

**INVESTIGATING THE RELATIONSHIP BETWEEN “EFFECTIVE”
TEACHERS AND THEORETICAL NOTIONS OF EFFECTIVE TEACHING: AN
ANALYSIS OF WHOLE-GROUP DISCUSSIONS**

by

Jimmy Scherrer

B.S., Education, Indiana University of Pennsylvania, 2000

M.A., Educational Administration, California State University Northridge, 2003

Submitted to the Graduate Faculty of
the School of Education in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Learning Sciences and Policy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Jimmy Scherrer

It was defended on

July 29, 2013

and approved by

Dr. Kevin Ashley, Professor, School of Law

Dr. Richard Correnti, Associate Professor, Learning Sciences and Policy

Dr. James Greeno, Professor, Instruction and Learning

Dr. Margaret Smith, Professor, Instruction and Learning

Dissertation Advisor: Dr. Mary Kay Stein, Professor, Learning Sciences and Policy

Copyright © by Jimmy Scherrer

2013

**INVESTIGATING THE RELATIONSHIP BETWEEN “EFFECTIVE”
TEACHERS AND THEORETICAL NOTIONS OF EFFECTIVE TEACHING: AN
ANALYSIS OF WHOLE-GROUP DISCUSSIONS**

Jimmy Scherrer, PhD

University of Pittsburgh, 2013

The literature base on teacher effectiveness has rapidly expanded in the past decade. Once dominated by social scientists, the literature is now heavily influenced by economists. Utilizing value-added modeling, economists have mainstreamed attempts to isolate the effect that a teacher has on student achievement. Findings from these attempts, however, confuse an improvement in measuring teachers with an improvement in measuring teaching. The field of education is still missing transparent, debatable measures of teaching practices. This study proposes a new measure, the adjusted whole-group discussion score, for one teaching practice—conducting whole-group discussions. It then uses that measure on a purposefully sampled group of teachers, and investigates its relationship to statistically derived measures of teacher effectiveness—that is, value-added estimates. This study is one of the first to go inside the classrooms of teachers being labeled effective through value-added modeling and to shed light on their classroom practices. In so doing, the study highlights different aspects of good teaching, which include being accountable to facts and procedures and being accountable to authentic discourse. Moreover, it articulates specific classroom discussion moves that can be used in targeted interventions. Findings illustrate that teachers with similar value-added scores can have markedly different teaching practices and that high value-added estimates do not necessarily reflect a full range of classroom teaching practices. The main policy implication of these findings

is that, similar to students who need *individualized instruction*, teachers need *individualized intervention*.

TABLE OF CONTENTS

PREFACE.....	XI
1.0 INTRODUCTION.....	1
1.1 QUANTIFYING EFFECTIVENESS: THE RISE OF VALUE-ADDED MODELING	4
1.1.1 What is value-added modeling.....	5
1.1.2 Limitations of value-added modeling.....	5
1.2 BEYOND VALUE-ADDED ESTIMATES: THEORIZING AND MEASURING QUALITY TEACHING.....	7
1.3 PURPOSE OF THE STUDY	8
2.0 CONCEPTUALIZING AND MEASURING STUDENTS’ OPPORTUNITIES IN THE MATHEMATICS CLASSROOM	10
2.1 LEARNING AND INSTRUCTION.....	10
2.2 LEARNING TO PARTICIPATE IN MATHEMATICAL PRACTICES	14
2.3 THE TEACHER’S ROLE IN LEARNING	16
2.4 CLASSROOM DISCOURSE	20
2.5 CONDUCTING WHOLE-GROUP DISCUSSIONS	23
2.5.1 Various dimensions of effective whole-group discussions	26

2.6 MEASURING STUDENTS' OPPORTUNITIES IN THE MATHEMATICS CLASSROOM	27
2.6.1 The situative perspective of learning: selecting a grain size	28
2.6.2 The Analyzing Teaching Moves Guide: theorizing effective discourse	30
2.6.3 Conceptualization of the whole-group discussion score	35
2.7 RESEARCH QUESTIONS.....	37
2.7.1 Hypotheses	37
3.0 METHODOLOGY	40
3.1 SAMPLE	40
3.2 CODING.....	45
3.3 ANALYSIS.....	45
3.3.1 Data sources for research question 1.....	46
3.3.2 Data sources for research question 2.....	58
4.0 RESULTS	60
4.1 THE RELATIONSHIP BETWEEN WHOLE-GROUP DISCUSSION SCORES AND VALUE-ADDED ESTIMATES.....	60
4.1.1 Comparing <i>amount</i> of whole-group discussion with value-added estimates	62
4.1.2 Comparing teachers' WGD and WGD_Adj scores	63
4.1.3 Comparing teachers' VAM estimates with their adjusted whole-group discussion scores	65
4.1.4 Summary: Research question 1	66

4.2	USING DESCRIPTIONS OF CLASSROOM PRACTICE TO HELP EXPLAIN WHY SOME OF THE PARTICIPANTS' MEASURES DO NOT CONVERGE	67
4.2.1	Teachers accountable to facts and procedures and authentic discourse ..	69
4.2.2	Teachers accountable to facts and procedures but not authentic discourse	83
4.2.3	Teachers accountable to authentic discourse but not facts and procedures	87
4.2.4	Teachers not accountable to facts and procedures or authentic discourse	90
4.2.5	Summary: Research question 2	93
5.0	DISCUSSION	96
5.1	POLICY IMPLICATIONS	98
5.2	LIMITATIONS.....	100
5.2.1	Limitations of the adjusted whole-group discussion score.....	100
5.2.2	Limitations of the study	102
5.3	NEXT STEPS.....	103
5.4	CONCLUSION	105
	APPENDIX A	106
	APPENDIX B	107
	APPENDIX C	110
	APPENDIX D	114
	BIBLIOGRAPHY	124

LIST OF TABLES

Table 3.1 MQI and Value-Added Percentile Rank for Participants	42
Table 3.2: Scoring Guidelines: Whole-Group Discussion Score	49
Table 4.1: Whole-Group Discussion, MQI, and VAM scores for Teachers	61

LIST OF FIGURES

Figure 3.1: Mathematical Quality of Instruction (MQI) versus value-added estimate percentile rank.	44
Figure 4.1: Average percent of each teacher’s lesson that was whole-group discussion (% WGD) versus value-added percentile rank (VAM).....	63
Figure 4.2: Whole-Group Discussion score (WGD) versus Adjusted Whole-Group Discussion score (WGD_Adj).....	64
Figure 4.3: Adjusted whole-group discussion score (WGD_Adj) versus value-added estimate percentile rank (VAM).....	66
Figure 4.4: Pictures of Vince's discussions.....	73
Figure 4.5: Pictures of Dolly's Discussions	79
Figure 4.5: Pictures of Dolly's Discussions (continued).....	80
Figure 4.6: Pictures of Hanna's discussions.....	86
Figure 4.7: Pictures of Alberto's discussions	92

PREFACE

My dissertation committee—Kevin Ashley, Richard Correnti, James Greeno, Margaret Smith, and Mary Kay Stein—sacrificed a lot of time and energy helping me complete this document. I apologize if I took more than my fair share; each minute was appreciated. Mary Kay’s guidance as committee chair was remarkable. These would be blank pages without her.

Deanna Weber Prine provided clarity and support on a daily basis. Again, I apologize for taking more than my fair share.

Other scholars outside of my own institution also took the time to sit with me and provide encouragement throughout this process. Whether they opened the door to their office, or opened the door to their home, their thoughts were invaluable: Claude Goldenberg, Stanford; Heather Hill, Harvard; Jim Stigler, UCLA; and Etienne Wenger.

Individuals situated outside of my academic life also played a large role helping me achieve this milestone: Bill Saunders, co-founder of the Talking Teaching Network, has served as a model of excellence for how to bring research to practice. Bill Bushaw, executive director of Phi Delta Kappa, whose optimism seems to bring hope to any education issue. Aurora Arreola, Caroline Piangerelli, and Barbara Thibodeau all served as wonderful mentors and leaders while I was working in the Los Angeles Unified School District.

Most importantly, the beautiful children of Pacoima, California deserve to be acknowledged. Their strength and courage in the face of adversity is extraordinary. They are my inspiration.

1.0 INTRODUCTION

The literature base on teacher effectiveness has rapidly expanded in the past decade. Once dominated by social scientists, the literature is now heavily influenced by economists. Utilizing sophisticated statistical models, economists have mainstreamed attempts to isolate the effect that a teacher has on student achievement. Findings from these attempts suggest that there exists wide variation in teacher effectiveness, usually measured by a teacher's contribution to standardized student assessment outcomes. Moreover, these studies find that "observable" characteristics of teachers such as experience and education explain very little of the variation (Hanushek & Rivkin, 2006; Rivkin, Hanushek, & Kain, 2005). Thus, it is commonly assumed that much of the unexplained variance can, in part, be attributed to teachers' classroom practices.

This assumption has led to many well-publicized claims about the effect of teachers. However, because the national debate about teacher effectiveness now seems to be dominated by those outside the arena of practice, effectiveness tends to be conceived not as what it actually is but as how it can be measured. For example, in a highly cited article,¹ Rivkin, Hanushek, and Kain (2005, p.419) assert that, "high quality instruction throughout primary school could substantially offset disadvantages associated with low socioeconomic background." While this may be true, the authors make this assertion without actually studying, measuring, or even

¹ At the time of this writing, in just over eight years since its publication, the Rivkin et al. (2005) article has been cited 2,271 times. That is an average of about 24 cites per month.

defining high quality instruction (i.e., teaching). This is not a unique instance. Most, if not all, of the economics-of-education literature dealing with teacher effectiveness measures *teachers*, not *teaching*. This study adopts Hiebert and Morris's (2012, p.92) definition of teaching—"the methods used to interact with students about content."

The emerging literature on teacher quality continues to confuse an improvement in measuring teachers with an improvement in measuring teaching. A quote from the well-known education economist Eric Hanushek is illustrative, "I use a simple definition of teacher quality: good teachers are the ones who get large gains in student achievement for their classes; bad teachers are just the opposite" (Hanushek, 2002, p.3).²

While investigating gain scores can be useful, the practice is very limited. Gain scores alone create a black box—we have no idea what teaching practices are related to high gain scores, and we have no idea what teaching practices are related to low gain scores.

Hanushek (2009) has also proposed a policy of *teacher deselection*, where the worst 6-10 percent of teachers are simply eliminated. He argues, in a statistically sound way, that this could increase student achievement (on standardized assessments) by one-half standard deviation.

The field of education would be remiss if it believes teaching will improve by removing "ineffective" teachers without articulating what it is about their teaching that is ineffective.³

²Many are willing to adopt this simple definition of teacher quality asserting that it is empirically-based and it avoids defining quality through any "particular type" of teaching (see Cochran-Smith & Zeichner, 2005). However, this assertion is problematic—valuing particular measures implicitly values approaches that increase those measures. By defining quality through increased student achievement on traditional standardized assessments, one is implicitly valuing types of teaching that best increase these assessments.

³ Healthy systems are able to convert the failure of individuals into gains for the entire system. For example, in engineering, every bridge that collapses makes future bridges safer. In aviation, every plane that crashes makes future flights safer. After rigorous study, healthy systems are able to uncover and articulate "what went wrong."

Equally unhelpful is the practice of labeling teachers “effective” without articulating what it is about their teaching that is effective.

The field of education faces a problem: current measures of teacher effectiveness are incomplete. They tell us which teachers are accountable to facts and procedures—that is, which teachers are able to produce high scores on standardized assessments. But they do not tell us which teachers are accountable to other mathematical practices—that is, which teachers have classroom instruction that goes beyond the teaching of basic facts and procedures. A solution to this problem is for the field to develop measures for specific teaching practices that theory suggests are beneficial. This will result in a few major benefits. First, the field will have debatable, transparent measures of teaching practices that education researchers can use when discussing “effectiveness.” Second, these measures will articulate different teaching practices that can be used in targeted interventions with practitioners. Third, these measures can be used to test the validity of the current practice of using measures of *teacher* effectiveness (e.g., value-added modeling) as a proxy for *teaching* effectiveness.

The present study proposes a measure of one teaching practice, conducting whole-group discussions, and investigates its relationship to current notions of effective teachers. In so doing, the study highlights different aspects of effective teaching. Moreover, it proposes a debatable and transparent measure of one teaching practice, including the articulation of specific classroom discussion moves that practitioners can use in targeted interventions.

