discussion . . . facilitate and direct student understanding and ownership . . . and continually refocus student thinking" (CMI Framework, 2008). The Develop and Practice portions of the CMI Framework contain similar descriptions of teacher roles. In order to "facilitate" ownership and "refocus" thinking, a teacher needs to be tailoring responses that allow students to push ahead in their construction and connections for meaning. This includes teachers "reading" students' understanding, knowing their expected trajectory through the mathematics, and adjusting the pace and strategies they use. These actions describe the construct flexible teaching for understanding.

Maximizing student thinking. Maximizing student thinking is an integral part of the CMI Framework and CMI instruction. Guided inquiry instruction has at its core not only that students are constructing meaning, but that teachers serve as experienced guides to assist students in making sense of the constructions. Throughout the teaching and learning cycle phases, there is an emphasis on what teachers and students do to become aware of thinking in order to make the best use of student thinking. This construct is predominately described by the questions students ask themselves as well as the questions teachers ask students. There are also the ideas of significant, worthwhile mathematics and purposeful struggle in allowing adequate time for engaging in the task before the teacher interacts with students that are captured by this construct.

Organization for supporting student engagement. This construct is described by materials, tools, and structures being available and used to help students work with the mathematics of the lesson. CMI instruction, guided by the Framework, requires teachers to mindfully plan to facilitate student engagement. In order to fully engage in the mathematics, students need appropriate materials and tools to make their thinking visible. They need to have optimal groupings that support the work they undertake, pushing them to think without overwhelming them. Unless students are engaging with the mathematics, they will have little chance of developing understanding. The groupings, materials, and tools help create the organizing structure for supporting student engagement.

Discourse. Separating the items into the two constructs of *student discourse* and *teacher discourse* seemed consistent with the CMI Framework's student role and teacher role delineations. However, the two roles intersect, and the high correlation in these data between the constructs reflected that intersection. The single construct of teacher and student discourse includes both roles. A key feature of CMI instruction is that students question themselves and others, and explain, justify, and otherwise interactively talk with peers and the teacher to understand the mathematics. Likewise, teachers are to orchestrate a discussion, probe students to draw out their thinking, and otherwise create discourse that allows students to connect ideas within the lesson and across lessons. CMI instruction is discourse-rich instruction.

Instructional climate. In the conceptual model of instruction outlined in the introduction, Cohen, Raudenbush, & Ball's (2003) "environment" was a sort of matrix into which all instruction was set. Embedded in the guided inquiry instructional model the Framework promotes are social constructivist ideals. These ideals include peer to peer interactions and adult to student interactions which cannot effectively occur in the absence of respect and trust. The guided inquiry model also includes clarity of expectations. Although the task deployed may be an open or more constrained task, students need to grapple with the mathematics rather than ambiguity in behavioral expectations or task instructions. The eight items that were included in the CMI Observation Protocol's Classroom Climate section and approved by the expert panel were there because of the importance of a instructional climate in supporting rigorous work and in developing the "positive disposition," "persistence," and "effort" that reform mathematics calls for (NCTM, 2000; NRC, 2001; National Mathematics Advisory Panel, 2008). Six of the eight Classroom Climate items loaded on the instructional climate construct, along with six items from throughout the Protocol.

Math content. In a measure of mathematics instruction one would expect to find mathematics content in the mix of what is being measured! A focus of CMI instruction is on having a clear mathematical purpose for every lesson. There is an understanding that in CMI instruction the mathematics should be important and coherent, as the NCTM principles recommend (NCTM, 2000). Additionally, there is an expectation that the teacher has sufficient depth of understanding of the mathematics to accurately teach and communicate about it. This construct in the modified model only contains items about the accuracy of the mathematics being taught and communicated, and the teachers' understanding of the mathematics. Since math content is quite important, the items from the overall model that loaded on this construct should be re-examined for modification in order to more fully capture the clarity of purpose, importance, and coherence of the mathematics.

Theory Development

This dissertation research, though largely a development and validation study, also contributed to theory development for the CMI Framework. Each of the constructs uncovered by the EFA and tested in single factor CFA models and the multiple-factor modified model are constructs that make sense in terms of reform mathematics and match the CMI Framework's conceptions of instruction. 1) *Flexible instruction for understanding*, 2) *maximizing student thinking*, 3) *organization for supporting student engagement*, 4) *teacher & student discourse*, 5) *instructional climate*, and 6) *math content* are all important facets of CMI instruction.

Whetten (1989) asserts that there are four building blocks to theory development:

1) identifying what factors are involved in describing the phenomenon, 2) determining how the factors are related, 3) positing assumptions about the theory, and 4) explicating the who, where, and when of generalizability. This research has identified at least some factors—the six constructs—that help describe the CMI Framework. The validation work also revealed how the constructs are related empirically. A full analysis and explication of these relationships is beyond the scope of this dissertation. However, the modified model fits imply that the constructs are related and that together they underlie CMI instruction.

The importance of uncovering constructs and the beginning of theory development cannot be overstated. This research revealed key constructs that underlie CMI instruction. Knowing the constructs brought to light here will enable further work. As the CMI Observation Protocol is used to collect additional data, there will be the opportunity to verify and extend the theory that the six constructs named in this study are those that CMI instruction rests upon.

Validity and the Use of Multiple Categories of Evidence

Obtaining and reporting multiple measures of validity is both important and uncommon. The CMI Observation Protocol research used multiple categories of validity evidence, which is supported by the measurement concept of validity (American Educational Research Association, et al., 1999). According to Kilday and Kinzie (2009) in their review of nine recent protocols for measuring instruction, including the Inside the Classroom Observation Protocol, eight reported only one category of evidence, while one reported no evidence. The validity evidence from these studies came from the response process and internal structure categories, reporting inter-rater reliability or internal consistency reliability. The problem with using only one measure of validity is that it gives to narrow a view of the scope of issues in validity. For example inter-rater reliability gives assurance that the raters are consistent in their measurement, but gives no evidence that they are measuring important constructs. The same problem applies to internal consistency reliability, which tells whether like items are being rated in similar ways to each other and dissimilar items are being rated differently from each other. By obtaining and reporting multiple categories of evidence a more complete view of validity is acquired.

The measures of validity evidence for the CMI Observation Protocol came from three of the five AERA, et al. (1999) recommended categories. The three types of evidence, content, response process, and internal structure, were selected for their salience to the CMI Observation Protocol. It was important to be reasonably certain that the content of the CMI Framework was captured and well represented in the measure. Because multiple observers may introduce larger measurement error and reduce validity, two measures for response processes were used to strengthen the validity evidence. Good training increases the chances of multiple observers producing scores similar to each other, thus the training methods and processes were reported as evidence of the ability of the observers to become nearly interchangeable. The ICC was offered as quantitative evidence of the exchangeability of observers. The third category, internal structure evidence, also was strengthened by the use of two measures. Factor analysis served to show that the CMI Observation Protocol has a factor structure consistent with principles that describe the CMI Framework and that the items from the protocol group together sensibly. Internal consistency reliability was used to strengthen this category as well. However, by itself, internal consistency reliability is not enough to support a claim of validity.

Limitations

There is always a first iteration of any creation. The CMI Observation Protocol presented in this dissertation was not technically a first iteration because of the numerous changes it incurred before being used for classroom observation. However, this research represents the first time the protocol was used for data collection. A primary limitation to this research it is that the protocol has only been used in a relatively small number of classrooms. From both the internal structure and the theory-building aspects more data collection needs to be done to see if the constructs uncovered here hold true. A second important limitation is that the modified model limits the validity claim for the entire protocol. The adequate internal structure validity evidence can only be claimed for the 57 items included in the modified model.

Future Research

As mentioned in the limitations, future research needs to continue with the theory building that began in this study. Further data collection with the protocol will enable new results to be compared with these to determine if the same factor structure emerges. The relationships among the constructs need to be more closely analyzed as well and connected with the CMI Framework and the research literature. Building on this study, items that were eliminated from the models because of cross-loadings or low factor-loadings should be reviewed by the expert panel and be considered for revision. In that vein, the *math content* construct should be reviewed to determine if there is adequate representation of the CMI Framework in the CMI Observation Protocol. Additionally, the items that did not load adequately in these data need to be closely examined to determine if they represent constructs not revealed by the data from this study.

Another area for consideration in future research is the composition of the expert panel for reviewing the items. As teachers become more grounded in CMI instruction, they should be added to the expert panel. Teachers may be able to offer a more nuanced perspective of instruction that would be valuable in modifying or adding items to more accurately measure CMI instruction.

Conclusion

It was critically important that the CMI Observation Protocol was developed. The CMI Framework's potential for impacting instruction and student achievement could not be gauged without a measure specifically aligned to the Framework. Without accurately measuring instruction, the CMI Framework's impact would be a hunch. The Framework recommendations seem logical and there appears to be a connection between student scores and CMI instruction implementation. However, until instruction could be specifically observed and quantified, the impact of Framework instruction would remain a gut feeling. To be fair, instructional measurement had been occurring prior to the development of the CMI Observation Protocol. The results with a measuring tool that was not absolutely aligned were quite good. Now, with a protocol that shows strong content evidence, and adequate internal structure evidence and strong internal consistency reliability, the CMI Framework's power to transform instruction may be more fully known.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: Author.

Arbuckle, J. L. (2008). *Amos 17.0 user's guide*. Chicago, IL: SPSS.

Ball, D. L., & Cohen, D. K. (1996). Reform by the Book: What Is or Might Be the Role of Curriculum Materials in Teacher Learning and Instructional Reform? *Educational Researcher, 25*(9), 6-14. doi: 10.3102/0013189x025009006

Ball, D. L., & Rowan, B. (2004). Measuring instruction. *The Elementary School Journal, 105*(1).

Banilower, E. R. (2005). A study of the predictive validity of the LSC Classroom Observation Protocol. Chapel Hill NC: Horizon Research, Inc.

Blank, R. K. (2002). Using surveys of enacted curriculum to advance evaluation of instruction relation to standards. *Peabody Journal of Education, 77*(4), 86-121.

Borko, H., Kuffner, K. L., Arnold, S. C., Creighton, L., Stecher, B. M., Martinez, F., et al. (2007). Using artifacts to describe instruction: Lessons learned from studying reform-oriented instruction in middle school mathematics and science. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation (CSE), Graduate School of Education and Information Studies, University of California, Los Angeles.

Brophy, J., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 328-375). New York: MacMillan Publishing.

Brown, T. A. (2006). Confirmatory factor analysis for applied research. New York: Guilford Press.

Camburn, E., & Barnes, C. A. (2004). Assessing the validity of a language arts instruction log through triangulation. *The Elementary School Journal 105*(1), 49-73.

Cohen, D. K., & Ball, D. L. (1999). Instruction, capacity, and improvement (CPRE Research Report Series RR-043). In C. f. P. R. i. Education (Ed.), *CPRE Research Report Series RR-043*. Philidelphia: University of Pennsylvania Consortium for Policy Research in Education.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance:The mathematics reform in California. *Teachers College Record, 102*(2), 294-343.

Cohen, D. K., Peterson, P. L., Wilson, S., Ball, D. L., Putnam, R., Prawat, R., et al. (1990). Effects of state-level reform of elementary school mathematics curriculum on classroom practice (Final report, U.S. Department of Education, OERI Grant R117P8004). East Lansing: Michigan State University, Center for the Learning and Teaching of Elementary Subjects and National Center for Research on Teaching.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*(2), 119-142.

Cohen, D. K., & Spillane, J. P. (1992). Policy and Practice: The Relations between Governance and Instruction. *Review of Research in Education, 18,* 3-49.

Coleman, J., Campbell, E., Hobsen, C., McPartland, J., Mood, A., Weinfeld, F., et al. (1966). Equality of educational opportunity survey. Washington, DC: U.S. Govenment Printing Office.

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

Corcoran, T. B. (2007). Teaching Matters: How State and Local Policymakers Can Improve the Quality of Teachers and Teaching. CPRE Policy Briefs RB-48. Philadelphia, PA: Consortium for Policy Research in Education.

Corey, D. L. (2007). *Constructing and using a measure of teaching for mathematical proficiency*. Unpublished doctoral dissertation. University of Michigan. Ann Arbor.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. Retrieved October 8, 2009, from Education Policy Analysis Archives http://epaa.asu.edu/epaa/v8n1/

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199. doi: 10.3102/0013189x08331140

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112.

Fosnot, C. T. (1996). Constructivism: Theory, perspectives, and practice. New York: Teachers College, Columbia University. in the following EDTC 618 On-Line Reader Constructivism Chapter Four in Dani Baylor, Pavel Samsonov and Noel Smith (eds) A Collaborative Class Investigation into Telecommunications in Education 24 Jun 2000 retrieved August 24, 2010 from http://www.univie.ac.at/constructivism/edu.html.

George, G., & *Mallery*, P. (*2003*). SPSS for windows step by step: A simple guide and reference, 11.0 update. Boston, MA: Allyn & Bacon.

Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the New Standards for Educational and Psychological Teasting: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36*(3), 181-191.

Hanushek, E. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1*(1), 19-41.

Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis, 19*(2), 141-164.

Hendrickson, S., Hilton, S. C., & Bahr, D. (2008). The Comprehensive Mathematics Instruction (CMI) Framework: A new lens for examining teaching and learning in the mathematics classroom. *Utah Mathematics Teacher, Fall*, 44-52.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In J. Frank K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: a project of the National Council of Teachers of Mathematics*. Charlotte, NC: Information Age.

Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): Results from the TIMSS 1999 video study. *Educational Evaluation and Policy Analysis, 27*(2), 111-132.

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

Hilton, S. C., Hendrickson, S., and Bahr, D. (2010) *The Comprehensive Mathematics Instruction (CMI) Framework: A new lens for examining teaching and learning*. Paper presented at the Fourteenth Annual Conference of the Association of Mathematics Teacher Educators, January 30, 2010, Irvine, California.

Horizon Research Inc. (2000). Inside the Classroom Observation and Analytic Protocol. Chapel Hill, North Carolina.

Kilday, C. R., & Kinzie, M. B. (2009). An analysis of instruments that measure the quality of mathematics teaching in early childhood. *Early Childhood Education Journal*, *36*, 365-372.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions “at-scale”'. *Educational Assessment*, *13*(4), 267 — 300.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., et al. (2006). Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment. *CSE Technical Report #681*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

Mayer, D. P. (1999). Measuring instructional practice: Can policy makers trust survey data? . *Educational Evaluation and Policy Ananlysis*, *21*(1), 29-45.

McDiarmid, D., & Ball, D. (1989). The Teacher Education and Learning to Teach study: An occasion for developing a conception of teacher knowledge. Retrieved from http://ncrtl.msu.edu/http/tseries/ts891.pdf

National Council of Teachers of Mathematics. (1980). An agenda for action. Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation. Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). Professional standards for teaching mathematics. Reston, VA: Author.

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

National Council of Teachers of Mathematics. (2009). Guiding Principles for Mathematics Curriculum and Assessment. Retrieved August 25, 2010 from http://www.nctm.org/uploadedFiles/Math_Standards/NCTM Guiding Principles 6209.pdf

National Mathematics Advisory Panel. (2008). Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education.

National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, Swafford, J. and Findell, B. (Eds.). Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education (Ed.). Washington, DC: National Academy Press.

Ornstein, A. C., & Hunkins, F. P. (2009). *Psychological Foundations of Curriculum* (Ornstein & Hunkins, 2009)

Piaget, J. (1964). *Development and learning*. In R. E. Ripple and V. N. Rockcastle (Eds.), Piaget Rediscovered, pp. 7-20.

Piburn, M. D., & Sawada, D. (2000). Reformed Teaching Observation Protocol (RTOP): Reference manual (ACEPT Technical Report No. IN00-3). Tempe, AZ:

Arizona Collaborative for Excellence in the Preparation of Teachers.

Porter, A. C. (1989). A curriculum out of balance: The case of elementary mathematics. *Educational Researcher, 18*(5), 9-15.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrica, 73* 417–458.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *The Elementary School Journal 105*(1), 75-101

Rowan, B., & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the Study of Instructional Improvement. *Educational Researcher, 38*(2), 120-131.

Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of elementary schools. *Teachers College Record, 104*(8), 1525-1567.

Rowan, B., Harrison, D. M., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *The Elementary School Journal 105*(1), 103-127.

Rowan, B., Jacob, R., & Correnti, R. (2009). Using instructional logs to identify quality in educational settings. *New Directions for Youth Development, 121*(Spring 2010), 13-31.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., et al. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics, 102*(6), 245-253.

Sawada, D., Piburn, M. D., Turley, J., Falconer, K., Benford, R., Bloom, I., et al. (2000). Reformed Teaching Observation Protocol (RTOP) Training Guide (ACEPT Technical Report No. IN00-2). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Silver, E. A., Mesa, V. M., Morris, A. K., Star, J. R., & Benken, B. M. (2009). Teaching mathematics for understanding:An analysis of lessons submitted by teachers seeking NBPTS certification. *American Educational Research Journal*, *46*(2), 501-531.

Slaven, R. E., Lake, C., & Groff, C. (2009). Effective programs in Middle and High School mathematics: A best-evidence synthesis. *Review of Educational Research*, *79*(2), 839-911.

Spillane, J. P., & Zeuli, J. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, *21*(1), 1-27.

Stecher, B., Le, V.-N., Hamilton, L., Ryan, G., Robyn, A., & Lockwood, J. R. (2006). Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educational Evaluation and Policy Analysis*, *28*(2), 101-130. doi: 10.3102/01623737028002101

Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. Mathematical Thinking and Learning, 10(4), 313-340. DOI: 10.1080/10986060802229675

Stein, M.K., M.S.Smith, M.A. Henningsen, E.A. Silver (2000). Implementing Standards-Based Mathematics Instruction: A Case for Professional Development.

Stigler, J. W., Gonzales, P. A., Kawanka, T., Knoll, S., & Serrano, A. (1999). The TIMSS videotape classroom study: Methods and findings from an exploratory research project on Eighth-Grade mathematics instruction in Germany, Japan, and the United States. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999074 doi:http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=1999074

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: The Free Press.

Truxaw, M. P. & DeFranco, T. C. (2008). Mapping mathematics classroom discourse and its implications for models of teaching. *Journal for Research in Mathematics Education, 39*, 489-525.

VanTassel-Baska, J., Quek, C., & Feng, A. X. (2007). The Development and Use of a Structured Teacher Observation Scale to Assess Differentiated Best Practice. *Roeper Review, 29*(2), 84-92.

Von Glasersfeld, E. (1992). Questions and Answers about Radical Constructivim' In M.K. Pearsall (Ed.), Scope, Sequence and Coordination of Secondary School Science. Volume II. Relevant Research. Washington,DC: The National Science Teachers Association, pp. 169–182.

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

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). Looking inside the classroom: A study of K-12 mathematics and science education in the United States. Chapel Hill, NC: Horizon Research, Inc.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review*, 14(4), 490-495.

Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, *17*, 89-100.

Appendices

Appendix A: Comprehensive Mathematics Instruction Framework

DEVELOP UNDERSTANDING The goal of *Develop Understanding* is to surface student thinking, which leads to understanding of ideas, strategies, and representations relative to a selected mathematical purpose.		
LAUNCH	**EXPLORE**	**DISCUSS**
PURPOSE: Pose a task with: • Clear mathematical purpose aligned with a state or national standard or objective. • Multiple paths to solutions, and/or multiple solutions. **TEACHER ROLE:** ***Before*** 1. Identify mathematical objective(s) for lesson. 2. Select or design an appropriate task, e.g. new tasks, previously posed tasks, student-generated ideas, misconceptions, or questions. ***During*** 3. Activate student background knowledge. 4. Launch and clarify the task. **STUDENT ROLE:** 1. Actively listen. 2. Ask clarifying questions. 3. Access background knowledge	**PURPOSE:** Allow students to build understanding of the selected mathematical purpose by engaging in the task through: 1. Developing ideas. 2. Developing problem solving strategies. 3. Developing multiple representations using: • Appropriate manipulatives and/or technology • Charts, tables, diagrams, pictures, etc. **TEACHER ROLE:** ***Before*** 1. Anticipate student thinking. 2. Determine student grouping (individuals, pairs, small groups) ***During*** 3. Allow student exploration and discourse. 4. Facilitate exploration by asking questions to: • engage students in the task, • prompt or guide student exploration, • clarify mathematical thinking, • deepen student thinking. 5. Assess and select 3 to 5 ideas, strategies, and/or representations to share during *Discuss* phase. • Order by level of complexity to develop connections between ideas, strategies, and/or representations. • May choose incorrect examples to illustrate common misconceptions. **STUDENT ROLE:** 1. Engage in task. 2. Reflect on individual or group work by questioning, describing, explaining, and justifying thinking. 3. Ask "Does this make sense?"	**PURPOSE:** Develop student understanding of ideas, strategies, and representations by having students communicate, explain, and support their own thinking and interact with the thinking of their peers. **TEACHER ROLE:** ***Before*** 1. Anticipate the structure and flow of the discussion of selected ideas, strategies, and/or representations. ***During*** 2. Orchestrate discussion of selected ideas, strategies, and/or representations. 3. Help students understand criteria for judging ideas, strategies, and/or representations. 4. Assess while helping students clarify mathematical reasoning behind ideas, strategies, and/or representations. 5. Assess while helping students compare and connect ideas, strategies, and representations, using appropriate mathematical vocabulary. 6. Help students summarize and connect discussion to the selected mathematical purpose. ***After*** 7. Determine the next phase: • Remain within *Develop Understanding* phase, or • Move to *Solidify Understanding* phase. **STUDENT ROLE:** 1. Share and explain thinking. 2. Actively participate by listening, describing, complimenting, or comparing student work. 3. Question to clarify understanding.

<table>
<tr><th colspan="3">SOLIDIFY UNDERSTANDING
The goal of Solidify Understanding is to examine and extend student ideas, strategies, and representations, which leads to a development of concepts, algorithms, and tools.</th></tr>
<tr><th>LAUNCH</th><th>EXPLORE</th><th>DISCUSS</th></tr>
<tr>
<td>PURPOSE:
1. Pose a task with a focused
• idea,
• strategy, and/or
• representation
2. Designed to
• confirm,
• connect,
• generalize, and/or
• transfer
mathematical understanding.

TEACHER ROLE:
Before
1. Select an idea, strategy, and/or representation for focused instruction by:
• choosing a string of related problems, or
• choosing a problem with a string of related questions, or
• choosing a string of related tasks.
During
1. Activate student’s background knowledge from Develop Understanding phase.
2. Launch and clarify the task(s).

STUDENT ROLE:
1. Actively listen.
2. Ask clarifying questions.
3. Access background knowledge.</td>
<td>PURPOSE:
Engage students in task(s) to solidify understanding and gain ownership of the selected idea, strategy, and/or representation.

TEACHER ROLE:
Before
1. Anticipate student thinking and misconceptions and plan responses to guide focused discussion.
2. Determine structure of teaching cycle(s) and student grouping (individuals, pairs, small groups, whole group).
During
3. Facilitate and direct student understanding and ownership by:
a) exposing and eliminating misconceptions, and
b) asking questions to:
• prompt,
• clarify,
• guide,
• scaffold,
• probe, and/or
• connect mathematical thinking.
4. Continually refocus student thinking.

STUDENT ROLE:
1. Engage in task.
2. Raise “how,” “why,” “what if,” “so what,” “does this make sense,” “have I seen something like this before” questions.
3. Question, explain, and justify individual or group work using proper vocabulary.
4. Make connections among the string of related problems, questions, or tasks.
5. Make connections with previous learning.</td>
<td>PURPOSE:
Use student understanding of ideas, strategies, and representations to move students to an emerging understanding of concepts, algorithms, and tools.

TEACHER ROLE:
Before
1. Purposefully structure the focused discussion of ideas, strategies, and/or representations.
During
2. Ask probing/directed questions to draw out explicit and specific connections.
3. Confirm correct thinking.
4. Use direct instruction as appropriate.
5. Use language, conventions, and symbols of mathematicians.
6. Assess student understanding.
7. Help students recognize emerging concepts, algorithms, and tools.
After
8. Determine the next phase of the learning cycle:
• remain in Solidify Understanding,
• return to Develop Understanding with newly surfaced ideas, or
• move to Practice Understanding.

STUDENT ROLE:
1. Correctly use mathematical language, conventions, and symbols.
2. Explain and justify knowledge.
3. Describe connections between previous and current knowledge.
4. Generalize knowledge along a continuum from specific to abstract.
5. Transfer knowledge to new situations.</td>
</tr>
</table>

PRACTICE UNDERSTANDING

The goal of *Practice Understanding* is to allow students to refine and acquire fluency with concepts, algorithms, and tools, which leads to the development of definitions and properties, procedures, and models.

LAUNCH

PURPOSE:
1. Pose a task that re-engages students with one or more concepts, algorithms, or tools to acquire fluency as defined by:
- accuracy,
- efficiency,
- flexibility, and/or
- automaticity.

Note: Practice leads to uniqueness (individual refinement of concepts, algorithms, or tools) as well as sameness (fluency with common definitions, properties, procedures and models).

TEACHER ROLE:
Before
1. Identify the concept, algorithm, or tool to be practiced.
2. Select or design a vehicle with appropriate constraints to drive the practice, i.e., routines, games, worksheets, reviews, 10 minute math, etc., or
3. Embed the practice in the task of *Develop Understanding* or *Solidify Understanding.*

During
4. Connect the task to students' previous work.
5. Launch and clarify the task(s).

STUDENT ROLE:
1. Actively listen
2. Ask clarifying questions
3. Access background knowledge and learning from previous phases.