I begin by reviewing the current policy environment that contextualizes why this study is important to carry out—that is, why an improvement in measuring teachers is not enough, and perhaps dangerous. Then I review the literature on conceptualizing and measuring students’

opportunities to participate in high-quality, robust mathematics instruction. Together these sections motivate the specific research questions that this study aims to address.

1.1 QUANTIFYING EFFECTIVENESS: THE RISE OF VALUE-ADDED MODELING

The United States has historically used *status* measurement models when judging the performance of schools and, more recently, teachers. Status models, often referred to as cross-sectional, measure student performance at one point in time. These “snapshots” are generally compared to each other, or to some set target. For example, the controversial *No Child Left Behind* accountability system in the United States set a yearly target for the percentage of students at each school (and district) that had to be deemed “proficient” or “advanced.” Schools measured their student body, usually through end-of-the-year standardized testing, and compared the percent of their students at “proficient” or “advanced” to the nationally set target. Those who did not meet the target faced possible sanctions.

However, because variation in student achievement scores is largely explained by background characteristics of students (that are outside of a teacher’s control), status models reward schools (and teachers) for *who* they teach rather than *how* they teach (Harris, 2009; Rothstein, 2004).⁴

⁴ This is of great concern for the education of disadvantaged populations. Accountability systems using status models incentivizes educators to work with the most advantaged students. The general belief, birthed from reports using status measures, has traditionally been that inner city schools provide a much inferior education when compared to suburbia. However, VAM has suggested the average effectiveness of teachers in high poverty schools is only the slightest bit lower than the average effectiveness of teachers in much more advantaged schools (Sass, Hannaway, Xu, Figlio, & Feng, 2010).

Thus, statistical models designed to isolate teacher effects have garnered much interest.⁵ Most prominent among these methods is value-added modeling (VAM). The emergence of VAM in education, however, has been a rather contentious issue.

1.1.1 What is value-added modeling

In theory, VAM captures a fraction of achievement growth over time that can be attributed to a particular teacher (or school, program, counselor, and so on). VAM improves on the previously mentioned status models by controlling for non-school factors related to student achievement (e.g., socioeconomic status, health, parent education). These “controls” allow for a computation of *expected growth* for each student. Any deviation from this expected growth is attributed to the teacher. Averaging these deviations across many years for many students, a teacher is given a value-added estimate (often referred to as a value-added score).⁶ However, like all statistical methods, VAM brings with it a host of limitations.

1.1.2 Limitations of value-added modeling

Dubious Assumptions. All statistical models are designed around basic assumptions. If these assumptions are violated, the estimates generated by models may not accurately “describe” the

⁵ It is important to highlight this point: much of the push for the use of VAM is grounded in the history of status models’ inability to identify effective teachers (and schools). Surprisingly, this is rarely mentioned in current “value-added debates.” Pundits who highlight the limitations of value-added models fail to mention that the previously used status models were even worse at identifying effective teachers (and schools).

⁶ Value-added models produce a quantitative outcome that includes considerable measurement error. As such, I prefer using “estimate” over “score.” I feel the word “score” infers an exact quantitative outcome rather than the actual approximation that is yielded.

phenomena under study. In the case of value-added, it has been argued that that the underlying assumptions of the models are violated in real educational settings (Scherrer, 2011). For example, one underlying assumption in most commonly used value-added models is that *school administration and peer interactions among teachers do not impact the manner in which one teaches*. Empirical evidence suggests otherwise: principals, coaches, and colleagues, among others, have all been shown to have an effect on classroom teaching practices (e.g., Bryk & Schneider, 2003; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Penuel, Sun, Frank, & Gallagher, 2012). Often billed as social capital, these studies suggest individuals' social interactions within their surrounding community can augment, enable, constrain, and/or mediate their classroom practice (Resnick & Scherrer, 2012). Thus, as a result of the mediating effect of their surrounding communities, two teachers who bring the same skill set to teaching could be identified as not equal by a value-added estimate. This violation is just one example of why many view value-added estimates with skepticism.

Reliability. The weak stability of the estimates produced by VAM is another area of concern. Value-added estimates have been shown to be unreliable across models, between assessments, and over time. For example, Hill, Kapitula, & Umland (2011) found slightly altering their models produced a radically different “ranking” of teachers. When replacing a school fixed effect with a vector of student covariates, for example, the percentile rank for 23 of their 24 focus teachers changed: 14 increased, 9 decreased. One of their focus teacher's relative rank went from the 26th percentile to the 65th percentile (an *increase* of 39) while another went from the 83rd percentile to the 44th percentile (a *decrease* of 39). Value-added estimates have also been found to be unstable between test instruments (McCaffrey, Sass, Lockwood, & Mihaly,

2009), assessment forms (Lockwood, McCaffrey, Hamilton, Stecher, & Martinez, 2007), and over time (Koedel & Betts, 2005; Schochet & Chiang, 2010).

Validity. Most relevant to the present study is the validity of using value-added estimates to identify “effective” teachers. Even the most sophisticated statistical models can only produce estimates as good as the data used to create them. The most common value-added models use data from assessments that measure basic, procedural skills and not the more cognitively demanding thinking and reasoning valued by most educators (Rothman, Slattery, Vranek, & Resnick, 2002; Rothstein, Jacobsen, & Wilder, 2008; Shepard, 2000). Hence, even those scoring the highest on traditional measures of learning may not possess the skills we desire.⁷ This point is central to the present study’s argument, and will be discussed throughout.

1.2 BEYOND VALUE-ADDED ESTIMATES: THEORIZING AND MEASURING QUALITY TEACHING

With all of their limitations, why have value-added estimates proliferated? For one, although VAM may indeed be imperfect, there is no evidence that suggests the technique is any less reliable, valid, or imprecise than traditional measures of teachers (e.g., principal evaluations, status measures). Recent research suggests that value-added estimates have convergent validity with other measures of teacher “effectiveness.”⁸ A study by Jacob and Lefgren (2008) found that value-added estimates correlate highly with principals’ evaluations of teachers. Both value-

⁷ It should be noted that these “validity” issues exist regardless of the statistical model being used. In other words, the questionable validity is an artifact of the tests used to generate the data, not the statistical models used to analyze the data.

⁸ Achieving convergent validity suggests that the estimates are related to what they should theoretically be related to.

added estimates and subjective principal evaluations are generally effective at identifying the very best and worst teachers. Another study (Hill et al., 2011) found teachers' value-added estimates correlated with their mathematical knowledge and the mathematical accuracy of their classroom instruction.

Second, enthusiasm for value-added measures may be the result of a lack of agreed-upon measures that capture the practices of quality teaching. Without a common view of quality teaching, measures used to identify “quality teachers” will likely continue to dominate the policy discussion.

There are many dimensions to teacher quality (e.g., teacher knowledge, teaching practices, teacher beliefs), and many ways to measure each (Gitomer & Bell, 2013). Every measure brings with it a particular perspective that positions certain elements to the foreground and assigns other elements to the background.

While the current study acknowledges the importance of each dimension of teacher quality, it only focuses on one: teaching practices. Specifically, it focuses on the teaching practice of conducting whole-group discussions, and measures that practice using a coding protocol that brings certain patterns of teacher-student interactions to the foreground.

1.3 PURPOSE OF THE STUDY

As the use of VAM in education continues to expand, one outstanding question remains: What is the relationship between “effective” teachers—as determined by value-added estimates—and theoretical notions of effective teaching? The purpose of this study is (1) to go inside the black box of value-added estimates and describe the whole-group discussion practices of teachers

being labeled effective to determine if value-added estimates can indeed be used as a proxy for high quality instruction, (2) to develop and use a newly created coding and scoring scheme as the basis of a transparent, debatable measure of one specific teaching practice—conducting whole-group discussions, and (3) to broaden the notion of effective teachers beyond those who simply are accountable to facts and procedures (i.e., those with high VAM estimates), to include teachers that are accountable to a wide range of mathematical practices. Specifically, this study investigates the following research questions:

1. *To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers?*
2. *What is it about the participants' classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not converge with their value-added estimates?*

This study will contribute to the work of educational researchers, policy makers and practitioners. For educational researchers, the work will articulate and describe a debatable and transparent measure of one teaching practice: conducting whole-group discussions. For policy makers and practitioners, this study is one of the first to go inside the classrooms of teachers being labeled “effective” by value-added estimates (and their peers being labeled less-effective) and to shed light on their teaching. Specifically, this study investigates how these teachers conduct whole-group discussions and compares those practices to theoretical notions of effective teaching. In so doing, the study highlights various dimensions of quality teaching.

2.0 CONCEPTUALIZING AND MEASURING STUDENTS' OPPORTUNITIES IN THE MATHEMATICS CLASSROOM

I begin Chapter 2 by theoretically building the conception of how teaching can *create* opportunities for students to participate in various types of mathematical practices. I end Chapter 2 by theoretically building the conception of how to *measure* teaching practices that provide opportunities for students to participate in whole-group discussion of a mathematical task.

2.1 LEARNING AND INSTRUCTION

The notion that learning should be a primary outcome of schooling is widespread. Equally as widespread is the notion that learning occurs, at least in part, as a result of “instruction” (i.e., *teaching*). However, teaching practices vary widely, and the use of particular instructional strategies reflects certain pedagogical epistemologies (Alexander, 2004; Askew et al., 1997; Barnes & Todd, 1995; Moyles et al., 2003; Wells, 1999). What, it is fair to ask, is the best pedagogical approach? The answer to this question depends on *what* one wants students to learn. This section discusses three perspectives on learning: the behaviorist, the cognitive, and the situative. This study uses the latter—the situative perspective—to conceptualize how teachers

might provide opportunities for their students to participate in various mathematical practices during whole-group discussions of a mathematical task.⁹

Traditional instruction in the west has been informed by the behaviorist perspective of learning. This perspective focuses attention on the organization of associations and components of skills; learning is viewed as the process of accumulating and strengthening these associations and skills (Thorndike, 1906). Viewed from this perspective, teaching consists of the *direct instruction* of skills and follows a carefully sequenced curriculum of increasingly advanced skills. The hallmark of direct instruction is the active role assumed by the teacher and the passive role assumed by the student (Baumann, 1988). Evidence suggests that while this method of instruction might be an effective means of teaching factual content, it is less effective for building higher-order cognitive skills such as reasoning and problem solving, nor does it result in the flexibility needed to apply content in novel situations (Peterson & Walberg, 1979; these points are elaborated in the following sections). Thus, if one views learning as more than the accumulation of facts and strengthening of skills, the behavioristic perspective is limited.

The cognitive perspective on learning, on the other hand, focuses attention on individual cognitive structures; learning is viewed as a reorganization of concepts in the learner's mind that leads to increased "meaning making" (Bruner, 1990). Teaching that adopts this perspective is primarily associated with constructivist learning theories (Confrey, 1990). Influenced by Piaget's ideas about cognitive development, constructivist theories are grounded in the belief that a reorganization of cognitive structures occurs through intellectual activity rather than by absorption of information. Hence, what distinguishes constructivism from direct teaching is the

⁹ To use language from the situative perspective, this study conceptualizes how teachers might provide opportunities for their students to become legitimate peripheral participants in a community of mathematicians.

active role that the student plays in the learning process—e.g., more problem solving, comprehending (see Greeno, Collins, & Resnick, 1996, for more on the behaviorist and cognitive perspectives and their links to education).

While the cognitive perspective on learning shifts attention from the accumulation of facts to conceptual thinking, many learning scientists question its silence with respect to the impact that one's environment has on learning: If one believes that the locus of knowledge is outside the individual, then the cognitive perspective on learning is limited. As a result, more and more “postmodern constructivist” perspectives have emerged (Palincsar, 1998). The hallmark that unifies these perspectives “is the rejection of the view that the locus of knowledge is in the individual.” Rather, in these perspectives, “learning and understanding are regarded as inherently social,” and community activities and tools are regarded as “integral to conceptual development” (Palincsar, 1998, p.348). Thus, these are referred to as social constructivist perspectives.

This study adopts a social constructivist perspective known as the situative perspective of learning.¹⁰ The situative perspective of learning focuses attention on participation in communities in which people interact with each other and with material in their environment (Brown, Collins, Duguid, 1989; Greeno, 1997, 1998; Lave & Wenger, 1991; Resnick, Saljo, Pontecorvo, & Burge, 1991; Wenger, 1998). In this perspective, learning and cognition are inseparable from the activities in which students engage, and learning is not viewed as the

¹⁰ I adopt Greeno's (e.g., Greeno, 1997, 2011) use of the term situative. Greeno made the small syntactic change from the more traditional “situated” to “situative” to make less likely the misconception that only some action, cognition, or learning is situated. To preserve the original nature of some quotations, however, the terms “situative” and “situated” are used interchangeably throughout the paper.

acquisition of skills or structure, but rather as the increased ability to participate in the practices of a particular community.¹¹

From a situative perspective, learning is not absorbing, nor is it transmission or assimilation. Learning occurs when individuals change the way they think, engage, and participate in the practices of a specific community (Lave & Wenger, 1991; Rogoff, 1990; Wenger 1998); in other words, learning occurs as individuals grow into the intellectual life around them (Vygotsky, 1978). As such, the situative perspective on learning focuses primarily on interactive systems, which include individuals as participants. This interaction can be with others, tools, tasks, and so on (Lave & Wenger, 1991; Resnick, Saljo, Pontecorvo, & Burge, 1991). As individuals negotiate and renegotiate the meaning of their surrounding communities, they will naturally change the way they participate in it. This change in participation, from a situative perspective, is what signifies learning (Lave & Wenger, 1991).

It is important to highlight that the situative perspective on learning is agnostic as to the type of educational practices that should be adopted. However, it does assert that participation in practices is fundamental to what students learn—the skills that children learn are rooted in, and are inseparable from, their participation in the classroom community in which they interact. From a situative perspective, the paramount question in schooling is not *will* children learn, but rather *what* will children learn.

The situative perspective of learning suggests that if we want students to become fluent in various mathematical practices, they must be granted legitimate peripheral participation in communities where they have the opportunity to engage in these various practices (Lave &

¹¹ It is important to highlight here that if one believes learning is more than the acquisition of facts and procedures, then any measure of learning needs to contain more than just assessment of facts and procedures.

Wenger, 1991). For example, if there are certain ways that the community of mathematicians think and reason that we would like our students to mimic (e.g., justifying their work, examining constraints, connecting ideas, critiquing the reasoning of others), then students need the opportunity to participate in those ways of thinking and reasoning. The study discussed herein examines one particular practice valued in a community of mathematicians: discourse around tasks (Aleksandrov, Kolmogorov, & Lavrent'ev, 1999; van Kerkhove & van Bendegem, 2010). Specifically, this study examines the types of questions teachers ask during whole-group discussion of a mathematical task and how teachers use student responses to those questions. It goes on to discuss how different types of whole-group discussions provide students different opportunities to participate with the mathematics.

2.2 LEARNING TO PARTICIPATE IN MATHEMATICAL PRACTICES

Communities of practice are connected by socially constructed webs of beliefs and identities that largely determine what it is they do (Geertz, 1983; Wenger, 1998). Becoming a legitimate member of a community corresponds to an increase in participation of the community's practices (Lave & Wenger, 1991). Within these communities, it is not the acquisition of knowledge with a universal currency (e.g., textbook knowledge) that identifies the "competent" member. Rather, it is a demonstrated ability to participate in ways that are valued by other members of the community of practice that is all-important and constitutes learning and competence (Contu & Willmott, 2003; Wenger, 1998). The practices of a community, therefore, become an essential part of the learning curriculum.

Within mathematics, learning is indeed as much a matter of acquiring the practices of that community as of acquiring any defined set of knowledge (Resnick, 1989; Schoenfeld, 1992). Kitcher (1984) asserts that mathematical knowledge is to be understood as knowledge of mathematical practice, not simply knowledge of content. Halmos (1980) holds that problem solving is at the heart of the community of mathematicians. Polya (1945, 1954) and Schoenfeld (1992) view mathematics as a social activity, where engagement by the community includes discovery and opportunity to test conjectures.

However, engaging in mathematics as a social activity that includes discovery, for example, is not the norm for most students. Rather, developing fluency with a defined set of knowledge (i.e., computation) is the hallmark of traditional western mathematics lessons (Stigler & Hiebert, 1999). Many believe (e.g., Brown, Collins, & Duguid, 1989; Resnick, 1997) the traditional mathematics classroom teaches children more about how to be students (e.g., sit quietly, listen, reproduce) than how to be mathematicians (e.g., problem solve, reason, conjecture). As a result, children learn to participate in a type of inauthentic community—school mathematics. This is problematic: research suggests that the way students interact during mathematics is a powerful influence on their identity within that community (Boaler, 2002; Bishop, 2012). Students of mathematics, should participate and negotiate mathematical ideas in a similar way as mathematicians.¹²

¹² To be clear, I am not asserting computation and basic skills are not important. “Learning to participate in the discourse of conceptual meaning and learning basic routines of symbol manipulation can both be seen as significant assets for student participation, rather than being orthogonal objectives” (Greeno, 1998, p.17). Indeed, both basic skills and conceptual understanding are needed for students to understand what mathematics is: one needs to engage in the mathematical practices to understand the content as much as one needs content to use the practices.

The recently crafted Common Core State Standards (National Governors Association Center for Best Practices, 2010, pp.7-8) includes a list of mathematical practices that serve as an example of the type of thinking and reasoning needed to be mathematically literate:

1. Make sense of problems and persevere in solving them
2. Reason abstractly and quantitatively
3. Construct viable arguments and critique the reasoning of others
4. Model with mathematics
5. Use appropriate tools strategically
6. Attend to precision
7. Look for and make use of structure
8. Look for and express regularity in repeated reasoning

As suggested by these recommended practices, in order to be mathematically literate, students need to be able to do more than just compute facts and execute procedures. But in order to learn these other practices, students need to be given repeated opportunities to participate in them. Teachers play a large role in enabling or constraining these opportunities.

2.3 THE TEACHER'S ROLE IN LEARNING

From a situative perspective, students' cognitive development is inseparable from the way they are guided through material. Thus, the situative perspective on learning has a strong affinity with the idea of cognitive apprenticeship (Collins, Brown, & Newman, 1989; Dennen, 2004; Rogoff, 1990, 1991). Barbara Rogoff (1990, p.39) defines this approach:

“The apprenticeship system often involves a group of novices (peers) who serve as resources for one another in exploring the new domain and aiding and challenging one another. Among themselves, the novices are likely to differ usefully in expertise as well. The “master,” or expert [teacher], is relatively more skilled than the novices, with a broader vision of the important features of the culturally valued activity... The model provided by apprenticeship is one of active learners in a community of people who support, challenge, and guide novices as they increasingly participate in skilled, valued sociocultural activity.”

The term apprenticeship helps to emphasize the centrality of participation in the situative perspective on learning. In this approach, the students’ and teacher’s roles are entwined: students do not behave as “students,” in the traditional sense, but as active participants working with the teacher in authentic practices. Concepts are developed out of and through continuing authentic activity. Understanding of knowledge is conceived as a process that unfolds through connected discourse that is more like natural conversation than traditional transmission styles of teaching (Tharp & Gallimore, 1988).

Various interventions designed around cognitive apprenticeship have yielded positive results with respect to traditionally valued educational outcomes: increased learning gains and transfer (Coltman, Petyaeva, & Anghileri, 2002; King, 1999; Palinscar & Brown, 1984), higher levels of engagement (Webb, Troper, & Fall, 1995), more favorable reaction to instruction (Hendricks, 2001), and higher levels of mastery (Chang, Sung, & Chen, 2001). A common characteristic in these interventions is the belief that students’ cognitive development is inseparable from the way they are guided through material.

For example, studies have reported that the ways teachers use talk as a tool in the classroom have a significant effect on how their students use talk as a tool in their own learning (e.g., Fisher & Larkin, 2008; King, 1994; Webb, Nemer, & Ing, 2006). King (1994) discusses a study in which teachers modeled how to guide discussions using “integration” questions – questions meant to promote connections among ideas. When comparing students from this “apprenticeship” to a control group, post-intervention analyses on student verbal interaction reported no differences in the *number* of questions students asked each other while working without the teacher. However, there were indeed differences in the *kinds* of questions students asked. Students in the control group asked significantly more factual questions, while the treatment group asked more “integration” questions (the same type of questioning techniques that were part of their apprenticeship). Furthermore, a post-treatment comprehension test revealed that the apprentices scored significantly better than controls on literal comprehension and inference/integration items. The apprentices also scored significantly higher on a separate assessment of retention.

Other evidence suggests students will interact with each other at a rather unproductive, basic level in the absence of the apprenticeship structure (Britton, Van Dusen, Glynn, & Hephill, 1990; Ellis & Rogoff, 1986; Hogan, Nastasi, & Pressley, 2000; King, 1999; Koester & Bueche, 1980; McLane, 1987; Nathan & Knuth, 2003). These findings will be further discussed in Section 2.5.

Social constructivist perspectives of learning have shed light on how teachers use discourse in the apprenticeship model as a tool to mediate students’ participation (e.g., Daniels, 2001; Halliday, 1993; Hicks, 1995; Moll, 1990; Tharp & Gallimore, 1988; Vygotsky, 1978,

1987; Wells, 1994). For example, the teacher can press students to explain their different perspectives, highlight important findings, and connect multiple ways of thinking.

Teachers often successfully mediate learning and increase students' participation through the use of *scaffolded dialogue*—achieving common understanding through structured and sequenced questioning of joint activity (Alexander, 2000; Edwards & Mercer, 1987; Halliday, 1993; Tharp & Gallimore, 1988). Through *scaffolded dialogue*, students are given a sense of agency, where they have the opportunity to take control of their own mental activity (Bruner, 1996; Palincsar & Brown, 1984). As the “more experienced” member of the community, the teacher uses such dialogue to navigate students through tasks by providing opportunities for discussion and reflection of the knowledge being negotiated. In this sense, knowledge becomes bi-directional, where both the teacher and student contribute to the shared knowledge and idea creation of the classroom (Edwards & Mercer, 1987; Lave & Wenger, 1991). Rogoff (1990, p.196) describes this bi-directional knowledge creation: “The participants gain in understanding and may have difficulty determining ‘whose’ idea an insight was; many claim an insight as their own and cannot trace it to the group discussion. Indeed, it was theirs, but not theirs alone. The insights of such coordinated discussion are theirs as participants in the process.”

Participating in the discourse surrounding relevant tasks of a community enable novices to become legitimate participants in that community (Driver et al., 1994; Lave & Wenger, 1991). Thus, a reasonable place to begin analyzing how teachers provide opportunities for students to participate as mathematicians is at the level of classroom discourse.¹³

¹³ Although not the focus of this study, it is important to note teachers' ability to provide students with opportunities to participate in various mathematical practices is largely shaped, enabled, and constrained within relations of power. For example, a teacher may be forced to teach tasks from a curriculum that severely limits opportunities for problem solving, reasoning, and so on. Furthermore, a teacher may be evaluated in such a manner that distorts her teaching in a way that ensures “high marks.” In short, when

2.4 CLASSROOM DISCOURSE

It is at the level of classroom discourse that one can begin to understand how very different classroom communities can be from one another, despite the structural similarities. Two classrooms can have the same seating arrangement with the same amount of students. Both classrooms can be observed to have teacher-student interactions. The classrooms could even be working on the very same task. It may not be until you arrive at the level of discourse that one will “see” true variation in students’ opportunities to participate. Different teacher questions produce different responses (Franke et al., 2009; King, 1999) and take students through different intellectual terrain (Hogan, Nastasi, & Pressley, 2000; Lampert, 2001).

Analyzing classroom discussion is certainly nothing new for education researchers. Throughout the years, two robust findings appear: teachers ask a lot of questions (e.g., Floyd, 1960; Gall, 1970; Stevens, 1912) and the majority of those questions ask for a simple recall of facts (e.g., Boaler & Brodie, 2004; Corey, 1940; Haynes, 1935; Kawanaka & Stigler, 1999; Lee & Kinzie, 2012; Stevens, 1912).

In her meta-analysis of teachers as “professional question makers,” Gall (1970) points out that most early studies of classroom discourse report on what questions teachers ask, but typically do not discuss the type of questions theory suggests ought to be asked. Moreover, Gall points out, these early studies are silent on which questions are linked to which student outcomes. Further, Gall reports, despite the growing body of literature on classroom discourse, the field of educational research knows very little about different patterns of teacher discourse.

studying *why* opportunities to participate may or may not exist, which the present study does not, the teacher cannot be the only unit of analysis. The hegemony over opportunities to participate in various mathematical practices might actually lie somewhere higher in the education nest.

“In fact, most question-classification systems do not take them into account since the systems are not concerned with question sequence(s)” (p. 712).