EXPLORE ⟷ DISCUSS

EXPLORE

PURPOSE:
Engage students in task(s) with appropriate constraints to hone, shape, and maintain concepts, algorithms, and tools.

TEACHER ROLE:
Before
1. The teacher determines when to monitor student work for accuracy, efficiency, and flexibility, e.g. during or after the exploration.

When Monitoring During Exploration
1. Monitor student work for fluency by:
- asking brief questions,
- eavesdropping on conversations, and
- visually scanning student work.

2. Monitor student work for opportunities for refinement by:
- asking questions to help students become more aware of what they are thinking and/or doing.
- encouraging efficient or flexible use of strategies.

3. Determine when to move to individual or group discussions.

When Monitoring After Exploration
1. Review student work for fluency by:
- correcting student work for accuracy,
- looking for common themes (conceptions and misconceptions) across samples of students' work.

STUDENT ROLE:
2. Reflect on work by asking:
- "Is this accurate?"
- "Do I understand this?"
- "Can I explain this?"
- "Where would I use this?"

DISCUSS

PURPOSE:
Give students personalized feedback that leads to independent fluency and to move students to an emerging understanding of definitions and properties, procedures, and models.

TEACHER ROLE:
Before
1. The teacher needs to be aware of the possible refinement that can occur during practice and ask questions to guide, mentor, and document this refinement.

When Giving Feedback During Exploration
1. Coach and mentor student work.
2. Provide individualized feedback.
3. Reinforce communication skills and computation.
4. Help students recognize emerging generalizations, procedures, and models.

When Giving Feedback After Exploration
1. Provide individualized feedback.
2. Identify emerging generalizations, procedures, and models.

After
1. Determine the next phase of the learning cycle
- remain in *Practice Understanding,*
- return to *Solidify Understanding,* or
- move on to *Develop Understanding.*

STUDENT ROLE:
1. Practice standard and invented algorithms, problem solving strategies, multiple representations, higher-level thinking, communication and fact recall.

3. Increase efficiency, flexibility, automaticity, and ability to justify work.

Appendix B: Examples of Measures

Table 11

A sampling of questions from the Teacher Questionnaire of TIMSS, Spillane & Zeuli, (1999)

Questions
In your mathematics lessons, how often do you usually ask students to do the following? (a) Explain the reasoning behind an idea (b) Represent and analyze relationships using tables, charts, or graphs (c) Work on problems for which there is no immediately obvious method of solution (d) Practice computational skills
In your mathematics lessons, how frequently do you do the following when a student gives an incorrect response during a class discussion? (d) Call on other students to get their responses and then discuss what is correct
In mathematics lessons, how often do students... (c) Work together as a class with the teacher teaching the whole class? (d) Work together as a class with students responding to one another?

Note: item response choices are listed as reported in Appendix A of Spillane, J. P., & Zeuli, J. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis, 21*(1), 1-27.

Table 12

Survey items about instructional practices, Cohen & Hill (2000)

Survey items
About how often do students in your class take part in the following activities during mathematics instruction:
Make conjectures and explore possible methods to solve a mathematical problem
Discuss different ways that they solve particular problems
Work in small groups on mathematics problems
Work on individual projects that take several days
Work on group investigations that extend for several days
Write about how to solve a problem in an assignment or test
Do problems that have more than one correct solution
Practice or take tests on computational skills
Work individually on mathematics problems from the text/workbook
Which statement best describes your use of a mathematics textbook? (Circle one)
A textbook is my main curriculum resource
I use other resources as much as I use the text
I mainly use curriculum resources other than the text
I do not use a textbook. I only use supplementary resources.

Note: 13 of 14 available in article

Table 13

Constructs forming scales on the SEC Survey, Blank (2000)

Constructs
Problem solving in mathematics.
Instructional activities in classrooms (e.g., small groups, manipulatives, investigations).
Mathematics and science content in classrooms (topics by cognitive demand or expectations).
Multiple assessment strategies in math and science.
Use of education technology and equipment.
Teacher preparation in subject and professional development.
Influences of policies and standards on practice.
Alignment of content taught with state assessments.
School and classroom conditions for teaching.

Table 14

Example of questions from the 3rd Grade Mathematics Log, Number Concepts Section A, (Rowan, Harrison, & Hayes, 2004)

Questions	Response choices
What did the target student work on today?	Writing, reading, or recognizing whole numbers, decimals, or fractions
	Counting
	Comparing or ordering two or more quantities
	Properties of whole numbers
	Factors, multiples, or divisibility with whole numbers
	Composing or decomposing (grouping) whole numbers or decimals into tenths, ones, tens, etc
	Identifying the values of the pieces in whole numbers or decimals
	The meaning of fractions
	Understanding equivalent fractions or working on reducing fractions
	Relationships between decimals and fractions
	Estimating the size of quantities or rounding off numbers
What did you or the target student use to work on the aspects of number concepts you checked?	Numbers or symbols
	Concrete materials
	Real-life situations or word problems
	Pictures or diagrams
	Tables or charts
	I made explicit links between two or more representations

Table 14 (continued)

Example of questions from the 3rd Grade Mathematics Log, Number Concepts Section A, *(Rowan, Harrison, & Hayes, 2004)*

Questions	Response choices
What was the target student asked to do during the work on number concepts?	Listen to me present the definition for a term or the steps of a procedure
	Perform tasks requiring ideas or methods already introduced to the student
	Assess a problem and choose a method to use from those already introduced to the student
	Perform tasks requiring ideas or methods NOT already introduced to the student
	Explain an answer or a solution method for a particular problem
	Analyze similarities and differences among representations, solutions, or methods
	Prove that a solution is valid or that a method works for all similar cases
Did the target student's work in number concepts include any of the following?	Orally answering recall questions
	Working on textbook, worksheet, or board work exercises for practice or review
	Working on problem(s) that have multiple answers or solution methods, or involve multiple steps
	Discussing ideas, problems, solutions, or methods in pairs or small groups
	Using flashcards, games, or computer activities to improve recall or skill
	Working extended explanations of mathematical ideas, solutions, or methods
	Working on an investigation, problem, or project over an extended period of time

Table 15

Elements of Mathematical Quality of Instruction, (Hill, et al., 2008)

1. Mathematics errors—the presence of computational, linguistic, representational, or other mathematical errors in instruction; Contains subcategory specifically for errors with mathematical language

2. Responding to students inappropriately—the degree to which teacher either misinterprets or, in the case of student misunderstanding, fails to respond to student utterance;

3. Connecting classroom practice to mathematics—the degree to which classroom practice is non-mathematical focus, such as classroom management, or activities that do not require mathematical thinking, such as students following directions to cut, color, and paste, but with no obvious connections between these activities and mathematical meaning(s);

4. Richness of the mathematics—the use of multiple representations, linking among representations, mathematical explanation and justification, and explicitness around mathematical practices such as proof and reasoning;

5. Responding to students appropriately—the degree to which teacher can correctly interpret students' mathematical utterances and address student misunderstandings;

6. Mathematical language—the density of accurate mathematical language in instruction, the use of language to clearly convey mathematical ideas, as well as any explicit discussion of the use of mathematical language.

Table 16

Reform Teaching Observation Protocol Domains and Items, (Piburn, M. D., & Sawada, D. 2000)

LESSON DESIGN AND IMPLEMENTATION

1. The instructional strategies and activities respected students' prior knowledge and the preconceptions inherent therein.
2. The lesson was designed to engage students as members of a learning community.
3. In this lesson, student exploration preceded formal presentation.
4. This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.
5. The focus and direction of the lesson was often determined by ideas originating with students.

CONTENT Propositional knowledge

6. The lesson involved fundamental concepts of the subject.
7. The lesson promoted strongly coherent conceptual understanding.
8. The teacher had a solid grasp of the subject matter content inherent in the lesson.
9. Elements of abstraction (i.e., symbolic representations, theory building) were encouraged when it was important to do so.
10. Connections with other content disciplines and/or real world phenomena were explored and valued.

CONTENT Procedural Knowledge

11. Students used a variety of means (models, drawings, graphs, concrete materials, manipulatives, etc.) to represent phenomena.
12. Students made predictions, estimations and/or hypotheses and devised means for testing them.
13. Students were actively engaged in thought-provoking activity that often involved the critical assessment of procedures.
14. Students were reflective about their learning.
15. Intellectual rigor, constructive criticism, and the challenging of ideas were valued.

CLASSROOM CULTURE Communicative Interactions

16. Students were involved in the communication of their ideas to others using a variety of means and media.
17. The teacher's questions triggered divergent modes of thinking.
18. There was a high proportion of student talk and a significant amount of it occurred between and among students.
19. Student questions and comments often determined the focus and direction of classroom discourse.
20. There was a climate of respect for what others had to say.

CLASSROOM CULTURE Student/Teacher Relationships

21. Active participation of students was encouraged and valued.
22. Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.
23. In general the teacher was patient with students.
24. The teacher acted as a resource person, working to support and enhance student investigations.
25. The metaphor "teacher as listener" was very characteristic of this classroom.

Table 17

Inside the Classroom Observation Protocol dimensions, (Weiss, et al., 2003)

Design

1. The design of the lesson incorporated tasks, roles and interaction consistent with investigative mathematics/science.
2. The design of the lesson reflected careful planning and organization.
3. The instructional strategies and activities used in this lesson reflected attention to students experience, preparedness, prior knowledge, and /or learning styles.
4. The resources available In this lesson contributed to accomplishing the purposes of the instruction.
5. The instructional strategies and activities reflected attention to issues of access, equity, and diversity for students (e.g., cooperative learning, language-appropriate strategies/materials).
6. The design of the lesson encouraged a collaborative approach to learning among the students.
7. Adequate time and structure were provided for "sense-making."
8. Adequate time and structure were provided for wrap-up.

Implementation

1. The instructional strategies were consistent with investigative mathematics/science.
2. The teacher appeared confident in his/her ability to teach mathematics/science.
3. The teacher's classroom management style/strategies enhanced the quality of the lesson.
4. The pace of the lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.
5. The teacher was able to "read" the students' level of understanding and adjusted instruction accordingly.
6. The teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem solving (e.g., emphasized higher-order questions, appropriately used "wait time," identified prior conceptions and misconceptions).

Mathematics/Science content

1. The mathematics/science content was significant and worthwhile.
2. The mathematics/science content was appropriate for the developmental levels of the students in this class.
3. Teacher-provided content information was accurate.
4. Students were intellectually engaged with important ideas relevant to the focus of the lesson.
5. the teacher displayed an understanding of mathematics/science concepts (e.g. in his/her dialogue with students).
6. Mathematics/science was portrayed as a dynamic body of knowledge continually enriched by conjecture, investigation analysis, and/or proof/justification.
7. Elements of mathematics/science abstraction (e.g., symbolic representations, theory building) were included when it was important to do so.
8. Appropriate connections were made to other areas of mathematics/science to other disciplines, and/or to real world contexts.
9. The degree of "sense-making" of mathematics/science content within this lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.

Classroom Culture

1. Active participation of all was encourages and valued.
2. There was a climate of respect for students' ideas, questions, and contributions.
3. Interactions reflected collegial working relationships among students (e.g., students worked together, talked with each other about the lesson).
4. Interactions reflected collaborative working relationships between teacher and students.
5. The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/or propositions.
6. Intellectual rigor, constructive criticism, and the challenging of ideas were evident.

Table 18

Inside the Classroom Observation Protocol Rating Structure

Level 1: Ineffective Instruction

There is little or no evidence of student thinking or engagement with important ideas of mathematics/science. Instruction is highly unlikely to enhance students' understanding of the discipline or to develop their capacity to successfully "do" mathematics or science. Lesson was characterized by either (select one below):

a. Passive "Learning"

Instruction is pedantic and uninspiring. Students are passive recipients of information from the teacher or textbook; material is presented in a way that is inaccessible to many of the students.

b. Activity for Activity's Sake

Students are involved in hands-on activities or other individual or group work, but it appears to be activity for activity's sake. Lesson lacks a clear sense of purpose and/or a clear link to conceptual development.

Level 2: Elements of Effective Instruction

Instruction contains some elements of effective practice, but there are serious problems in the design, implementation, content, and/or appropriateness for many students in the class. For example, the content may lack importance and/or appropriateness; instruction may not successfully address the difficulties that many students are experiencing, etc. Overall, the lesson is very limited it its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully "do" mathematics/science.

Level 3: Beginning Stages of Effective Instruction. (select one below)

Low Solid High

Instruction is purposeful and characterized by quite a few elements of effective practice. Students are, at times, engaged in meaningful work, but there are weaknesses, ranging from substantial to fairly minor, in the design, implementation, or content of instruction. For example, the teacher may short-circuit planned explorations by telling students what they "should have found"; instruction may not adequately address the needs of a number of students; or the classroom culture may limit the accessibility or effectiveness of the lesson. Overall, the lesson is somewhat limited in its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully "do" mathematics/science.

Level 4: Accomplished, Effective Instruction

Instruction is purposeful and engaging for most students. Students actively participate in meaningful work (e.g., investigations, teacher presentations, discussions with each other of the teacher, reading). The lesson is well-designed and the teacher implements it well, but adaptation of content or pedagogy in response to student needs and interests is limited. Instruction is quite likely to enhance most students' understanding of the discipline and to develop their capacity to successfully "do" mathematics/science.

Level 5: Exemplary Instruction

Instruction is purposeful and all students are highly engaged most or all of the time in meaningful work (e.g., investigation, teacher presentations, discussions with each other or the teacher, reading). The lesson is well-designed and artfully implemented, with flexibility and responsiveness to students' needs and interests. Instruction is highly likely to enhance most students' understanding of the discipline and to develop their capacity to successfully "do" mathematics/science.

Table 19

Reform elements measured by artifact collection in a "Scoop Notebook", Borko et al. (2007)

Elements
1. Grouping
2. Structure of Lessons
3. Multiple Representations
4. Use of Mathematical Tools
5. Cognitive Depth
6. Mathematical Discourse Community
7. Explanation/Justification
8. Problem Solving
9. Assessment
10. Connections/Applications

Table 20

Artifact analysis criteria, (Silver, et al., 2009)

Topics	Cognitive Demand of Tasks	Pedagogical Features of Tasks
Number and operations	High demand: Students • Explain, justify, compare, assess • Make decisions and choices to plan or formulate questions/problems	Require multi-person collaboration and discourse
Algebra and functions	• Are creative • Translate one representation to another or interpret meaning across two or more representations	Require mathematical reasoning and explanation
Measurement		Consider applications outside of mathematics
Geometry	Low demand: Students • Make routine application of known procedures • Have challenging tasks made routine	Employed technology
Data analysis, statistics, and probability	• Describe procedures or non-mathematical aspects	Employed hands-on materials

Appendix C: Protocol Development Tools

CMI Framework Examination Table			
Learning Cycle Phase	**→ Develop Understanding**	**Solidify Understanding**	**Practice Understanding**
Teaching Cycle Phase ↓	Purpose: to surface student thinking leading to understandings of ideas, strategies, & representations relative to the selected mathematical purpose	Purpose: to examine and extend ideas, strategies & representations to develop concepts, algorithms, & tools	Purpose: to refine and acquire fluency with concepts, algorithms, & tools to develop definitions, properties, procedures, & models
Launch	**Develop Understanding**	**Solidify Understanding**	**Practice Understanding**
	Pose task(s) appropriate for the purpose of surfacing student ideas, (aligned with a state or national standard or objective)	Pose a task with a focus on an idea, strategy, or representation to solidify (confirm, connect, generalize, transfer) mathematical understanding (still aligned with a state or national standard or objective)	Pose a task that re-engages students with one or more concepts, algorithms, or tools to acquire fluency (accuracy, efficiency, flexibility, automaticity) (still aligned with a state or national standard or objective)
	Pose a task with multiple paths to solutions, and/or multiple solutions	Pose task(s) with a string of related problems, a problem with a string of related questions, or a string of related tasks	Pose a task with appropriate constraints to drive the practice [Embed the practice in the task of *Develop Understanding* or *Solidify Understanding*—how to capture this?]
	Clarify the task (teacher)	Clarify the task (teacher)	Clarify the task (teacher)
	Ask clarifying questions (students)	Ask clarifying questions (students)	Ask clarifying questions (students)
	Teacher activates	Teacher activates	Teacher connects the

	student background knowledge	student background knowledge from *Develop Understanding*	task to students' previous work
	Students access background knowledge	Students access background knowledge from *Develop Understanding*	Students access background knowledge and learning from previous phases
	Students actively listen	Students actively listen	Students actively listen
Explore	**Develop Understanding**	**Solidify Understanding**	**Practice Understanding** The ***Explore*** **and** ***Discuss*** phases of the teaching cycle have flow between them at the Practice Understanding phase. There is little need to have a whole group discussion as students are individually refining while simultaneously gaining fluency. Discuss phase takes place primarily with individuals and small groups.
	Teacher anticipates student thinking	Teacher anticipates student thinking and misconceptions and plans responses to guide focused discussion	The teacher needs to be aware of the possible refinement that can occur during practice and ask questions to guide, mentor, and document this refinement.
	Determine student grouping(s)	Determine student grouping(s)	Determine when to monitor student work for accuracy, efficiency, and flexibility (during or

			after the exploration)[both?]
	Allow student exploration		
	Allow student discourse		
		Facilitate and direct student understanding & ownership by exposing and eliminating misconceptions	
	Teacher questions for engagement, guidance, clarity of student thinking, depth of student thinking	Facilitate and direct student understanding & ownership by asking questions	
		Teacher continually refocuses student thinking	
	Engage students in task Students engage in task	Engage students in task Students engage in task	Engage students in task Students engage in task (Students practice standard and invented algorithms, problem solving strategies, multiple representations, higher-level thinking, communication and fact recall)
	Students reflect on	Students question,	

	work by questioning, describing, explaining, and justifying thinking	explain, justify individual or group work using proper vocabulary	
		Students make connections among the string of related problems, questions, or tasks	
		Students make connections with previous learning	
	The teacher assesses usefulness of student ideas, strategies, and representations and selects 3-5 to share during the *discuss phase* The teacher orders selected student ideas, strategies, representations by complexity to develop connections		Determine when to move to individual or group discussions
			Monitor for ***fluency*** **during** exploration by: • asking brief questions • Eavesdropping on conversations • Visually scanning student work
			Monitor for ***fluency*** **after** exploration by: • Correcting student work for accuracy • Looking for

			common themes across samples of students' work
			Monitor for ***refinement*** **during** by: Asking questions to help students become more aware of what they are thinking and/or doing Encouraging efficient or flexible use of strategies
Discuss	**Develop Understanding**	**Solidify Understanding**	**Practice Understanding** The ***Explore*** **and** ***Discuss*** phases of the teaching cycle have flow between them at the Practice Understanding phase. There is little need to have a whole group discussion as students are individually refining while simultaneously gaining fluency. Discuss phase takes place primarily with individuals and small groups.
	Teacher anticipates the structure of the discussion	Purposefully structure the focused discussion of ideas, strategies, and/or representations	The teacher needs to be aware of the possible refinement that can occur during practice and ask questions to guide, mentor, and document this refinement
	Teacher orchestrates a discussion of the ideas, strategies, and representations		
	Students communicate, explain, and support their own thinking		
	Students interact with the thinking of peers		

	The teacher helps students understand criteria for judging ideas, strategies, and/or representations		
	Teacher assesses student understanding of ideas, strategies, representations	Teacher assesses student understanding Teacher confirms correct thinking	Giving **feedback *during*** exploration: Coach and mentor student work Provide individualized feedback
	Teacher and/or students help other students clarify mathematical reasoning behind ideas, strategies, & representations	Teachers ask probing/directed questions to draw out explicit and specific connections	
	Teacher helps students compare and connect ideas, strategies, & representations	The teacher helps students recognize emerging concepts, algorithms, tools	The teacher helps students recognize emerging generalizations, procedures, and models
	The teacher helps students use appropriate mathematical vocabulary	The teacher and students use language, conventions, and symbols of mathematicians	The teacher reinforces communication skills and computation
	Teacher helps students summarize and connect discussion to the selected mathematical purpose	Students generalize knowledge along a continuum from specific to abstract	

	Teacher determines the next phase: moving to *Solidify* or remaining in *Develop*		
		Use direct instruction as appropriate	
			Students increase efficiency, flexibility, automaticity, and ability to justify work

Interview Protocol
Former Observers

Thinking about your observation experiences using the Inside the Classroom Observation Protocol, and after reviewing the Inside the Classroom Observation Protocol:

1. What would you say were the strengths, if any, of the Inside the Classroom Observation Protocol?

Possible follow-ups:

a. Was the structure or lay-out a strength? In what way(s)?
b. What parts of observed reform-type mathematics instruction did it seem to measure well? That is, when you saw particular practices were there items on the protocol that clearly described those practices?
c. Which items do you recall as being very descriptive?

2. What would you say were the weaknesses, if any, of the Inside the Classroom Observation Protocol?

Possible follow-ups:

a. Was the structure or lay-out a weakness? In what way(s)?
b. What parts of observed reform-type mathematics instruction did it <u>NOT</u> seem to measure well? That is, are there reform-type mathematics practices you observed that you felt should be recorded, but that the Inside the Classroom Observation Protocol had no items for, or for which items only partially described the practice?

3. What insights or advice might you give as a new observation protocol is being developed that would make it “observer-friendly”?

Expert Panel Focus Group Protocol

In a CMI launch the teacher should pose a task. There are teacher & student roles in clarifying the task and activating background knowledge & knowledge from previous phases.

1. Focusing on just the items in the Launch section, how well (to what degree) do those items represent what you would expect to see in a launch of a CMI lesson (including all phases of the Learning Cycle)?

In a CMI exploration sense-making is emphasized. The teacher uses discourse to facilitate student movement with increasing understanding through the task(s), students use discourse to increase understanding as they move through the task(s), teachers make use of student thinking and there are teacher & student roles for engaging in the task(s).

2. Focusing on just the items in the Explore section, how well (to what degree) do those items represent what you would expect to see in the explore phase of a CMI lesson (including all phases of the Learning Cycle)?

A CMI whole class discussion has a coherent structure focused on important mathematics, makes use of student thinking uncovered in the explore phase, students and the teacher question, explain, or justify ideas, strategies, and representations, etc.(from the math understanding continuum).
Also in the discussion, the teacher confirms correct thinking, connections are made explicit, appropriate mathematical language is used, and the discussion concludes with the students and teacher knowing what has been discussed.

3. Focusing on just the items in the Discuss section, how well (to what degree) do those items represent what you would expect to see in the discuss phase of a CMI lesson (including all phases of the Learning Cycle)?

4. Focusing on just the items in the Classroom Climate section how well (to what degree) do those items represent what you would expect to see in a classroom implementing CMI?

5. Focusing on just the items in the Mathematical Content section how well (to what degree) do those items represent what you believe are important in a CMI lesson?

6. Focusing on just the items in the Lesson Coherence/General Pedagogy section how well (to what degree) do those items represent what you would expect to see in a CMI lesson?

7. Overall, what would your recommendations be for further development of the protocol?
   a. Are there any sections or items that might be eliminated?
   b. Are there any sections or items that could be incorporated into another section?
   c. If there are areas you feel are important in measuring CMI, but missing, what would those be?

Appendix D: The CMI Observation Protocol

Comprehensive Mathematics Instruction (CMI) Observation Protocol
Pre-observation information

Teacher name___________________________________

Gender F M

Grade level(s)___________

Teacher's statement:

My purpose for today's lesson is (what I want students to walk away with):

It fits into the State core in the following way(s):

Briefly tell about students in your class who may have special learning needs that you will accommodate in this lesson: (e.g. "One boy is ELL in the silent stage. One girl already demonstrates understanding of this topic.")

Comprehensive Mathematics Instruction (CMI) Observation Protocol

Observation Date____________________________ Time: Start_________End_________

School____________________________________ District_________________________

Teacher name___________________________________ Observer _______________________

Grade level_______

Lesson Narratives

1. **Lesson Structure (running record)**

2. **Lesson Summary (description & observer commentary)**

Ratings: Launch

Indicator Item		Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
Task(s)							
	1. The task selected had multiple entry points, accommodating a variety of student needs	**1**	**2**	**3**	**4**	**5**	**6**
D	2. The task posed had multiple solutions or multiple paths to solutions	**1**	**2**	**3**	**4**	**5**	**6**
S	3. The task posed consisted of a string of related problems, questions, or tasks	**1**	**2**	**3**	**4**	**5**	**6**
P	4. The task posed consisted of a game or worksheet	**1**	**2**	**3**	**4**	**5**	**6**
D	5. The task potentially promoted the surfacing of numerous student ideas or misconceptions	**1**	**2**	**3**	**4**	**5**	**6**
S	6. The task potentially focused students on examining a selected idea, strategy or representation to confirm, connect, generalize and/or transfer mathematical understanding	**1**	**2**	**3**	**4**	**5**	**6**
P	7. The task potentially promoted practice of a(n) concept, algorithm or tool to acquire fluency (accuracy, efficiency, flexibility, and/or automaticity)	**1**	**2**	**3**	**4**	**5**	**6**
Teacher							
	8. The teacher activated or built students' **mathematical background knowledge** necessary for the task	**1**	**2**	**3**	**4**	**5**	**6**
	9. The teacher activated or built students' **contextual background knowledge** necessary for the task	**1**	**2**	**3**	**4**	**5**	**6**
	10. The teacher activated or built students' **connections** between the current lesson and previous work	**1**	**2**	**3**	**4**	**5**	**6**

11. The teacher clarified the task for students (to a degree matched to the purpose of the phase) before they began the exploration	**1 2 3 4 5**	**6**
Students		
12. Students responded appropriately during the launch: they listened actively (eyes on speaker, etc.) and/or they asked clarifying questions before beginning, if needed	**1 2 3 4 5**	**6**
13. Students appeared to know what the task entailed, (i.e. affect, body language) and were prepared to immediately engage in the task (participated/supported participation)	**1 2 3 4 5**	**6**

Launch Synthesis Rating
(This rating is holistic, based on how well the launch used CMI principles to prepare students for exploring)

1	2	3	4	5
Not at all The launch did not reflect use of CMI Principles	**Slightly** The launch reflected *slight* use of CMI principles	**Adequately** The launch reflected *adequate* use of CMI principles	**Commendably** The launch reflected *commendable* use of CMI principles	**Optimally** The launch reflected *optimal* use of CMI principles

Observer notes: (If there is a mismatch between item ratings and synthesis rating, what was the determining factor(s) leading to the synthesis rating?