Four decades have past since Gall’s report; yet the majority of mathematics education research on classroom discussion continues to focus on the number and type of question teachers ask. For example, Brodie (2008) developed a set of codes that describe how teachers interact with their students’ contributions. She illustrates how one teacher changed his questioning habits as he began to enact a new “reform” curriculum. Brodie reports that in the first week of her observation the teacher followed up about half (48%) of the student responses with low-level “elicit” moves. “Elicit” moves funnel students’ responses and put constraints on their thinking. Brodie notes that during the second week of her observations the teacher only followed up about a quarter (23%) of the student responses with “elicit” moves. Instead, he began to “press” his students by asking them to explain their thinking, for example.

By focusing on teacher questions in isolation from one another, such analyses do not recognize that a question’s surrounding context may change its utility and position students (Harré & van Lagenhove, 1999) into different opportunities to participate.

Analyses that move beyond isolated questions tend to focus on triadic dialogues (Lemke, 1985). The modal triad in many countries consists of a teacher asking a question to a student, the student responding, and the teacher evaluating the response (Abd-Kadir & Hardman, 2007; Nystrand et al., 2003; Smith, Hardman, Wall, & Mroz, 2004). This triad of discourse is often referred to as the Initiation-Response-Evaluation (IRE) pattern (Mehan, 1979) or the Initiation-Response-Feedback (IRF) pattern (Sinclair & Coulthard, 1975; Wells, 1993) of discourse.

Empirical studies suggest that moving beyond the IRE pattern of discourse achieves traditionally valued educational outcomes. For example, Adey, Shayer, & Yates (2001) report that, when compared to a control group, British national achievement test scores (for science, mathematics, and literacy) rose for students who participated for two consecutive years in a program focused on students articulating and explaining their solutions. In these treatments, instead of simply evaluating students' responses, teachers more often asked for further explanation, clarification, or challenged student responses over a series of turns. Similarly, Rojas-Drummond and Mercer (2003) found that teachers whose students achieved the highest outcomes, as measured by increased instances of reasoning and increased scores on the Raven's Progressive Matrices¹⁴, used question-and-answer sequences to guide the development of understanding, not just to evaluate knowledge. These studies, among others (e.g., Hiebert & Wearne, 1993; Howe et al., 2007; King, 1989; Lam, Law, & Shum, 2009), illustrate some of the benefits of participating in a discourse that more closely aligns with authentic discourse—the ordinary discourse practices of a community (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991; Tharp & Gallimore, 1988).

Similar to the analysis reported herein, Nathan and Knuth (2003) focused on teacher-student interactions during whole-group discussions of mathematical content. They found interactions, such as who talks to whom and who is accountable for what, are largely shaped by the decisions made by the classroom teacher. Nathan and Knuth's analysis, however, differs from the present study in many ways. Their analysis centered on the relationship between teachers' instructional decisions during whole-group discussion and their personal beliefs (the present analysis centers on the relationship between teachers' instructional decisions during

¹⁴ The Raven's Progressive Matrices is a widely used measurement of general reasoning.

whole-group discussion and an underlying theory of learning). Moreover, while the present analysis intends to frame teaching moves with respect to different types of participation afforded to students, Nathan and Knuth framed teaching moves with respect to the type of scaffolding they provided (e.g., analytic scaffolding, social scaffolding). Further, Nathan and Knuth did not examine the relationship of types of interactions and student outcomes; thus, the authors were silent on differences between “effective” teachers and their “less effective peers.”

Stronge, Ward, and Grant (2011) did examine the different questioning practices between “effective” teachers and their “less effective peers” (as measured by value-added estimates) and found no significant difference. However, they used very broad buckets to categorize questions—low, intermediate, and high cognitive demand—and only counted the number of questions per minute. The authors did not differentiate between questions asked during work on academic content and questions asked during other moments of the lesson. Further, Stronge et al. did not provide any detail regarding how teachers were, or were not, using student responses.

2.5 CONDUCTING WHOLE-GROUP DISCUSSIONS

Teacher-guided whole-group discussions can take on the elements of a good apprenticeship: they are organized around the accomplishment of a task, where skilled teachers can make usually hidden processes overt and connect shared elements to the underlying mathematical concept. The discourse choices teachers make during these discussions affect students’ opportunities to learn (Gee, 2008). If negotiated properly, whole-group discussions provide opportunities to participate in various mathematical practices (e.g., justifying work, connecting ideas, critiquing the reasoning of others). To be clear, it is not just the act of having a whole-group discussion

that is of concern. Rather, it is how the teacher interacts with students during the discussion that is of interest to this study.

Whole-group discussions that focus on conceptual understanding provide an opportunity to enable student reflection (Cobb et al., 1997; King, 1999). During group discussions, students may be confronted with ideas that diverge from their way of thinking. This creates a discrepancy and a need to reevaluate (and sometimes defend) their own thinking while simultaneously trying to grasp a new perspective. When interacting with others during whole-group discussions, students often discover that their own assumptions and understandings of the material differ to a greater or lesser extent from those of their peers with whom they are interacting. “When these conceptual discrepancies emerge, individuals often feel the need to reconcile them. To do so, they must negotiate understanding and meaning with each other. This meaning negotiation, this construction of knowledge, occurs through individuals’ explaining concepts to each other... working alone would not result in the same extent of cognitive exchange” (King, 1999, p.89).

This “cognitive exchange” often leads students to better understand their initial thinking (Mirza & Perret-Clermont, 2009; Mungy et al., 1981). Lefstein (2010) refers to this cognitive exchange as a sharing of horizons, where one uses another’s perspective as leverage for self-understanding and possible revision of her original horizon (see also Gadamer, 1998). Indeed, discussing differences in perspectives is fundamental to the act of meaning making and cognitive development (Scott et al., 2010; Voloshinov, 1986; Wegerif, 2007).

However, as alluded to earlier, students negotiate these cognitive exchanges rather poorly without the guidance of a teacher (Britton et al., 1990; Ellis & Rogoff, 1986; Hogan, et al., 2000; King, 1999; Koester & Bueche, 1980; McLane, 1987; Nathan & Knuth, 2003). For example, students find it difficult to scaffold their peer’s thinking, preferring instead to focus on having

their partner complete the task rather than understand it (Koester & Bueche, 1980; McLane, 1987). Moreover, when reasoning without a teacher, it takes students more turns to shape their thoughts (Hogan et al., 2000). Nathan and Knuth (2003) found students opting to take a democratic vote on what the correct answer is, rather than challenging each other to explain their thinking. Furthermore, when compared to peer discussions, teacher-guided discussions are a more effective means of attaining higher-levels of reasoning, higher quality explanations, and providing links between current and new knowledge (Ellis & Rogoff, 1986; Hogan, Nastasi, & Pressley, 2000). “Teachers typically (initiate) several responsive sequences in a row, whereas students often (drop) the questioning after a single response by a peer” (Hogan, Nastasi, & Pressley, 2000, p.417). Thus, while peer interaction plays a large part in knowledge creation, it is the teacher, acting as the more experienced other, who has the ability to properly guide and scaffold students through the negotiation of a task (Mercer & Littleton, 2007).

Guiding students through whole-group discussions that focus on more than just the correct answer is challenging for the teacher mathematically, pedagogically, and personally: mathematically, teachers must be able to anticipate student thinking; pedagogically, teachers must learn how to appropriately negotiate student thinking; personally, teachers must be willing to enter a space of uncertainty (Burkhardt, 1988). Indeed, guiding students in whole-group discussion is as demanding on the teacher as it is on the students (Lampert, 2001; Leinhardt & Steele, 2005; O’Connor, 2001; Staples, 2007). As such, one can reasonably hypothesize that teachers will vary in the way they interact with their students during whole-group discussions. This study aims to capture and describe that variance.

2.5.1 Various dimensions of effective whole-group discussions

While orchestrating whole-group discussions, teachers need to be able to juggle many aspects of good teaching (e.g., teachers need to provide sufficient wait time, they need to try including as many participants as appropriate, they need to manage behavior). Two aspects of good teaching that are of interest in this study are the need for teaching to be (1) *accountable to facts and procedures*, and (2) *accountable to authentic discourse*.

By *accountable to facts and procedures*, I mean teachers' instruction needs to be mathematically correct: their teaching should be free of mathematical errors; it should be clear and complete.

By *accountable to authentic discourse*, I mean teachers need to orchestrate discussions that align with the ordinary discourse practices of the underlying community (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991; Tharp & Gallimore, 1988). Whole-group discussions should provide students the opportunity to participate in various types of mathematical thinking. Beyond just being mathematically correct (i.e., being accountable to facts and procedures), authentic discourse should provide students with opportunities to make sense of mathematics, to justify their solutions, to critique the reasoning of others, and to create and test generalizations, among other practices.

While these definitions are admittedly incomplete, they are useful for the present study insofar as they help highlight the notion that quality teaching involves more than having mathematically correct instruction. As such, it becomes easier to see how measuring one aspect of quality teaching (e.g., accountability to facts and procedures) might not tell us much, if anything, about other aspects of quality teaching (e.g., accountability to authentic discourse).

For example, student scores on traditional standardized assessments can be expected to align with the extent to which a teacher's instruction is accountable to facts and procedures. Traditional standardized assessments test basic facts and procedures. For the most part, if a teacher has mathematically correct instruction (on the same basic facts and procedures that are on the assessment), they should be expected to receive higher value-added estimates than those teachers with mathematically incorrect instruction.¹⁵

The examination of student scores on standardized assessments that test basic facts and procedures, however, cannot be expected to shed light on the extent to which students were engaged in other practices such as advanced thinking and reasoning. As a result, we cannot determine if a teacher's teaching is accountable to other practices, such as authentic discourse, simply by investigating their value-added estimates. Other measures of specific classroom practices are needed to gauge different dimensions of quality teaching.

2.6 MEASURING STUDENTS' OPPORTUNITIES IN THE MATHEMATICS CLASSROOM

As discussed in Chapter 1, value-added estimates are currently being used to make inferences about quality teaching. However, these estimates are generated using data from traditional standardized assessments, which measure a narrow range of skills. The situative perspective on

¹⁵ There are, of course, exceptions. For example, a fifth-grade teacher could have mathematically correct instruction of fifth-grade standards. But if the students in her classroom are, say, two years behind grade level (i.e., they do not possess the required background knowledge), the mathematically correct fifth-grade instruction might be too advanced for them. Thus, students may perform poorly on a standardized assessment of fifth-grade standards (even though their teacher had mathematically correct instruction). Moreover, excessive classroom behavior issues could mediate the effect of mathematically correct instruction (in a negative manner).

learning suggests that it is possible for students to perform well on such assessments without participating in a broad range of mathematical practices. Currently, we do not know much about the relationship between specific teaching practices and value-added measures of teacher quality.

Value-added estimates are created using scores from what Shepard (2000) refers to as objective, one-skill-at-a-time type tests. These types of tests do not capture behaviors beyond recall of facts and execution of procedures. Measures of effective teaching, rather, should look like what theory tells us effective teaching should look like and measure actual teaching practices (Knight et al., 2012).

In an attempt to create a theoretically based measure of one specific teaching practice—conducting whole-group discussions—this study uses a coding scheme, the Analyzing Teaching Moves Guide¹⁶, to capture, and score, various ways teachers guide students through whole-group discussions of a mathematical task. It then goes on to examine the relationship between this newly constructed measure and teacher value-added estimates. In so doing, the study will help us better understand the relationship between teaching that is *accountable to facts and procedures* and teaching that is *accountable to authentic discourse*.

2.6.1 The situative perspective of learning: selecting a grain size

Finding the right level of analysis (or grain size) for any scientific study is a major question. Different theories of learning have different grain sizes and, often, define learning in different

¹⁶The Analyzing Teaching Moves guide was developed at the University of Pittsburgh, Learning Research and Development Center by Kevin Ashley, Richard Correnti, Moddy McKeown, Peg Smith, Mary Kay Stein, James Chisholm, and Jimmy Scherrer.

ways. For example, in the cognitive perspective of learning, the level of analysis is the individual. And, from a cognitive perspective, learning is viewed as an unobservable process that causes unobservable changes in knowledge in an individual's mind (Koedinger, Corbett, & Perfetti, 2012). Instead of being observed, learning, from a cognitive perspective, is *inferred* from various assessments.

On the other hand, the situative perspective of learning has a larger grain size. The level of analysis focuses attention on participation in communities in which people interact with each other and with material in their environment (Brown, Collins, & Duguid, 1989; Greeno, 1997, 1998; Lave & Wenger, 1991; Resnick, Saljo, Pontecorvo, & Burge, 1991; Wenger, 1998). And, from a situative perspective, learning, for the most part, is observable and is viewed as the increased ability to participate in ways that are valued by other members of the community (Contu & Willmott, 2003; Wenger, 1998).

The usefulness of a particular perspective of learning depends on what specific questions are being asked. The present study adopts the situative perspective of learning because its grain size—social interactions—and its view of learning—changes in participation—align with the study's research questions (discussed below).

What is learned depends heavily on what occurs during instruction. Classroom instruction (i.e., teaching) enables or constrains the opportunities students have to engage in various practices. Thus when trying to measure these opportunities, we need an instrument that highlights the social interactions between teachers and students during instruction. Specifically, this study is interested in teacher-student interactions during whole-group discussion of a mathematical task.

2.6.2 The Analyzing Teaching Moves Guide: theorizing effective discourse

Teacher-student interactions around discourse provide an important methodological lens for understanding relationships between teaching and learning. Through teacher-student interactions, pupils are positioned into, and take up, identities of agency or passivity (Solomon & Black, 2008).

The Analyzing Teaching Moves guide was designed on the premise that the teacher and student are mutually involved in knowledge construction. This belief precludes regarding either actor as independently definable. Teaching moves are thus defined with respect to the context provided by the students' previous contributions and interactions with the task. As a result, this study relies on a contextualist hypothesis (Pepper, 1942), asserting that consideration of how moves are situated in their surrounding context will paint a better picture of the interactions occurring within a community than analyzing questions in isolation. The following piece of transcript from Chapin, O'Connor and Anderson (2003, pp.12-15) illustrates.

- (1) Ms. D: Philipe, is 24 even or odd?
- (2) Philipe: Well, if we could use 3, then it could go into that, but 3 is odd. So then if it was ... but ... 3 is even. I mean odd. So if it's odd, then it's not even.
- (3) Ms. D: OK, so let me see if I understand. So you're saying that 24 is an odd number?
- (4) Philipe: Yeah, because 3 goes into it. Because 24 divided by 3 is 8.
- (5) Ms. D: Can anyone repeat what Philipe just said in his or her own words? Miranda?
- (6) Miranda: Um, I think I can. I think he said that 24 is odd because it can be divided by 3.
- (7) Ms. D: Is that right, Philipe? Is that what you said?
- (8) Philipe: Yes.

- (9) Ms. D: Miranda, do you agree or disagree with what Philipe said?
- (10) Miranda: Well, I sort of... like, I disagree?
- (11) Ms. Davies: Can you tell us why you disagree with what he said? What's your reasoning?
- (12) Miranda: Because I thought we said yesterday that you can divide even numbers by 2. And I think you can divide 24 by 2, and it's 12. So isn't that even?
- (13) Ms. D: So we have two ideas here about the number 24. Philipe, you're saying that 24 is odd because you can divide it by 3?
- (14) Philipe: Uh-huh.
- (15) Ms. D: And Miranda, you're saying it's even because you can divide it by 2? Is that correct?
- (16) Miranda: Yes.
- (17) Ms. D: OK, so what about other people? Who would like to add to this discussion...

In this piece of transcript, the teacher and students are mutually involved in knowledge construction. Simply counting the number of “low-level” and “high-level” questions in this example hides the nature of the interactions. The teacher presses students to explain their thinking (line 11), provides opportunities for students to agree or disagree with what has been said (line 9), offers space for different opinions (line 17), and has “set-the-table” for possible connections of student ideas. How the teacher is negotiating the mathematical ideas in this example and how the moves are situated in their surrounding context paints a better picture of the discourse than analyzing questions in isolation.

If opportunities to participate in mathematical practices are developed through socially supported interactions, then a tool is needed to capture those interactions. The Analyzing Teaching Moves guide is used in the present study to describe how whole-group discussion is being, or not being, socially supported by the classroom teacher.

The Analyzing Teaching Moves guide is divided into three categories: (a) initiating moves, (b) rejoinders, and (c) other.

Initiating Moves. Initiating moves include *Launch*, *Re-initiate* and *Literal*. These moves are used to describe how a teacher starts a discussion.

Launch. *Launch* refers to an open-ended question that invites student thinking. *Launch* questions are the building blocks of dialogue (Alexander, 2003) and provide the opportunity for students to formulate a response that makes sense to them. The following teacher utterance is an example *Launch*: “How might you solve this problem?”

Re-initiate. There are times when teachers repeat or rephrase their initial *Launch* question. These instances are coded as *Re-initiate*.

Literal. *Literal* refers to questions that ask for a retrieval of factual information: the “known-answer variety.” The following teacher utterance is an example *Literal*: “How many sides does a triangle have?”

Rejoinders. The second category of the Analyzing Teaching Moves guide is “rejoinders.” Rejoinders include: *Uptake*, *Collect*, *Connect*, *Lot*, and *Repeat*. In contrast to the traditional IRE pattern of discourse (discussed above), a rejoinder indicates that the teacher is not evaluating the student response in the traditional “right” or “wrong” manner. Rather than “allying” with some external standard which judges the appropriateness of a student contribution (i.e., evaluating), the use of a rejoinder signals to the students that the teacher is taking their view of the subject seriously, even though the teacher may want to extend or modify their views through further discussion (Barnes, 1976; Barnes & Todd, 1995; Mercer & Littleton, 2007).

Uptake. *Uptake* is an example of a *Launch*—open-ended question that invites student thinking—that uses a student response, usually in an attempt to extend, deepen, clarify, or

elaborate the discussion. The following teacher utterance is an example *Uptake*: “Juan just said that the answer must be wrong because the number is not even. Why do you think Juan believes the answer should be even?”

Collect. *Collect* is an instance where a teacher attempts to gather additional responses to a question. The following teacher utterance is an example *Collect*: “I would like someone else to share how they solved the problem?”

Connect. *Connect* is an instance where a teacher tries to get students to make an explicit connection. The following teacher utterance is an example *Connect*: “What do you notice that is similar in Maria and Rob’s solutions?”

Lot. *Lot* are instances where the teacher acknowledges a student contribution and indicates that the class will discuss it at a later time. The following teacher utterance is an example *Lot*: “I hear what you are saying, but we are going to come back to that later.”

Repeat. *Repeat* is an instance where a teacher echoes a student response.

Other Moves. The third category of the Analyzing Teaching Moves guide is “other.” Other includes: *Provides Information*, *Think Aloud*, *Terminal*, and *NC*.

Provides Information. *Provides Information* are instances where the teacher “tells” students information related to the instructional task. The following teacher utterance is an example *Provides Information*: “Use multiplication to solve these problems.”

Think Aloud. *Think Aloud* are instances when the teacher talks about how she is thinking about a problem. *Think Aloud* can be thought of a “verbal metacognition” in which the students have an opportunity to “see” inside the teacher’s head as she negotiates a decision (e.g., deciding how to enter a problem, developing solution paths, struggling, considering alternatives, measuring reasonableness of results). This is different than telling students information and

procedures, as in *Provides Information*. The following teacher utterance is an example *Think Aloud*: “Okay, so I know this problem is similar to others that I have seen. I usually draw a picture when confronted with these types of problems, but this particular one deals with fractions. I wonder if using a number line would be easier to represent this problem? I am going to try both; sometimes I like to start by trying both and then think about the problem again once I have a little bit going.”

Terminal. *Terminal* is an utterance that discontinues a student’s response. An instance of *Terminal* often, but not always, implicitly or explicitly evaluates a student response. The following teacher-student exchange ends with a *Terminal* (that does *not* implicitly or explicitly evaluate a student response):

T: How might we start building a definition for parallelograms?

S: You could also call these shapes quadrilaterals.

T: I did not ask that. I asked, how might we start building a definition for parallelograms.¹⁷

NC. *NC* stands for Node Code and simply indicates that a given teacher turn cannot be labeled using one of the codes above. *NC*, for example, could be applied to an utterance dealing with discipline issues (e.g., “Sit down back there.”)

The categories (i.e., initiating moves, rejoinders, and other) are only used for organizational purposes and are not used in the analysis. The codes are not necessarily exclusive to one category. For example, a *Literal* question could be used to initiate a conversation, but it could also be used to respond to a student contribution.

¹⁷ Note that this teacher turn would be coded *Terminal*, *Reinitiate*. Indicating that the first utterance – *I did not ask for that* – is an instance of *Terminal* and the second – *I asked, how might we start building a definition for parallelograms* – is an instance of *Reinitiate*.

It is important to note that using the Analyzing Teaching Moves guide is not meant to highlight the student cognitive processes that the teacher attempts to set into motion; rather, the coding scheme is meant to highlight the type of participation afforded to students by specific teaching moves. For example, one instance of *Uptake* – an open-ended question that uses a student contribution – could very well be more cognitively demanding than another instance of *Uptake*. Regardless, the present study treats both instances as an opportunity to extend a student contribution, and does not evaluate the “quality” of each instance.

2.6.3 Conceptualization of the whole-group discussion score

While none of the codes on the Analyzing Teaching Moves guide (in isolation) are inherently good or bad, some moves are generally thought to do a better job of providing students the opportunity to participate in high-level mathematical practices, as opposed to simply recalling or computing facts. The ways in which teachers and students talk with each other should resemble the authentic practices of the underlying discipline (Shepard, 2000); in mathematics, this includes a range of practices. For the purpose of this study, different codes were awarded different points based on the type of move and how that move was situated in its surrounding social context.

The theoretically-based *whole-group discussion score* has two important characteristics. First, it identifies some codes as providing better opportunities to participate in the type of higher-level thinking that a community of mathematicians value. For example, it is important to provide students with the opportunity to discuss their work (captured in *Launch*), it is important to provide students with the opportunity to explain why they chose certain methods (captured in *Uptake*), and it is important to provide students with the opportunity to make connections

between ideas and solutions (captured by *Connect*). Therefore, some codes are awarded more “points” than others.¹⁸ Chapter 3 discusses the theoretical backing for the scores.

The second important characteristic of the whole-group discussion score is that it takes into consideration the context of the immediate surrounding codes. Thus, the whole-group discussion score moves beyond “question type” and begins to probe patterns of interaction. For example, the code *Uptake* represents an instance where a teacher is using a student response to further the discussion, perhaps by asking the student to justify her thinking or elaborate her idea. This move is generally thought to be useful in communities where members are held accountable for their thinking and where ideas and contributions are valued (Michaels, O’Connor, & Resnick, 2008). Hence, instances of *Uptake*, for example, are awarded more points than many other codes. But, pilot work on the coding scheme suggests that when *Uptake* is embedded in a series of moves within a single teacher turn, for example, the effectiveness of the *Uptake* is often lost. For instance, it is common to observe a teacher ask the class to elaborate on a student contribution (i.e., *Uptake*) and then immediately follow that move with a *Literal* yes/no question. As a result, the students tend to become cognitive economists (Hogan et al., 2000; Klaczynski, 1997) and direct their attention to the less cognitively demanding *Literal* question. In these instances, the original *Uptake* is “diluted” and would be awarded fewer points. The full set of scoring guidelines and theoretical backing can be found in Chapter 3.

¹⁸ It is also appropriate to sometimes ask simple recall of facts (captured in *Literal*). However, research has shown many teachers only ask their students simple recall of facts. For example, when analyzing tapes from the TIMSS study, Kawanaka and Stigler (1999) found that in 1/3 of the U.S. lessons that they coded there was not a single instance where the teachers asked a student to explain her thinking. Hence, beyond theoretical reasons (discussed in Section 3), some codes receive more points because they capture certain practices that students do not often have the opportunity in which to participate.

2.7 RESEARCH QUESTIONS

To extend beyond labeling teachers effective because they are accountable to facts and procedures (i.e., high VAM estimates), this study attempts to improve the measuring of one specific teaching practice—conducting whole-group discussions of a mathematical task. This study uses the Analyzing Teaching Moves guide and construction of whole-group discussion scores to answer the following research questions:

1. *To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers?*
2. *What is it about the participants' classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not converge with their value-added estimates?*

2.7.1 Hypotheses

Increased pressure to raise scores on standardized assessments has led some teachers to focus their instruction nearly exclusively on basic facts and procedures (Rothstein, Jacobsen, & Wilder, 2008). This is not surprising. There is a long history (in many fields) of individuals changing their behavior to perform well on quantitative measurements (Adams, Heywood, & Rothstein, 2009).¹⁹ The argument advanced herein is, although standardized assessments may do an adequate job of holding teachers accountable to facts and procedures, they do not hold teachers accountable to other practices, such as orchestrating authentic discourse. Stated differently, assessments of basic facts and procedures can convey how well a teacher teaches

¹⁹ In the social sciences, this phenomena is often referred to as Campbell's Law.

basic facts and procedures. But assessments of basic facts and procedures cannot convey how well a teacher teaches other mathematical practices.