Ratings: Explore

Indicator Item	Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
Teacher						
1. The teacher facilitated student engagement in exploration (e.g., through task structure, reminders of norms, pre-alerting)	1	2	3	4	5	6
2. The teacher made appropriate tools available to assist students with mathematical understanding	1	2	3	4	5	6
3. The teacher differentiated within the task to meet student needs	1	2	3	4	5	6
4. The teacher allowed adequate exploration prior to engaging with students	1	2	3	4	5	6
5. Student groupings (individual, pairs, small group) facilitated the exploration of the task(s)	1	2	3	4	5	6
6. The teacher maintained the task at an appropriate level to meet the purpose of the phase (e.g. supplying only necessary scaffolding, allowing purposeful struggle, avoiding "telling")	1	2	3	4	5	6
7. The teacher was able to "read" the students' level of understanding and responded appropriately (includes moving to discussion at appropriate times)	1	2	3	4	5	6
8. The teacher strategically monitored student exploration (e.g. not needing to get to every individual/group)	1	2	3	4	5	6
9. The teacher monitored student thinking during exploration by asking brief questions, eavesdropping on conversations, and/or visually scanning student work	1	2	3	4	5	6
10. The teacher noted/recorded student thinking pertinent to use in the discussion (e.g. examples, common misconceptions)	1	2	3	4	5	6
11. The teacher's questions facilitated exploration (e.g., by engaging students in the task, prompting or guiding exploration, clarifying student thinking, deepening student	1	2	3	4	5	6

thinking)						
12. The teacher's questions facilitated and/or directed student understanding & ownership (e.g., by prompting, clarifying, guiding, scaffolding, probing, and/or connecting mathematical thinking)	**1**	**2**	**3**	**4**	**5**	**6**
S 13. The teacher's questions facilitated student **recognition** of misconceptions	**1**	**2**	**3**	**4**	**5**	**6**
S 14. The teacher's questions advanced student thinking that facilitated the **elimination** of misconceptions	**1**	**2**	**3**	**4**	**5**	**6**
15. The teacher's questions helped students to become more aware of what they were thinking and/or doing	**1**	**2**	**3**	**4**	**5**	**6**
16. The teacher's questions helped students to refine their thinking/understanding	**1**	**2**	**3**	**4**	**5**	**6**
Students						
17. Students persist and maintain effort on the exploration task	**1**	**2**	**3**	**4**	**5**	**6**
18. Students display positive affect about the task (e.g. enthusiasm, curiosity, etc.)	**1**	**2**	**3**	**4**	**5**	**6**
19. The students used appropriate tools to assist with mathematical understanding (e.g. drawings, manipulatives, etc.)	**1**	**2**	**3**	**4**	**5**	**6**
20. Students engaged in sensemaking by Asking themselves, "Is this accurate," "do I understand this," "can I explain this," "where would I use this" (evidence in journal, work, conversation)	**1**	**2**	**3**	**4**	**5**	**6**
D S 21. Students engaged in sensemaking by questioning, describing, explaining, and/or justifying their own or others' work	**1**	**2**	**3**	**4**	**5**	**6**
D S 22. Students engaged in sensemaking by making connections among the string of related problems, questions, or tasks, and/or with previous learning	**1**	**2**	**3**	**4**	**5**	**6**

Explore Synthesis Rating:

1	2	3	4	5
Not at all The explore phase did not reflect use of CMI Principles	**Slightly** The explore phase reflected *slight* use of CMI principles	**Adequately** The explore phase reflected *adequate* use of CMI principles	**Commendably** The explore phase reflected *commendable* use of CMI principles	**Optimally** The explore phase reflected *optimal* use of CMI principles

Observer notes: (If there is a mismatch between item ratings and synthesis rating, what was the determining factor(s) leading to the synthesis rating?)

Ratings: Discuss

Indicator Item	**Not at all**	**Slightly**	**Adequately**	**Commendably**	**Optimally**	**Not applicable**
Teacher						
1. The content of the discussion reflected purposeful selection of student work from the exploration that focused on the lesson's purpose	**1**	**2**	**3**	**4**	**5**	**6**
2. The discussion had a logical, coherent structure that built on the exploration (e.g. ordered by level of complexity)	**1**	**2**	**3**	**4**	**5**	**6**
3. The discussion was ordered such that development of connections was facilitated (sequencing builds to the "aha")	**1**	**2**	**3**	**4**	**5**	**6**
4. Student thinking/work about the mathematics was represented in the whole class discussion	**1**	**2**	**3**	**4**	**5**	**6**
5. The teacher asked probing/directed questions to draw out explicit and specific connections (the teacher doesn't leave the "aha" to chance)	**1**	**2**	**3**	**4**	**5**	**6**
6. The teacher used appropriate "talk moves" (revoiced, used wait time, asked for students to restate, dis/agree, add to)	**1**	**2**	**3**	**4**	**5**	**6**
7. The teacher used mathematical language, conventions, and symbols accurately	**1**	**2**	**3**	**4**	**5**	**6**
8. Explicit/Direct instruction was used when appropriate (e.g. for giving information, conventions, norms)	**1**	**2**	**3**	**4**	**5**	**6**
S P 9. The teacher helped students identify correct thinking about the lesson's focus (e.g., through criteria, consensus building, etc.)	**1**	**2**	**3**	**4**	**5**	**6**
10. The discussion closed with a clear sense of what had been discussed	**1**	**2**	**3**	**4**	**5**	**6**

Ratings: Discuss, Students

Indicator Item	Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
Students						
11. Students communicated, explained, and/or supported their own thinking during the whole class discussion	**1**	**2**	**3**	**4**	**5**	**6**
12. Students questioned the thinking of peers during the whole class discussion	**1**	**2**	**3**	**4**	**5**	**6**
13. Students elaborated on the thinking of peers during the whole class discussion	**1**	**2**	**3**	**4**	**5**	**6**
14. Non-sharing students actively listened during the whole-class discussion (showed readiness to restate, explain, add to, dis/agree)	**1**	**2**	**3**	**4**	**5**	**6**
15. Students probed other students to articulate their reasoning behind ideas, strategies, representations, etc. for the purpose of advancing understanding	**1**	**2**	**3**	**4**	**5**	**6**
16. Students' used mathematical language, conventions, and symbols accurately (appropriate for grade level)	**1**	**2**	**3**	**4**	**5**	**6**
17. Students connected discussion points with an important mathematical purpose	**1**	**2**	**3**	**4**	**5**	**6**

Discuss Synthesis Rating:

1	2	3	4	5
Not at all The discuss phase did not reflect use of CMI Principles	**Slightly** The discuss phase reflected *slight* use of CMI principles	**Adequately** The discuss phase reflected *adequate* use of CMI principles	**Commendably** The discuss phase reflected *commendable* use of CMI principles	**Optimally** The discuss phase reflected *optimal* use of CMI principles

Observer notes: (If there is a mismatch between item ratings and synthesis rating, what was the determining factor(s) leading to the synthesis rating?)

Ratings: Classroom Climate

Indicator Item	Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
1. The teacher's classroom management style/strategies enhanced the quality of the lesson (e.g. clear procedures, effective transitions)	**1**	**2**	**3**	**4**	**5**	**6**
2. There was a climate of respect for students' ideas, questions, and contributions	**1**	**2**	**3**	**4**	**5**	**6**
3. The climate of the classroom encouraged students to generate ideas, questions, and/or conjectures	**1**	**2**	**3**	**4**	**5**	**6**
4. Previous student thinking is evident in the classroom (e.g. charts of previous discussion)	**1**	**2**	**3**	**4**	**5**	**6**
5. Interactions reflected collaborative working relationships **between teacher and students**	**1**	**2**	**3**	**4**	**5**	**6**
6. Interactions reflected collaborative working relationships **among students**	**1**	**2**	**3**	**4**	**5**	**6**
7. Active participation of **all** was encouraged and valued	**1**	**2**	**3**	**4**	**5**	**6**
8. Materials were organized and accessible to students	**1**	**2**	**3**	**4**	**5**	**6**

Classroom Climate Synthesis Rating:

1	2	3	4	5
Not at all The classroom climate did not reflect use of CMI Principles	**Slightly** The classroom climate reflected *slight* use of CMI principles	**Adequately** The classroom climate reflected *adequate* use of CMI principles	**Commendably** The classroom climate reflected *commendable* use of CMI principles	**Optimally** The classroom climate reflected *optimal* use of CMI principles

Observer notes: (If there is a mismatch between item ratings and synthesis rating, what was the determining factor(s) leading to the synthesis rating?)

Ratings: Mathematical Content

Indicator Item	Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
1. The teacher displayed confidence in his/her ability to teach mathematics	**1**	**2**	**3**	**4**	**5**	**6**
2. The content information provided by the teacher was accurate	**1**	**2**	**3**	**4**	**5**	**6**
3. The mathematics content was significant and worthwhile	**1**	**2**	**3**	**4**	**5**	**6**
4. The mathematics content was appropriate for the developmental levels of the students in the class	**1**	**2**	**3**	**4**	**5**	**6**
5. The teacher displayed an understanding of the mathematics being taught (e.g., in his/her dialogue with students)	**1**	**2**	**3**	**4**	**5**	**6**
6. The teacher displayed an understanding of student trajectory through the mathematics (e.g. by connections activated/built, selection of student work, discussion)	**1**	**2**	**3**	**4**	**5**	**6**

Mathematics Content Synthesis Rating:

1	2	3	4	5
Not at all The mathematics content did not reflect use of CMI Principles	**Slightly** The mathematics content reflected *slight* use of CMI principles	**Adequately** The mathematics content reflected *adequate* use of CMI principles	**Commendably** The mathematics content reflected *commendable* use of CMI principles	**Optimally** The mathematics content reflected *optimal* use of CMI principles

Observer notes: (If there is a mismatch between item ratings and synthesis rating, what was the determining factor(s) leading to the synthesis rating?)

Ratings: Lesson Coherence

Indicator Item	Not at all	Slightly	Adequately	Commendably	Optimally	Not applicable
1. There was a clear mathematical purpose intended for the lesson (adjustments may be made as the lesson unfolds)	**1**	**2**	**3**	**4**	**5**	**6**
2. The task(s) aligned with the mathematical purpose	**1**	**2**	**3**	**4**	**5**	**6**
3. The task was neither too easy nor too difficult for the majority of students)	**1**	**2**	**3**	**4**	**5**	**6**
4. The task's level of constraint was appropriate for the purpose of the lesson (from very open to very constrained)	**1**	**2**	**3**	**4**	**5**	**6**
5. The resources available in this lesson contributed to accomplishing the purposes of the instruction	**1**	**2**	**3**	**4**	**5**	**6**
6. The pace of the lesson was appropriate for the needs of the students (The right amount of time was provided for exploring the task, holding the discussion, etc.)	**1**	**2**	**3**	**4**	**5**	**6**
7. The structure of the task encouraged appropriate collaboration among students	**1**	**2**	**3**	**4**	**5**	**6**
8. The student groupings allowed for the mathematical purpose to occur (e.g. proximity, size of group)	**1**	**2**	**3**	**4**	**5**	**6**
9. The instructional strategies and activities used in the lesson reflected attention to students' experience, preparedness, prior knowledge, and/or learning styles	**1**	**2**	**3**	**4**	**5**	**6**
10. An appropriate balance of teacher-talk and student talk was achieved	**1**	**2**	**3**	**4**	**5**	**6**

Lesson Coherence Synthesis Rating:

1	2	3	4	5
Not at all The lesson coherence did not reflect use of CMI Principles	**Slightly** The lesson coherence reflected *slight* use of CMI principles	**Adequately** lesson coherence reflected *adequate* use of CMI principles	**Commendably** The lesson coherence reflected *commendable* use of CMI principles	**Optimally** The lesson coherence reflected *optimal* use of CMI principles

Capsule Rating: Overall Lesson Alignment with the CMI Framework

CMI Instruction is a balance of teacher directed and student centered instruction, with a task as the focal point from which important mathematics is explored and discussed, in ways consistent with the appropriate learning cycle phase. After the launch, control of the task should be released to students, with the teacher providing an active supporting role in helping students construct understandings consistent with the larger mathematical community.

____Level 1: No implementation of the CMI Framework (May still be "good" instruction)

This non-CMI lesson was characterized by (select one):

Teacher Centered
__a. Direct/explicit instruction
__b. Lecture/demonstration
__c. Traditional drill and practice
__d. Other teacher centered (describe)

Student Centered
__e. Activity unconnected to mathematics
__f. Discovery learning

__g. other student centered (describe)

Level 2: Very limited implementation of the CMI Framework

The launch, explore, and discuss phases of the teaching cycle may or may not *all* be present. If all of the teaching cycle phases are attempted, they **do not** accomplish their purposes and there are numerous and serious flaws throughout the lesson. For example, there are flaws in the task alignment; or tasks that are not likely to accomplish an important mathematical purpose; or inattention to student thinking revealed in the lesson; or consistently missing opportunities to press students to explain or justify; or incorrect mathematics; or classroom climate or pedagogy that impedes active student participation for some. Student control over the task and the discussion are very limited. Overall, students are *unlikely* to enhance their understanding of mathematics.

____Level 3: Emergent implementation of the CMI Framework

The teaching cycle structure (launch, explore, discuss), is attempted. The purposes of each phase are met to a limited degree. For example, the launch likely will need to be clarified before students are prepared for the explore. The explore will last too long or not long enough or have a task that is not aligned with a mathematical purpose or the teacher will tell the students "what they should have found". The discussion may be reduced to a summary statement by the teacher or a question and answer session. Student thinking is not probed or pressed. Student control over the task and the discussion are limited. Overall, students *may or may not* deepen their understanding of mathematics.

____Level 4: Adequate implementation of the CMI Framework

The launch, explore, and discuss phases of the teaching cycle are all present. The launch prepares students to successfully engage in the explore, although there may be some re-teaching once the explore has begun. In the explore the task is somewhat aligned to the mathematical purpose, or the teacher may "rescue" students by answering his/her own questions, or does not relinquish control of the task appropriately to students to explore. The discussion occurs, and student thinking is used on a limited basis, but there are many missed opportunities to use student thinking or press to have students justify or explain. Students are engaged but still largely rely on the teacher for direction. Students are *somewhat likely* to deepen their understanding of mathematics.

____Level 5: Proficient implementation of the CMI Framework

The teaching cycle structure (launch, explore, discuss) is solidly in use. The teacher prepares students well in the launch to engage in the explore. The task selected is well aligned to the mathematical purpose. The teacher appropriately relinquishes control to the students to explore the task. The teacher has anticipated student thinking but may struggle some to make full use of student thinking as it emerges in the lesson. However, the teacher makes use of many opportunities to use student thinking, and to press students to justify or explain. Students are engaged throughout the lesson in important mathematics. Students demonstrate ownership for learning for some of the lesson but also slip into relying on the teacher for direction at other times. Students are *likely* to deepen their understanding of mathematics.

____Level 6: Accomplished implementation of the CMI Framework

The launch, explore, and discuss phases of the teaching cycle are used with finesse. The task selection for alignment with the mathematical purpose, the attention to student thinking, and questioning are well done. There are a few flaws, but the teacher demonstrates smooth orchestration of the CMI framework elements. Students demonstrate ownership for learning and remain engaged in all phases of the lesson. Students are *highly likely* to deepen their understanding of mathematics.

____Level 7: Exemplary implementation of the CMI Framework

The teaching cycle structure (launch, explore, discuss) is masterfully used to promote understanding of a clear mathematical purpose. The lesson is nearly flawless. There is abundant evidence of teacher attention to pre-planning questions and anticipating student thinking. The teacher is able to quickly understand and flexibly meet the student needs "on the fly". The teacher allows sufficient struggle for students, providing support through questioning without providing the answers to the questions. The lesson is characterized by students' engagement in important mathematics, with students having ownership for learning, and making connections. Students are *almost certain* to deepen their understanding.

Appendix E: Confirmatory Factor Analysis Models

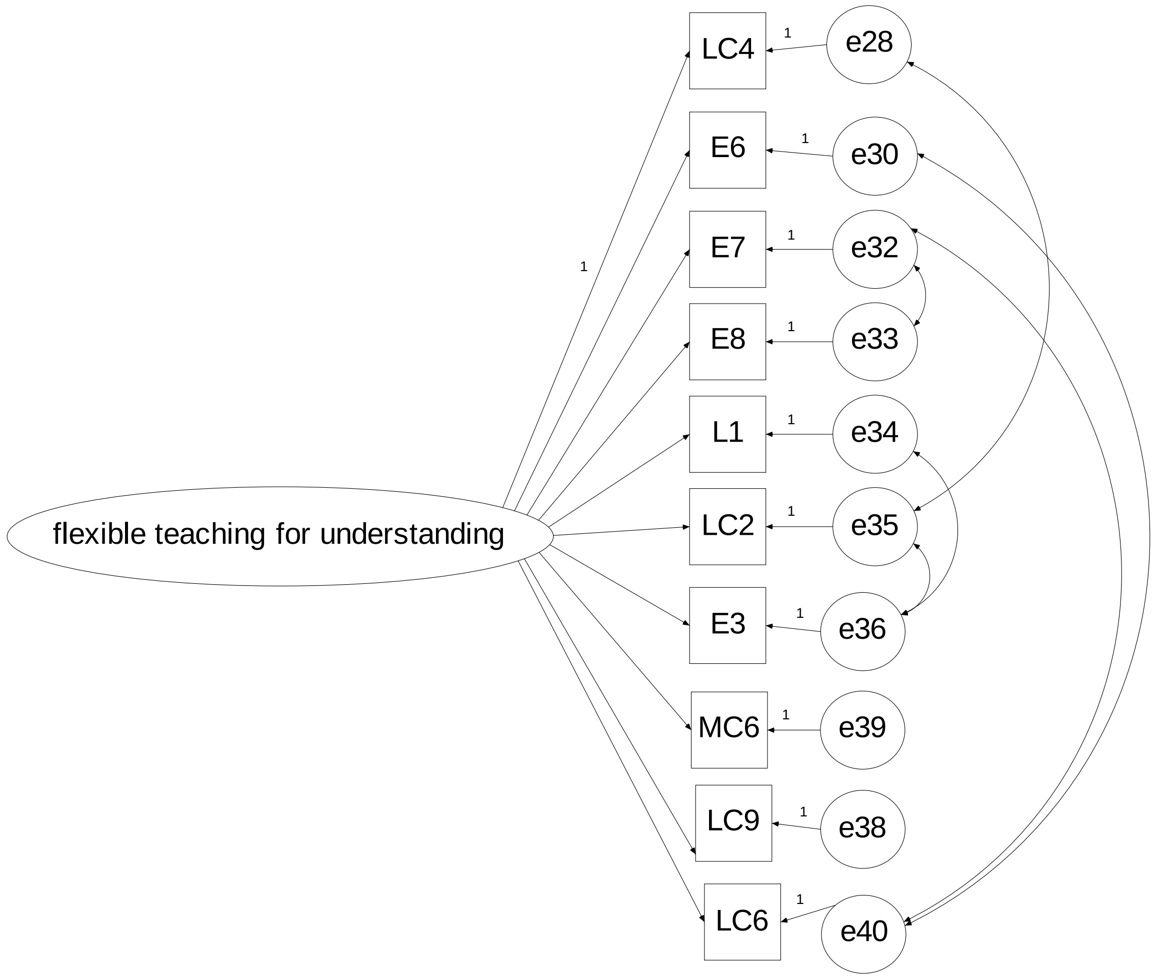

Figure 7. Single construct CFA model for flexible teaching for understanding.

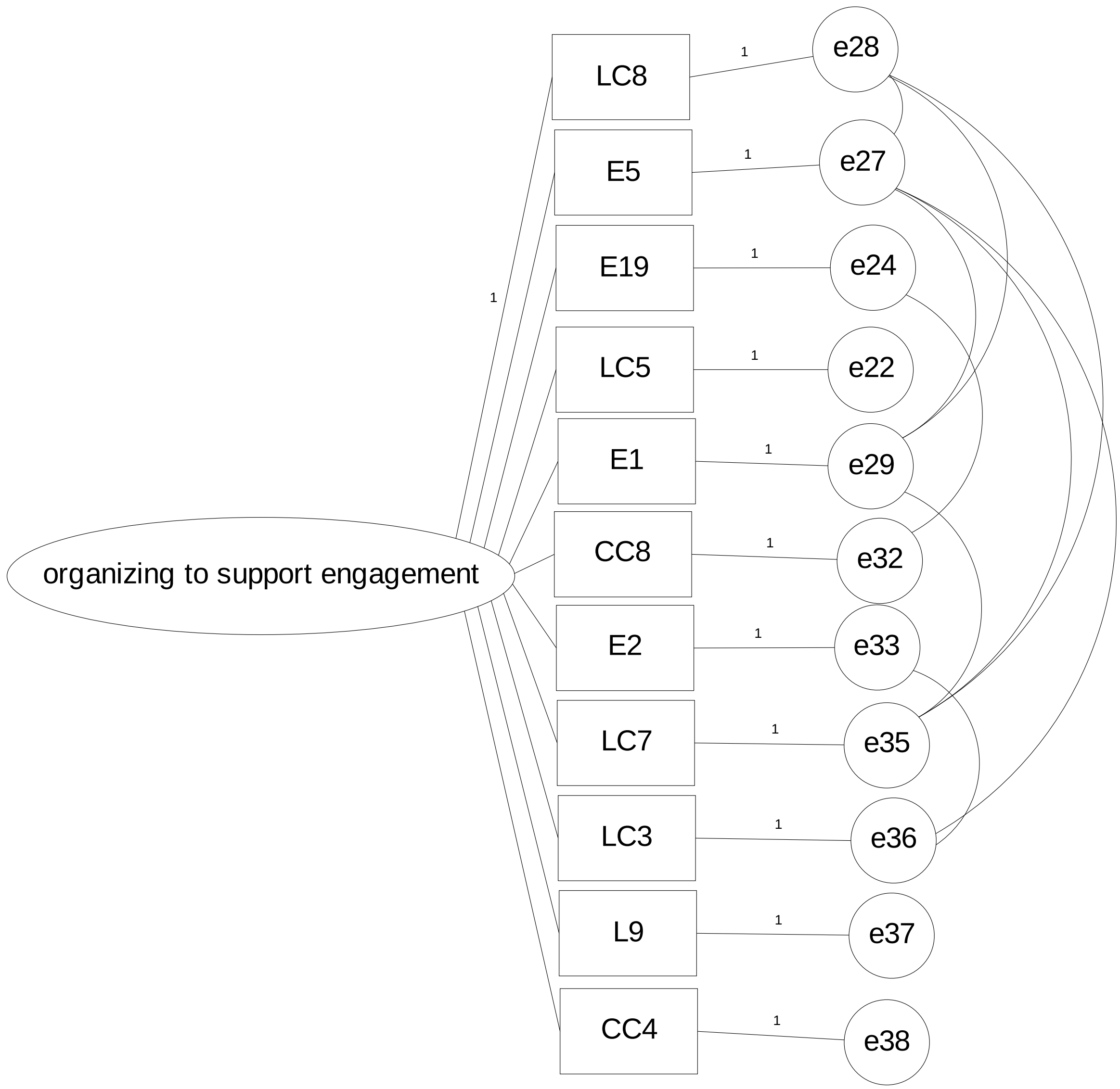

Figure 8. Single construct CFA model for organizing to support engagement.