As discussed in Section 2.5.1, *accountability to facts and procedures* means teachers have mathematically correct instruction: their instruction is free of errors; it is clear and complete. *Accountability to authentic discourse* means teachers orchestrate discussions that align with the discourse practices of the underlying community. Whole-group discussions should provide students the opportunity to participate in various types of mathematical thinking. Beyond just being mathematically correct (i.e., being accountable to facts and procedures), authentic discourse should provide students with opportunities to make sense of mathematics, to justify their solutions, to critique the reasoning of others, and to create and test generalizations, among other practices.

I hypothesize that by emphasizing instructional accuracy of basic facts and procedures with little or no attention to providing opportunities for students to participate in other mathematical practices, it is possible to have high standardized-assessment scores. That is, I hypothesize that a teacher's teaching can be accountable to facts and procedures, but not accountable to authentic discourse. Thus, using teachers' value-added estimates as the only proxy of effective teaching can be misleading.

With respect to accountability to facts and procedures and accountability to authentic discourse, there are four types of hypothesized teachers in this study: (1) teachers that are accountable to both facts and procedures and accountable to authentic discourse, (2) teachers that are accountable to facts and procedures but not authentic discourse, (3) teachers that are accountable to authentic discourse but not facts and procedures, and (4) teachers who are neither accountable to facts and procedures nor authentic discourse.

I predict that teachers who are accountable to facts and procedures, regardless if they are accountable to authentic discourse or not (i.e., type 1 and 2 teachers) will have higher value-added estimates when compared to teachers who are not accountable to facts and procedures (i.e., type 3 and 4 teachers). This prediction is based on the fact that traditional standardized assessments test basic facts and procedures. Thus, I hypothesize, teachers who teach those basic facts and procedures with precision will have higher value-added estimates than teachers who make a lot of mathematical errors when teaching those basic facts and procedures.

Teachers who pay little or no attention to mathematical practices beyond precision of facts and skills (i.e., type 2 teachers) can receive high value-added estimates. Since traditional standardized assessments do not test beyond precision of basic facts and procedures, teachers will not be penalized for ignoring other mathematical practices. I predict that this characteristic of traditional standardized assessments is what will cause some teachers' value-added estimates to not converge with their theoretically constructed whole-group discussion scores.

Teachers who are accountable to authentic discourse but not facts and procedures (i.e., type 3 teachers) will be the other source of non-convergence. This could be a result of the inability of the coding scheme presented herein to detect mathematically incorrect instruction. From the beginning stages of its design, the authors of the Analyzing Teaching Moves guide recognized the scheme was not meant to judge the mathematical correctness of instruction. Hence, the authors suggested a second coding be done by an expert in the field of mathematics education to determine the mathematical correctness of the instruction. Instead of an expert in the field of mathematics, the present study uses a teacher's Mathematical Quality of Instruction (MQI; Hill et al., 2008) score to determine the accuracy of instruction.

3.0 METHODOLOGY

Most current value-added estimates are generated using student achievement data on low-level standardized assessments. Socio-cultural theories of learning suggest that it is possible for students to perform well on these types of assessments without actually participating in ambitious teaching.

This study builds upon work completed by Hill and colleagues (Hill et al., 2011). This study investigates the whole-group discussion practices of the teachers in Hill et al.'s study that received high value-added estimates and high Mathematical Quality of Instruction scores (defined below). Its aim is to describe in detail how these teachers conduct whole-group discussions. Moreover, this study proposes a transparent, debatable measure of whole-group discussions. It goes on to investigate how well this measure converges with the value-added estimates of the teachers in Hill et al.'s study.

3.1 SAMPLE

The teachers used in this study come from a mid-sized district in the southwestern United States containing twenty-six middle schools ranging in size from fewer than 400 students to more than 1,100 students (Hill et al., 2011). The students at these middle schools are racially diverse (57%

Hispanic, 32% Caucasian, 5% American Indian, 4% African American, and 2% Asian) and 57% of them receive free or reduced-price lunch.

A small set of middle-school mathematics teachers ($n = 24$) were selected from four schools in the district. The four schools were purposefully selected so that they had diverse and stable value-added scores and similar demographic descriptors (Hill et al., 2011).²⁰ The teachers are similar to a national sample with respect to descriptive information (e.g., average years of experience, degrees); however, three of the teachers instructed their students in Spanish and several taught only gifted and talented students. Thus, it is possible the results reported herein do not generalize beyond this sample. The teachers in this sample already have been assigned two measures of “effectiveness”: a value-added modeling percentile ranking and a mathematical quality of instruction score.

Value-Added Modeling Percentile

Teacher VAM estimates were constructed for all middle school mathematics teachers in the district ($N = 222$). Specifically, three different covariate adjusted models were used to obtain teacher estimates (Hill et al., 2011): Model 1 (*Simple Model*) adjusted for prior student scores and a teacher random effect; Model 2 (*School Fixed Effects Model*) adjusted for prior student scores, a teacher random effect, and a school fixed effect; Model 3 (*Student Background Adjusted Model*) adjusted for prior student scores, a teacher random effect, and student background variables (see Appendix A for the actual models). As would be expected, all three models have high correlations but with some variability (see *Table 3.1*). The value-added estimates for the participants listed in *Table 3.1* were extracted from the entire data set ($N = 222$). Those teachers whose average VAM percentile rank is greater than or equal to 80 are classified

²⁰ It was expected that the variation in school value-added scores would increase the likelihood of variation in teacher quality.

as “effective” in the present study. This “cut-off” was selected based on previous research illustrating value-added modeling is best at identifying those teachers who are at the extremes (i.e., the top and bottom 20%), but have far less reliability distinguishing teachers in the middle of the distribution (Jacob & Lefgren, 2008). The teachers whose average VAM percentile rank is less than 50 are classified as “less effective peers” in the present study. This “cut-off” was raised from 20 in order to increase sample size. The list of participants and their VAM percentile rank for all three models can be found in *Table 3.1*.

Table 3.1 MQI and Value-Added Percentile Rank for Participants

Teacher	Value-Added Descriptives			Value-Added Percentile Ranks		
	MQI	n	FRL (%)	Model 1	Model 2	Model 3
Vince	2.50	110	39	90	84	83
Josephine	2.17	112	59	82	71	86
Dolly	2.25	116	41	95	93	95
Arthur	2.17	74	14	87	82	92
Gabrielle	2.17	95	52	89	87	87
Andrea	2.17	58	24	93	92	82
Hanna	2.17	31	55	96	96	93
Paloma	1.33	64	100	48	26	65
Alberto	1.33	92	65	31	51	25
Helene	1.33	74	38	9	4	3

Note: All teacher names are pseudonyms. MQI = mathematical quality of instruction score; *n* = number of students used to calculate value-added percentile ranks; FRL = percentage of students who qualified for free or reduced-price lunch; Model 1 = Simple Model; Model 2 = School Fixed Effects Model; Model 3 = Student Background Adjusted Model. All data used in *Table 3.1* come from Hill et al. (2011).

Mathematical Quality of Instruction Score

Hill et al. (2011) measured these same teachers using a classroom observation protocol known as the *Mathematical Quality of Instruction* (MQI; Hill et al., 2008). MQI scores (see

Table 3.1) are used in this study to differentiate teachers by the number of errors that they make during instruction. This differentiation plays an important role when trying to answer research question 2— *What is it about the participants’ classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not converge with their value-added estimates?*

The MQI observation protocol is intended to capture “the disciplinary integrity of the mathematics presented to students, including the degree to which teacher errors occur during the mathematics class”; it also captures, for example, precision in mathematical language and if correct generalizations are being developed in class (Hill et al., 2011, pp. 804-805). Scores are derived using a 3-point Likert-style rating: A score of one corresponds to lessons with significant teacher mathematical errors; a two corresponds to lessons with minor errors; a three is “reserved” for lessons that are relatively error free (Hill et al., 2008; Hill et al., 2011). Teachers were observed for six lessons and received an MQI score for each lesson. The MQI scores reported herein are averaged across the six lessons. In order to be classified as “effective” in the present study, teachers had to have an MQI score greater than or equal to 2.17 (i.e., relatively error free instruction). To be classified as a “less effective peer,” teachers had to have a MQI score less than or equal to 1.33 (i.e., instruction with significant errors). These “cut-offs” were selected to ensure discrimination of teachers: Mathematically, a teacher who receives an average MQI score of 2.17 scored a 2 or 3 in 67% of her lessons. A teacher who receives an average MQI score of 1.33 scored a 1 (the lowest possible score) in at least 67% of her lessons.

As can be seen in *Figure 3.1*, Hill et al. (2011) found a strong positive correlation between teachers’ value-added estimates and their MQI scores, providing evidence of convergent validity.

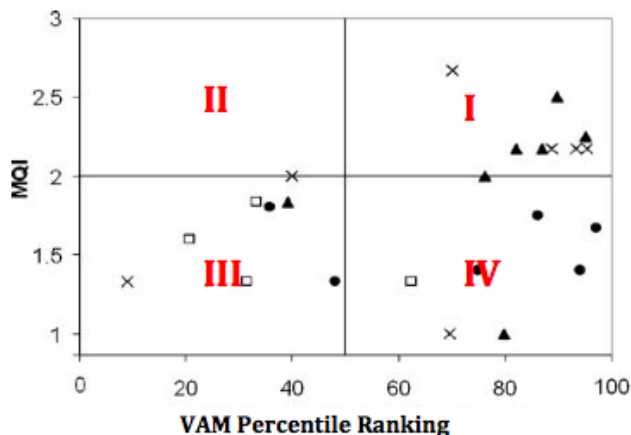


Figure 3.1: Mathematical Quality of Instruction (MQI) versus value-added estimate percentile rank.

Each symbol represents a different teacher; different symbols represent different schools.

From Hill et al.’s twenty-four subjects, the present study investigates the whole-group discourse practices of the most effective teachers ($n = 7$) and their least effective peers ($n = 3$). “Most effective” teachers are those with an average VAM percentile rank greater than or equal to 80 and an MQI score greater than or equal to 2.17 (i.e., the teachers situated on the right side of Quadrant I in *Figure 3.1*). Their “least effective peers” are those teachers in this sample with an average VAM percentile rank less than 50 and an MQI score less than or equal to 1.33 (i.e., the teachers situated in the bottom half of Quadrant III). The VAM percentile ranks and MQI score for each teacher can be found in *Table 3.1*.²¹

²¹ It is important to point out here that, if for some reason, an entire sample of teachers was determined to have unacceptable classroom instruction as measured by any observational instrument, a portion of that sample would still receive high value-added estimates. VAM uses a normative scale to rank teachers; as such, no matter how unacceptable classroom instruction may be, half of the teachers will be “above average.” Likewise, no matter how rich the instruction is, half the teachers will always be “below average.” This attribute of VAM is further motivation to always couple the estimates with other measures of “effectiveness.”

3.2 CODING

Classroom observations and video recording were scheduled based on simple criteria (no testing days, no field trips, “regular” instruction rather than special lessons designed for the study) and teacher convenience (Hill et al., 2011). For each teacher, half of the lessons were collected in January and half were collected in March. Six lessons per teacher were transcribed and coded for a total of sixty transcripts.

The sixty transcripts were first chunked into seven different segment types: *whole-class discussion*, *individual seatwork*, *group seatwork*, *review of work*, *guided practice*, *direct instruction*, and *transition/set up* (see Appendix B for a detailed description of each segment type). Then, two raters independently read and coded a subset of transcripts ($n = 12$).²² Coding of teachers moves during whole-group discussion was done through “systematic observation” (Mercer, 2010) using the Analyzing Teaching Moves guide. Although the raters read the entire transcript, only the *whole-group discussion* segments were coded. The two raters achieved 83% agreement; areas of disagreement were discussed and a consensus code was decided upon. The remaining transcripts were independently coded by the primary investigator.

3.3 ANALYSIS

A variety of data sources were used to answer the following research questions:

²²Eight of the lessons were from the January observations and four of the lessons were from the March observations. Two of the lessons did not include a whole-group discussion.

1. *To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers?*
2. *What is it about the participants' classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not converge with their value-added estimates?*

3.3.1 Data sources for research question 1

Two data sources were used to answer research question 1 – *To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers?* First, a whole-group discussion score was calculated for each teacher.