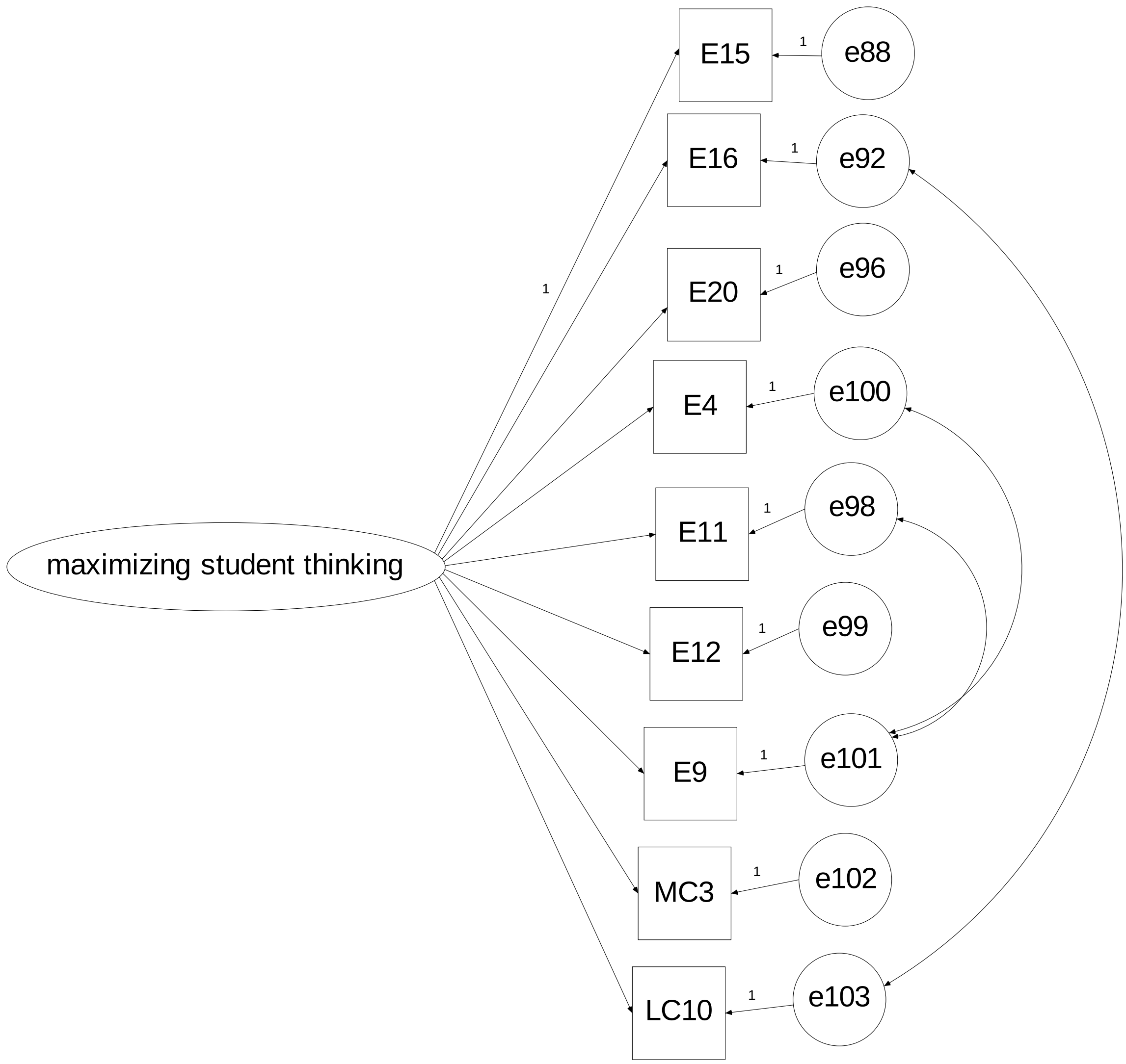

Figure 9. Single construct CFA model for maximizing student thinking.

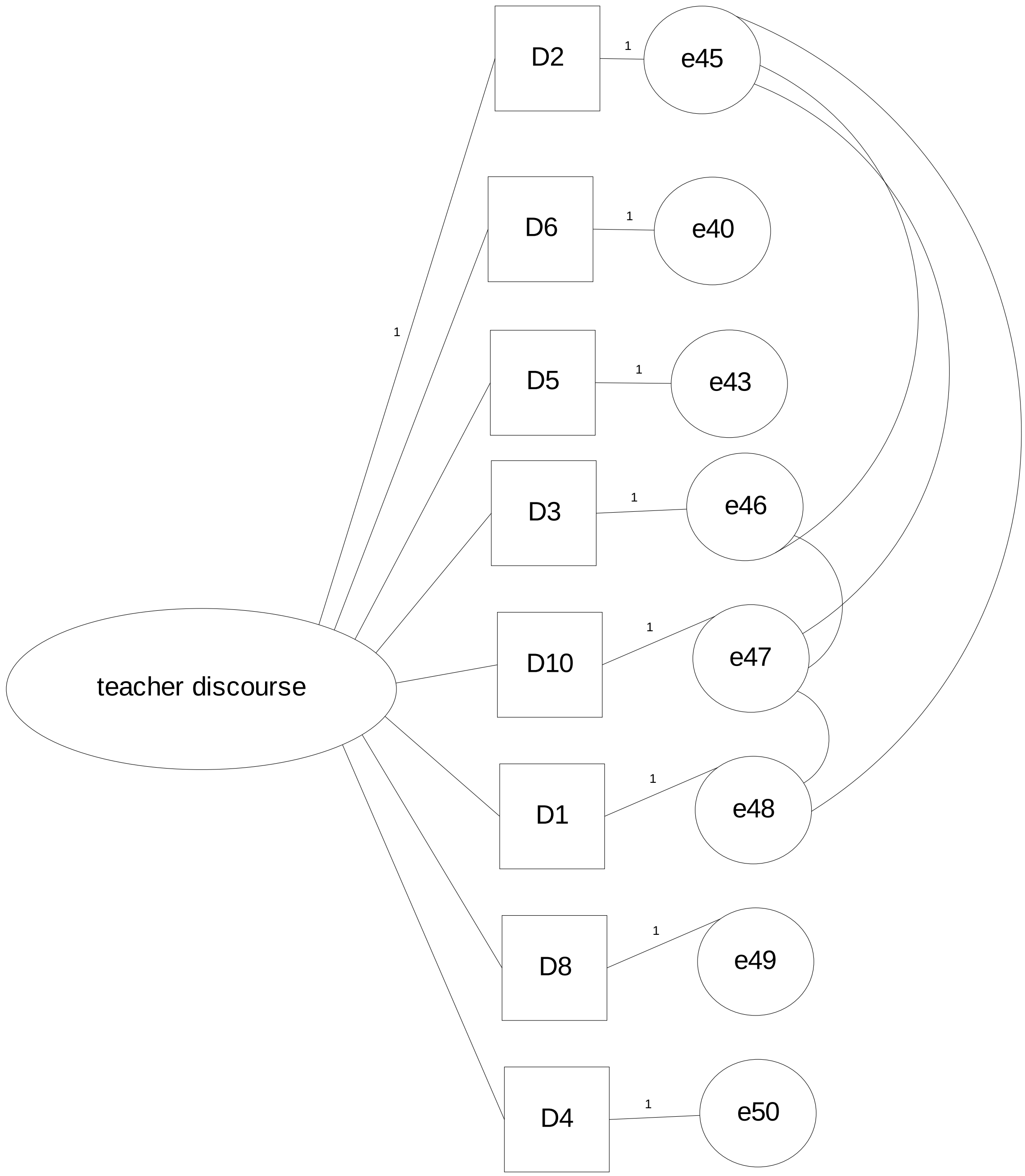

Figure 10. Single construct CFA model for teacher discourse.

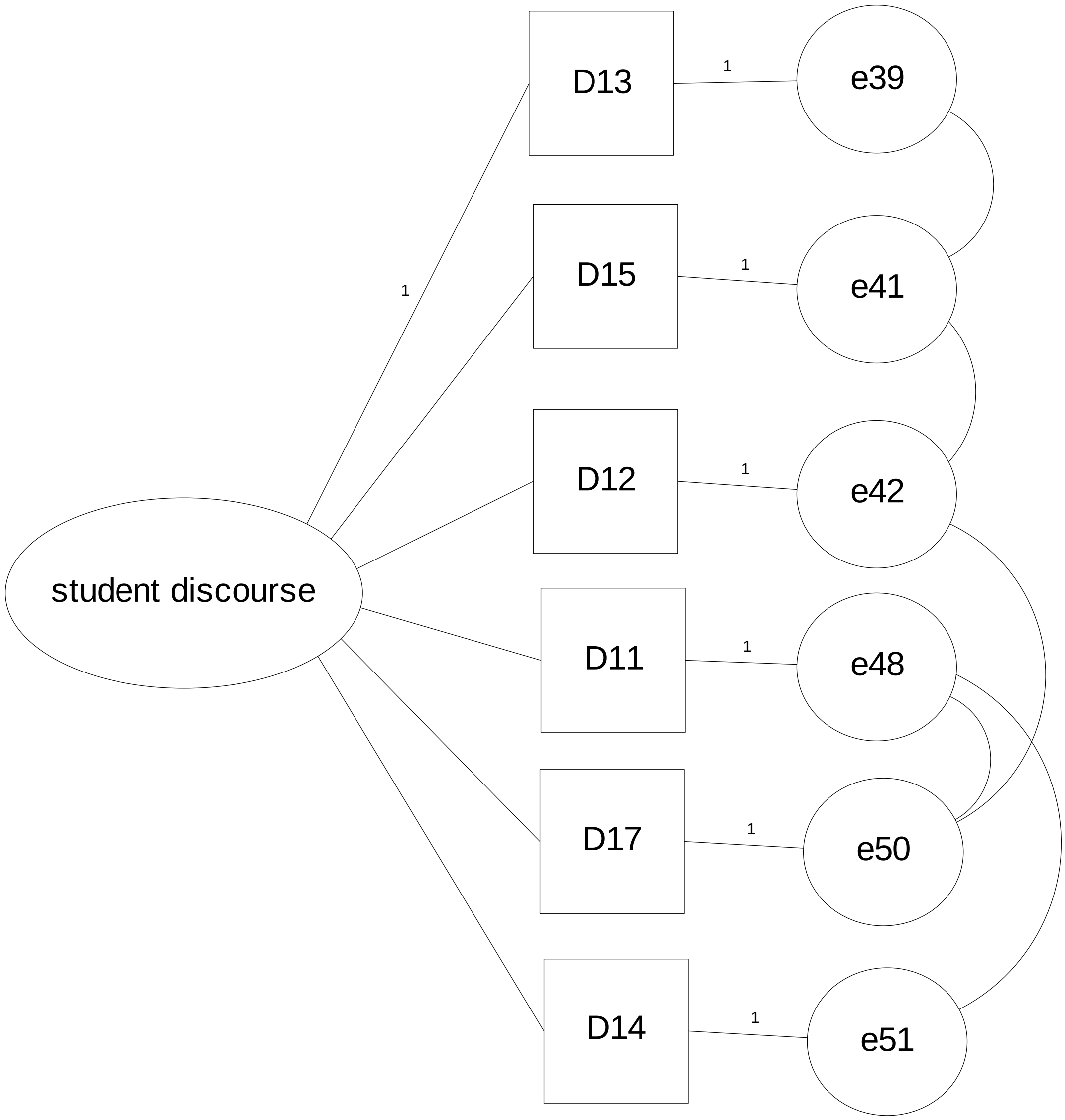

Figure 11. Single construct CFA model for student discourse.

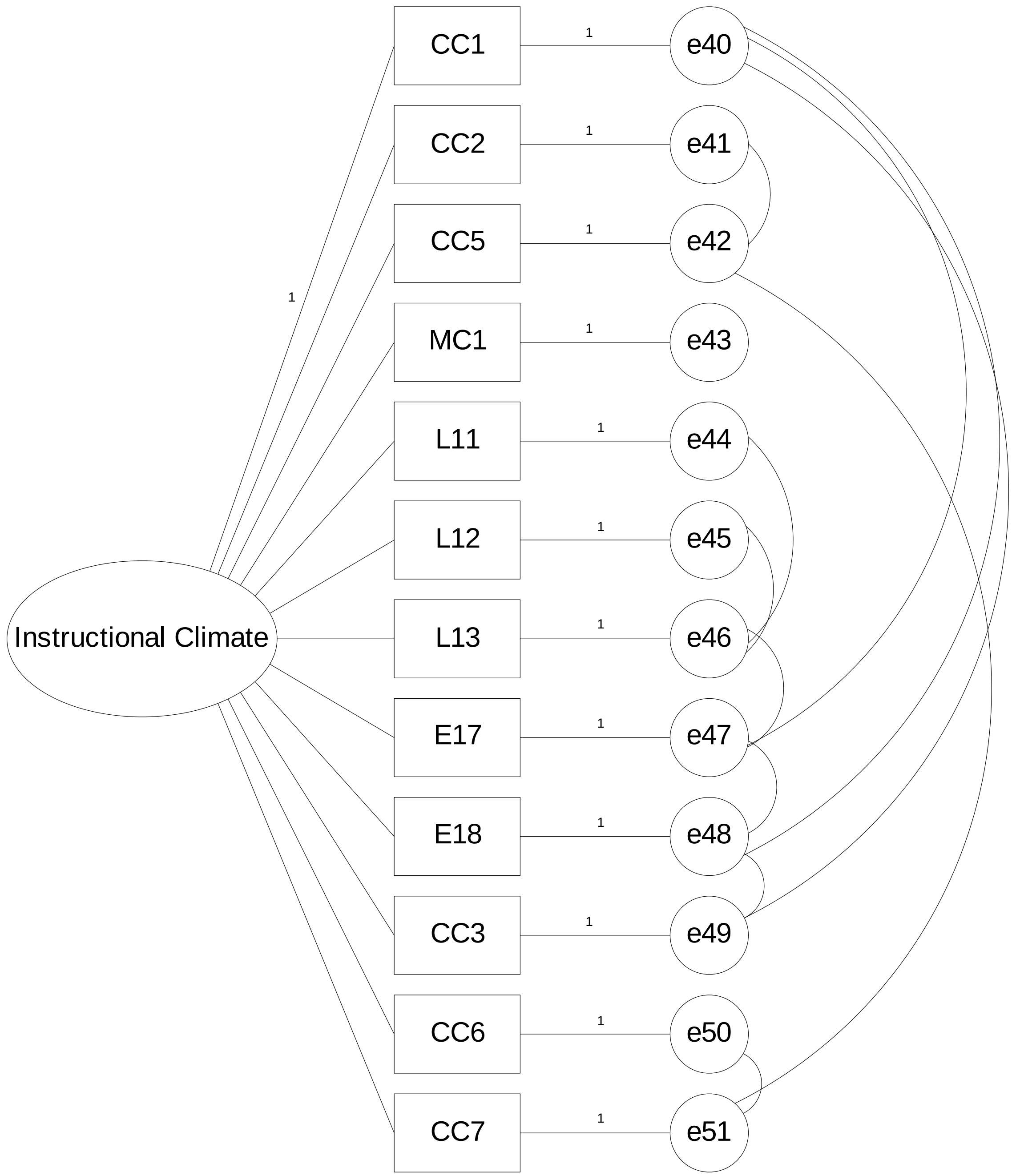

Figure 12. Single construct CFA model for instructional climate.

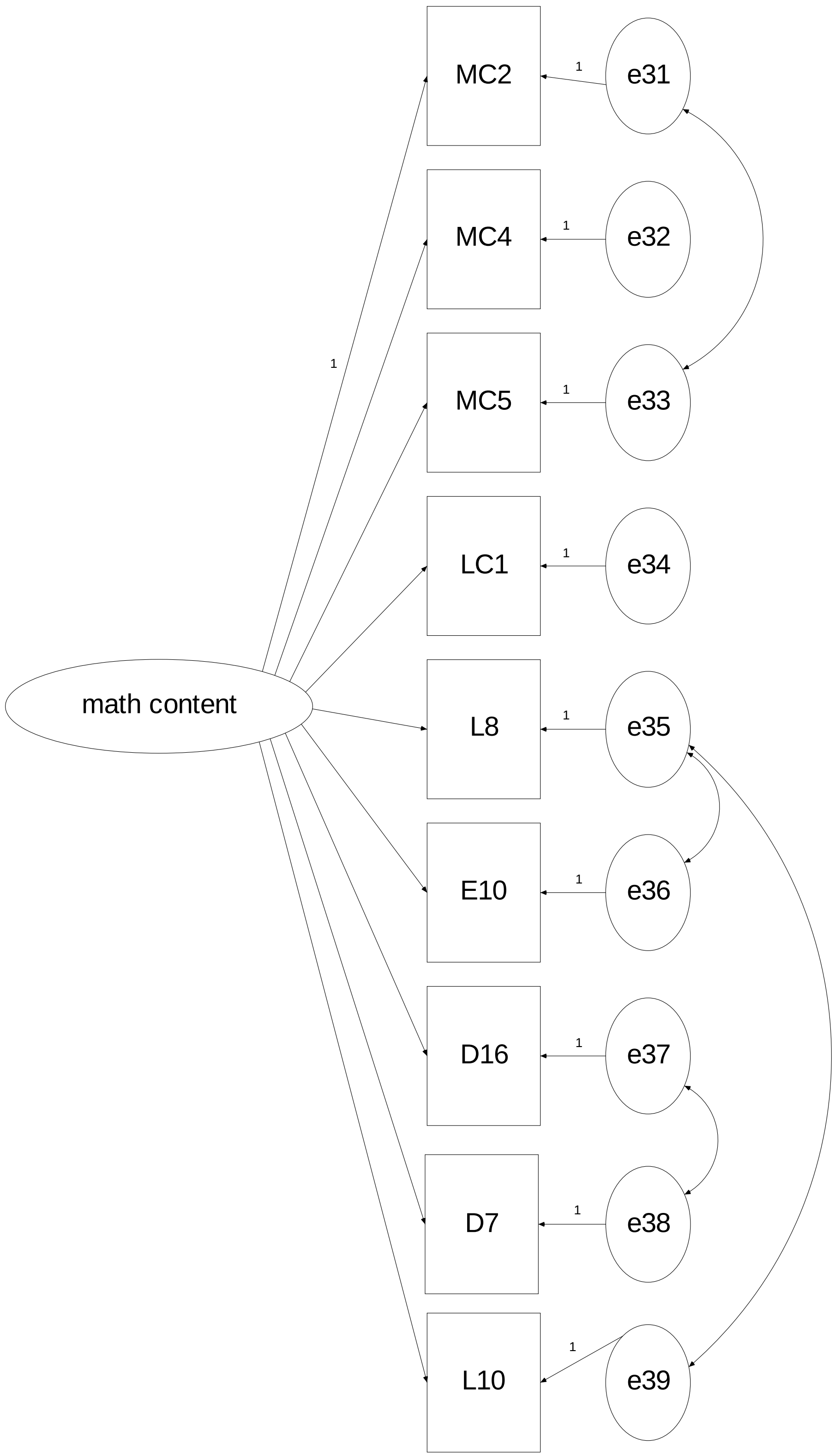

Figure 13. Single construct CFA model for math content.

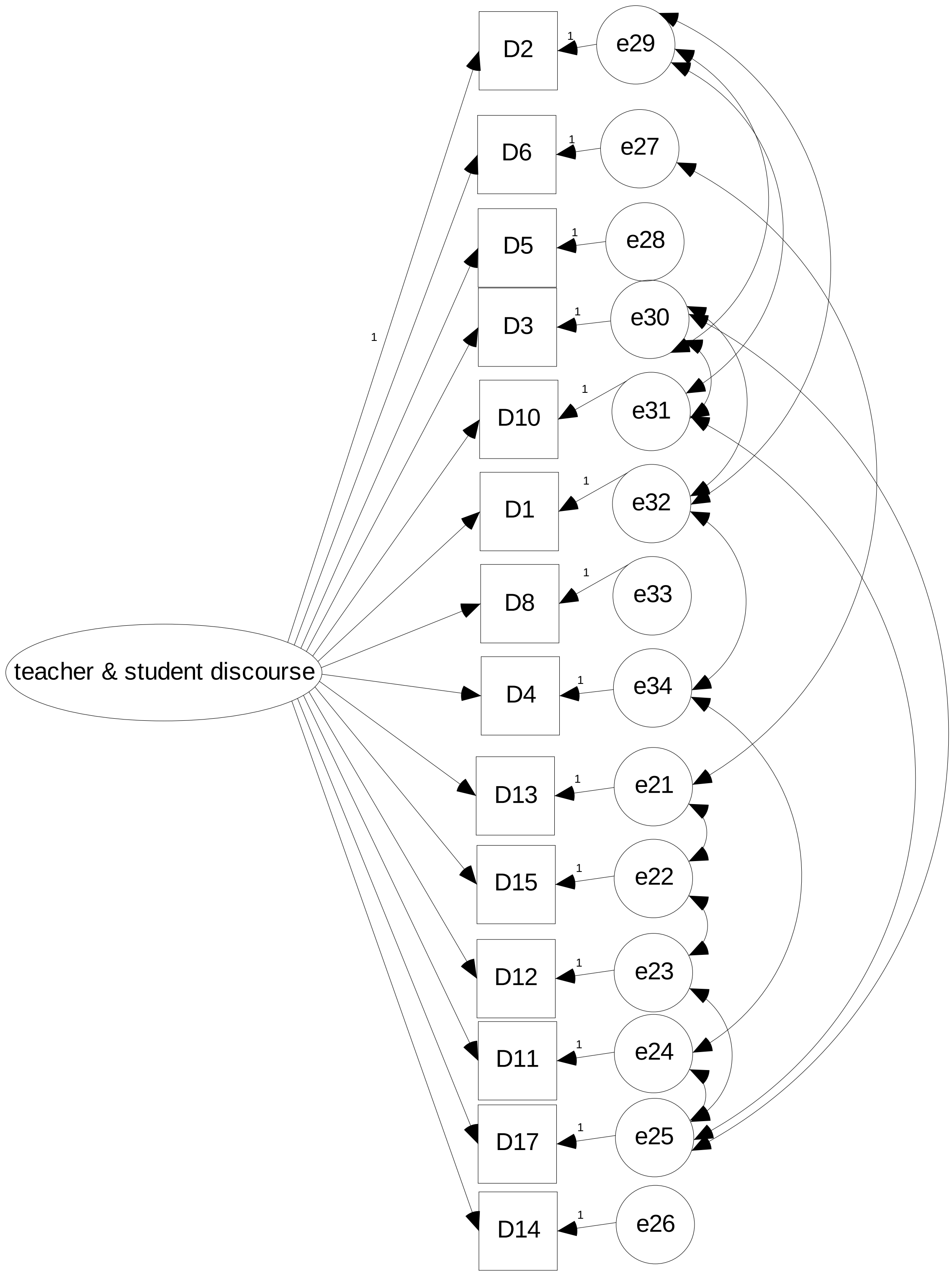

Figure 14. The single teacher and student discourse construct, collapsed from the student and teacher discourse constructs.

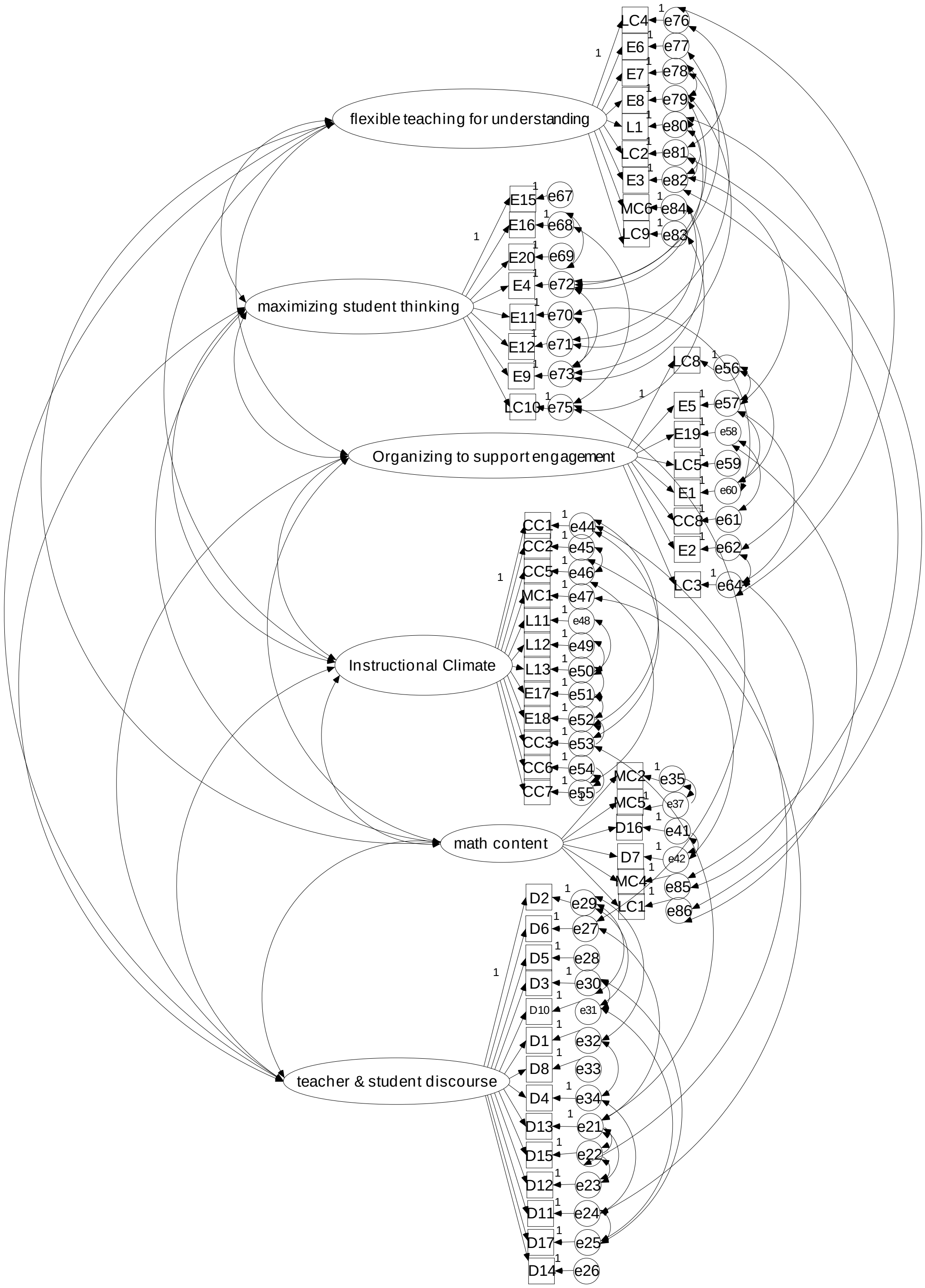

Figure 15. Modified 57-item CFA model.