Whole-Group Discussion Score Rationale

A preponderance of questions alone is not correlated with student achievement, but the types of questions are (Graesser & Person, 1994). Different questions position students into different types of thinking (Solomon & Black, 2008) and are associated with different responses (Franke et al., 2009; King, 1999; Lee & Kinzie, 2012). However, recent evidence suggests that teachers struggle with how to negotiate student responses, especially responses to higher-level questions (Franke et al, 2009; Peterson & Leatham, 2009; Stein, Engle, Smith, & Hughes, 2008). As such, the present study moves beyond analyzing isolated questions. The locus of attention for the whole-group discussion score shifts to how questions are situated in their surrounding context.

As attention is shifted from individual questions to “surrounding context,” the location of opportunities to learn changes as well. Instead of being located in one single moment, it becomes stretched across the field of social interaction, across initial and follow-up moves. The result is that a question’s “utility” is largely dependent on how it is used and positioned within the field of social interaction.

The following side-by-side comparison of two pieces of transcripts provides an example.

Class 1	Codes	Class 2	Codes
<i>T: How might I be able to use a t-table to help me find the slope?</i>	Launch	<i>T: How might I be able to use a t-table to help me find the slope? What if I plotted two of these ordered pairs, could I find the slope from that?</i>	Launch, Literal
<i>S: You can look at the change in y.</i>	Uptake	<i>S: Yes</i>	
<i>T: Say more about that.</i>		<i>T: How might I be able to do that?</i>	Uptake
<i>S: Every time x goes up one, y decreases by 3. So, the slope is 3.</i>	Literal	<i>S: You could determine, like if you move one x, what is happening to y.</i>	
<i>T: Three, or negative three?</i>		<i>T: Where could I find that same information here on the t-table? I could find it right here in the y-column, right. I could see what happens to y when x changes.</i>	Connection, Provides Information
<i>S: Negative three; I mean negative three.</i>	Provides Information, Connection		
<i>T: Okay, if y decreases when x increases, the slope will be negative. Where do you see that same slope on this graph?</i>			

If one is only interested in the number and types of moves, these two pieces of transcript would be indistinguishable. They both include the same five moves: 1 *Launch*, 1 *Uptake*, 1 *Literal*, 1 *Provides Information*, and 1 *Connection*. Using the default scores on the scoring guide presented in *Table 3.2*, both of these pieces of discussions would receive seven points. However, when the position of the codes is taken into consideration, the two pieces of transcripts look different. In the following example, the text from the transcripts has been removed to highlight the different positions of the codes:

Class 1	Class 2
Launch	Launch, Literal
Uptake	Uptake
Literal	Connection, Provides Information
Provides Information, Connection	

Using the full scoring criteria presented in *Table 3.2*, the scores for Class 1 and Class 2 are different, seven and five respectively. The present study asserts this difference is important: The learning opportunities in Class 1 and Class 2 are not the same, even though the same exact moves were used. What follows is the theoretical rationale for the points assigned in *Table 3.2*.

It is important to note that quantifying and scoring a teaching practice runs the risk of others using the measurement tool and resulting metric to evaluate teachers. When that occurs, those being evaluated commonly change their behavior in unintended ways to “game” the scoring (Adams, Heywood, & Rothstein, 2009; Scherrer, 2011). With this in mind, beyond the theoretical rationale, many of the score “adjustments” are in response to possible “gaming.”

Table 3.2: Scoring Guidelines: Whole-Group Discussion Score

Scoring Guidelines: Whole-Group Discussion Score						
Code	3	2	1	0	-1	-2
Launch	never	<i>default</i>	<ul style="list-style-type: none"> When it is followed by any other code in the same turn 	<ul style="list-style-type: none"> When the preceding code is also a <i>Launch</i>, regardless of whether it is in the same turn or not 	never	never
Re-initiate	never	never	never	<i>default</i>	never	never
Literal	never	never	never	<i>default</i>	<ul style="list-style-type: none"> When the preceding two codes are also <i>Literal</i>, regardless of whether they are in the same turn or not When it is buried in a turn of four or more moves, regardless of its position 	<ul style="list-style-type: none"> When the preceding three or more codes are also <i>Literal</i>, regardless of whether they are in the same turn or not
Uptake	never	<i>default</i>	<ul style="list-style-type: none"> When the subsequent move in the same turn is a <i>Literal</i> When it is buried in a turn of four or more moves, regardless of its position 	never	never	never

Scoring Guidelines: Whole-Group Discussion Score

Note: *default* indicates the amount of points a move "starts" with before considering the surrounding context. Each move can only receive one score; therefore, if a move satisfies the description in two separate boxes, the move will be awarded the lesser score.

Code	3	2	1	0	-1	-2
Collect	never	<ul style="list-style-type: none"> When it is the last (or only) code in a turn and the subsequent turn begins with a <i>Connect</i> 	<i>default</i>	<ul style="list-style-type: none"> When the preceding code is also a <i>Collect</i>, regardless of whether it is in the same turn When it is buried in a turn of four or more moves, regardless of its position 	never	never
Connect	<i>default</i>	<ul style="list-style-type: none"> When it is followed by any other one (1) move in the same turn When the preceding code in the same turn is also a <i>Connect</i> 	<ul style="list-style-type: none"> When it is followed by any other two (2) moves in the same turn When it is buried in a turn of four or more moves, regardless of its position 	never	never	never
Lot	never	never	never	<i>default</i>	never	never
Repeat	never	never	never	<i>default</i>	<ul style="list-style-type: none"> When the previous turn also included a <i>Repeat</i>, regardless of its position 	<ul style="list-style-type: none"> When the previous two or more turns also included a <i>Repeat</i>, regardless of their position

Scoring Guidelines: Whole-Group Discussion Score

Note: *default* indicates the amount of points a move "starts" with before considering the surrounding context. Each move can only receive one score; therefore, if a move satisfies the description in two separate boxes, the move will be awarded the lesser score.

Code	3	2	1	0	-1	-2
Provides Information	never	never	never	<i>default</i>	<ul style="list-style-type: none"> When the preceding two codes are also <i>Provides Information</i>, regardless of whether they are in the same turn or not When the preceding two turns included <i>Provides Information</i>, regardless of its position 	<ul style="list-style-type: none"> When the preceding three or more codes are also <i>Provides Information</i>, regardless of whether they are in the same turn or not
Think Aloud	never	<i>default</i>	<ul style="list-style-type: none"> When another <i>Think Aloud</i> appears anywhere in the previous two turns 	<ul style="list-style-type: none"> When two instances of <i>Think Aloud</i> appear anywhere in the previous three turns 	never	never
Terminal	never	never	never	<i>default</i>	<ul style="list-style-type: none"> When the preceding turn also contains a <i>Terminal</i>, regardless of its position 	<ul style="list-style-type: none"> When the preceding two or more turns also contain a <i>Terminal</i>, regardless of their position
NC	never	never	never	<i>default</i>	never	never

A WGD score is calculated for each whole-group discussion. The score reported herein is averaged across all of a teacher's whole-group discussions.

Launch

Launch questions are the building blocks of dialogue (Alexander, 2003): an open-ended initiating question (i.e., *Launch*) can take students through a radically different terrain when compared to a closed-ended initiating question (Hogan, Nastasi, & Pressley, 2000; King, 1999; Lampert, 2001). Students are positioned into, and take up, identities of agency (rather than passivity) through *Launch* questions (Solomon & Black, 2008). These questions provide an opportunity for students to formulate a response that makes sense to them. For these reasons, the default score for *Launch* is 2 points.

However, *Launch* questions can be diluted if a teacher makes another move before students have the opportunity to respond to the *Launch*. A diluted *Launch* is an example of *Multiple Questions*, and is awarded 1 point. Students are often unsure of what question to attend to when a teacher turn includes more than one move (i.e., *Multiple Questions*). This introduces a situation where there is extraneous cognitive load, which can affect cognition and learning (Sweller, 1994; van Merriënboer & Sweller, 2005). A close examination of the scoring guidelines will highlight that most codes “lose points” when included in a teacher turn that includes *Multiple Questions*.

To prevent teachers from simply asking a series of *Launch* questions (without “doing anything” with the responses), a *Launch* is not awarded any points when the preceding code is also a *Launch*, regardless of whether it is in the same turn or not.

Re-initiate

Instances of *Re-initiate* are not awarded any points. They simply signal that a teacher is attempting to “re-ask” a previous *Launch*, *Uptake*, or *Connect*. It could be argued that a teacher with a preponderance of *Re-initiates* throughout the lesson is not asking clear enough *Launch* questions. However, this behavior was never encountered during pilot work and, thus, does not warrant text in the actual scoring guidelines.

Literal

There are instances when a *Literal* question can be useful (e.g., when a teacher wants to create joint attention by focusing students’ attention on something specific, or when a teacher wants to quickly clarify something). However, a preponderance of fact-based questions place limits on student discourse (Graesser & Person, 1994; King, 1999); these questions are not generally associated with high levels of reasoning. When students are not pressed to reason, they become cognitive economists (Hogan et al., 2000; Klaczynski, 1997) and conserve their reasoning effort knowing that either the teacher will be forthcoming with the information, or it will simply be dropped. Thus, the default score for *Literal* is 0.

A *Literal* is assigned -1 point when it is preceded by two codes that are also *Literal*, regardless of whether they are in the same turn or not. A *Literal* is also assigned -1 point when it is buried in a turn of four or more moves, regardless of their position.

Furthermore, a *Literal* is assigned -2 when it is preceded by three or more moves that are *Literal*. Edwards and Mercer (1987) refer to this type of sequence as “cued elicitation,” in which a teacher tries to control student input through multiple questions with a pre-determined answer in mind.

Uptake

Evidence suggests that the use of *Uptake* provides access for students to have the opportunity to truly comprehend material (Chi, Lewis, Reimann, & Glaser, 1989; Collins, 1982). Rojas-Drummond and colleagues (Rojas-Drummond & Mercer, 2004; Rojas-Drummond, Mercer, & Dabrowski, 2001) have found higher achievement in classrooms where the teacher used, among other things, “why” questions, encouraged students to make explicit their thought processes, and challenged students to give reasons for their views (all examples of *Uptake*). Others have shown that extending student contributions and building upon their ideas (i.e., *Uptake*) is an effective way to scaffold their thinking (Bereiter & Scardamalia, 1989; Palincsar, 1986). Furthermore, *Uptake* has been shown to be a major facilitator of language development. Referred to as *semantic contingency* (Snow, 1982), using a student response as an opportunity to reinforce (or introduce) labels through clarifying questions enhances students’ ability to use correct language while simultaneously stressing the involvement of the student in knowledge creation (Snow, 1979, 1982). Thus, the default score for *Uptake* is 2 points.

However, similar to *Launch*, an instance of *Uptake* can be diluted. When *Uptake* is followed by a *Literal* in the same teacher turn, it is only awarded 1 point. *Uptake* is also only awarded 1 point when it is buried in a turn of four or more moves, regardless of its position.

Collect

Skillful teachers tend to stay with the same topic for multiple turns (Hogan, Nastasi, Pressley, 2000). One way of doing this is to *Collect* multiple responses to a question. This “collection” of ideas sets the table for opportunities to connect ideas. Thus, the default score for *Collect* is 1. However, when *Collect* is the last (or only) code in a turn and the subsequent turn

begins with a *Connect*, 2 points are awarded to the *Collect* (in general, this “pattern” signals that a teacher is connecting the ideas that were just collected).

To discourage teachers from asking a series of *Collects* without “doing anything” with the responses – a pattern that was referred to as *The Road to Nowhere* during pilot work (Scherrer & Stein, 2013) – when a *Collect* is preceded by another *Collect*, regardless of whether it is in the same turn or not, it is awarded 0 points (the first instance of *Collect* is still awarded 1 point). A *Collect* is also awarded 0 points when it is buried in a turn of four or more moves, regardless of its position (The exception to this is when *Collect* is the last move in a turn, and the next turn begins with a *Connect*. In this instance, as stated above, the *Collect* would be awarded 2 points.).

Connect

In classrooms striving for authentic discourse, students are encouraged to build on each other’s knowledge and experiences to connect new knowledge to what is already known (Kumpulainen & Lipponen, 2010; Mercer & Littleton, 2007). These connections facilitate transfer and enhance learning (Gick & Holyoak, 1983; King, 1994), but are seldom made independently (D’Andrade, 1981). Thus an important role of the teacher in guided participation is to provide opportunities to create links (Rogoff & Gardner, 1984; Staples, 2007). Thus, the default score for *Connect* is 3.

But, *Connect*, too, can be diluted. When an instance of *Connect* is followed by any other one move in the same turn, it is only awarded 2 points. An instance of *Connect* is also only awarded two points when the preceding code in the same turn is also a *Connect*.

An instance of *Connect* is awarded 1 point when it is followed by any other two moves in the same turn. Further, an instance of *Connect* is also awarded 1 point when it is buried in a turn of four or more moves, regardless of its position.

Lot

Instances of *Lot* are not awarded any points. They simply signal that the teacher has acknowledged a student contribution and indicated that the class will discuss it at a later time. It could be argued that a teacher with a preponderance of *Lots* throughout the lesson is not actually responding to her students. However, this behavior was never encountered during pilot work and, thus, does not warrant text in the actual scoring guide.

Repeat

The default score for *Repeat* is 0. However, when a teacher continually echoes (i.e. *Repeat*) what students say, the class learns that they do not have to listen to their classmates. This is another example of students becoming cognitive economists (Hogan et al., 2000; Klaczynski, 1997), where they understand the teacher will be forthcoming with the relevant information. In these situations, the mutual accountability needed for a successful community of learners is lost. Thus, a *Repeat* is assigned -1 points when the previous turn also includes a *Repeat*, regardless of its position. A *Repeat* is assigned -2 points when the previous two or more turns also include a *Repeat*, regardless of their position.

Provides Information

There are instances when *Provides Information* can be useful. However, a preponderance of *Provides Information* (during whole-group discussion) could signal that a teacher is not allowing her students to struggle. As early as the 1970's evidence was emerging that illustrated the benefits of student struggle. For example, a study by Wood, Wood and Middleton (1978)

reported children who were given the opportunity to productively struggle were more capable of successfully transferring knowledge when they were on their own when compared to children who were simply told the answer. A large body of recent research by Kapur and colleagues has also reported the benefits of allowing students to struggle with ideas (e.g., Kapur, 2008, 2010, Kapur & Bielaczyc, 2012). Thus, although the default score is 0, *Provides Information* is assigned -1 point when the preceding two codes are also *Provides Information*, regardless of whether they are in the same turn or not. An instance of *Provides Information* will also receive -1 point when the preceding two turns include *Provides Information*, regardless of its position.

An instance of *Provides Information* will be assigned -2 points when the preceding three or more codes are also *Provides Information*, regardless if they are in the same turn or not.

Think Aloud

Previous research has reported that students of teachers who were skilled at modeling their thought processes when challenged with difficult material (i.e., *Think Aloud*) show a better awareness of affordances and constraints of strategies and better understanding of the context, and are likely to use the same type of “metacognition” when the teacher is not present (Duffy et al., 1986; King, 1999; Schoenfeld, Minstrell, & van Zee, 1999). Thus, the default score of *Think Aloud* is 2.

To discourage teachers from overusing *Think Aloud*, an instance of *Think Aloud* is only awarded 1 point when another *Think Aloud* appears anywhere in the previous two turns and is assigned 0 points when two instances of *Think Aloud* appear anywhere in the previous three turns.

Terminal

The default score for *Terminal* is 0. In general, a preponderance of *Terminal* signals that a teacher is not “doing anything” with student responses. Thus, an instance of *Terminal* is assigned -1 point when the preceding turn also contains a *Terminal*, regardless of its position. An instance of *Terminal* is assigned a -2 when the preceding two turns also contain a *Terminal*, regardless of their position.

NC

Instances of NC are always assigned 0 points.

Each whole-group discussion was scored twice: once using only the default scores in *Table 3.2*, and once utilizing the entire scoring scheme presented in *Table 3.2*. Then, for each teacher, scores were averaged across the six lessons and each teacher received an unadjusted whole-group discussion score and an adjusted whole-group discussion score. The two scores were then plotted on a coordinate grid to examine their relationship. The relationship was examined to determine if the adjusted whole-group discussion score provided a unique measure that went beyond the traditional “number and type of question” used for the unadjusted score

To help answer research question 1—*To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers*—teachers’ adjusted whole-group discussion scores and value-added estimates were plotted on a coordinate grid to investigate the extent the measures converge (or diverge).

3.3.2 Data sources for research question 2

To answer research question 2 – *What is it about the participants’ classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not*

converge with their value-added estimates? – an infographic was generated for each teachers’ whole-group discussions. On the left side of the infographics, the codes from the Analyzing Teaching Moves guide were listed vertically. Teacher turns during a whole-group discussion were listed horizontally across the top of the infographic: each turn had its own column. Different colored squares were placed in the columns to represent the various codes being used in each turn. Together, these elements created a “picture” for each whole-group discussion (Example “pictures” are included in Chapter 4 and in Appendix D). This picture was used to investigate the types of questions teachers asked, the frequency of each type, and how a teacher’s moves were situated in their surrounding context. Attempts were made to find patterns across each teacher’s six lessons. Patterns, when found, were described using sample text from the teacher’s transcripts. Then, descriptions of teachers’ whole-group discussions were used to help explain why their adjusted whole-group discussion scores do, or do not, converge with their value-added estimates. Teachers with similar descriptions were grouped together using the hypothesized “types of teachers” discussed in Section 2.7.1.

4.0 RESULTS

The results of the data analysis are reported in this chapter. The chapter is organized into two sections that correspond to the two research questions. Section 4.1 describes the relationship between theoretically constructed whole-group discussion scores and other measures of effectiveness for the teachers used in this study. Section 4.2 uses descriptions of teachers' whole-group discussions to help explain why some teachers' whole-group discussion scores do not converge with their value-added estimates.

4.1 THE RELATIONSHIP BETWEEN WHOLE-GROUP DISCUSSION SCORES AND VALUE-ADDED ESTIMATES

The results presented in this section pertain to the study's first research question: *To what extent do theoretically constructed whole-group discussion scores converge with value-added estimates for this sample of teachers?*

The whole-group discussions in each teacher's lessons were coded using the Analyzing Teaching Moves guide (Appendix C includes a list of tables that report the relative frequency of moves during each teacher's discussions). Then using the scoring criteria presented in *Table 3.2*, each discussion was scored in two ways. First, discussions were scored by adding the *default* scores for each code (see *Table 3.2*). Teachers' scores were averaged across their six lessons and

a *whole-group discussion score* (WGD) was assigned to each teacher. Then, each discussion was scored a second time. The second scoring used the complete set of scoring criteria presented in *Table 3.2*—that is, the second scoring took into account a code’s surrounding context. These “adjusted” scores were averaged across each teacher’s six lessons, and subjects were assigned an *adjusted whole-group discussion score* (WGD_Adj). Note that not every lesson in the data set included a whole-group discussion. When this occurred, teachers received a 0—because their students were not given an opportunity to participate in a whole-group discussion—in both the whole-group discussion score and the adjusted whole-group discussion score for that particular lesson. The two discussion scores for the subjects in this study are presented in *Table 4.1* along with each subject’s MQI and VAM estimates. The first column in *Table 4.1* identifies teachers’ average percent of turns devoted to conducting whole-group discussions during their lessons.

Table 4.1: Whole-Group Discussion, MQI, and VAM scores for Teachers

Teacher	% WGD	WGD	WGD_Adj	MQI	VAM
Vince	34	25.5	14	2.50	86
Josephine	14	15	9	2.17	80
Dolly	32	21	4	2.25	94
Arthur	14	7	3.5	2.17	87
Gabrielle	16	12	7	2.17	88
Andrea	22	15	3	2.17	89
Hanna	15	8	-8	2.17	95
Paloma	3	2	1	1.33	46
Alberto	24	3.5	-1	1.33	36
Helene	23	15	5	1.33	5

Note: % WGD = percent of turns that was whole-group discussion, averaged across six lessons; WGD = whole-group discussion score based on *default* scores only; WGD_Adj = whole-group discussion score based on complete set of scoring criteria presented in *Table 3.2*; MQI = mathematical quality of instruction score; VAM = average value-added percentile across the three models (see *Table 3.1*).

4.1.1 Comparing *amount* of whole-group discussion with value-added estimates

First, the relationship between the average percent of each teacher's lesson that was whole-group discussion (%WGD) was compared to VAM estimates ($r = .11$). As can be seen in *Figure 4.1*, simply having a higher dosage of whole-group discussion is not predictive of high VAM. This demonstrates that teachers currently being labeled "effective" through value-added estimates use whole-class discussions to varying degrees; some teachers with high value-added estimates use a lot of whole-class discussion while others use very little (the same occurs with teachers who have low value-added estimates). This suggests that if we are to find a correspondence between value-added estimates and whole-group discussion it will need to be with regards to the *kind* of discussion they are conducting.

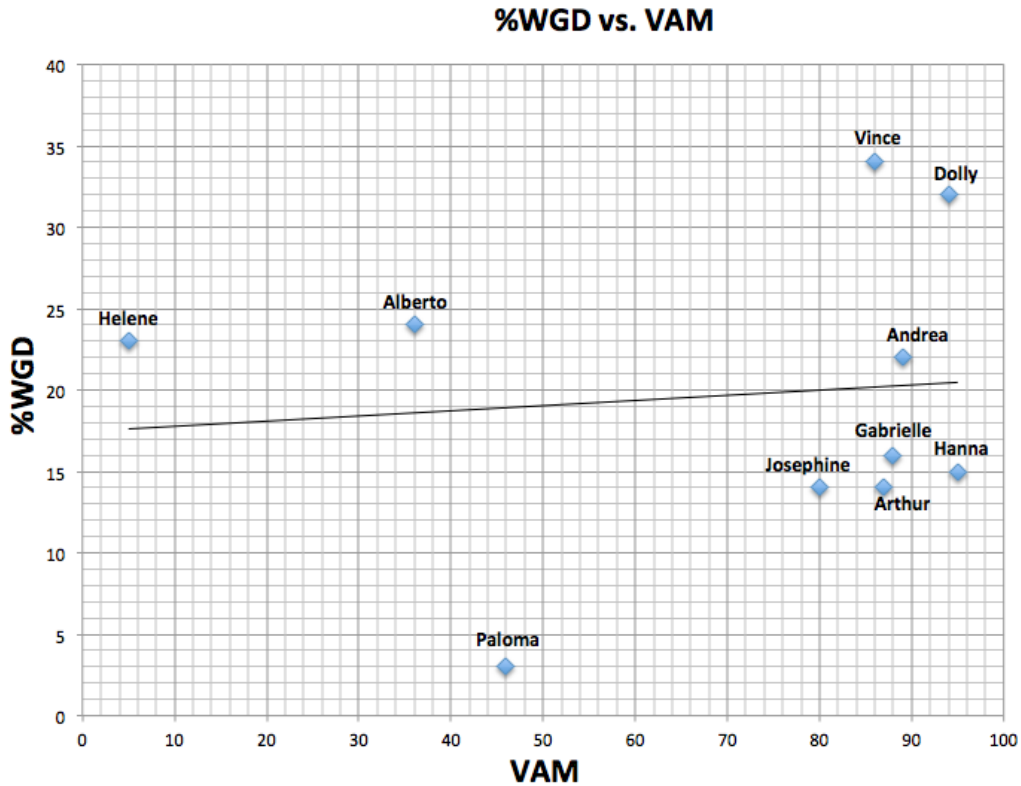


Figure 4.1: Average percent of each teacher’s lesson that was whole-group discussion (%WGD) versus value-added percentile rank (VAM)

4.1.2 Comparing teachers’ WGD and WGD_Adj scores

Figure 4.2 illustrates the relationship between teachers’ whole-group discussion scores (WGD) and teachers’ adjusted whole-group discussion scores (WGD_Adj). Not surprising, the sets of scores have a strong, positive correlation ($r = .69$). However, Figure 4.2 illustrates two important findings. First, the arithmetic range of the WGD scores is rather large (23.5), suggesting teachers conduct whole-group discussions in various ways (this is further investigated in Section 4.2). Second, while all teachers have an adjustment to their scores, the amount of adjustment differs. This holds true even among teachers who have the same WGD score (e.g., Andrea, Helene, and

Josephine). This suggests that the WGD score masks differences in discussion patterns. Hence, the adjusted whole-group discussion score (WGD_Adj) is a better metric for the *kind* of discussion that is occurring. That is, the adjusted whole-group discussion score allows the field of education to go beyond static descriptions of question type to describe various forms of whole-group discussion.

From this point forward I will restrict our attention to the adjusted whole-group discussion score (WGD_Adj).²³

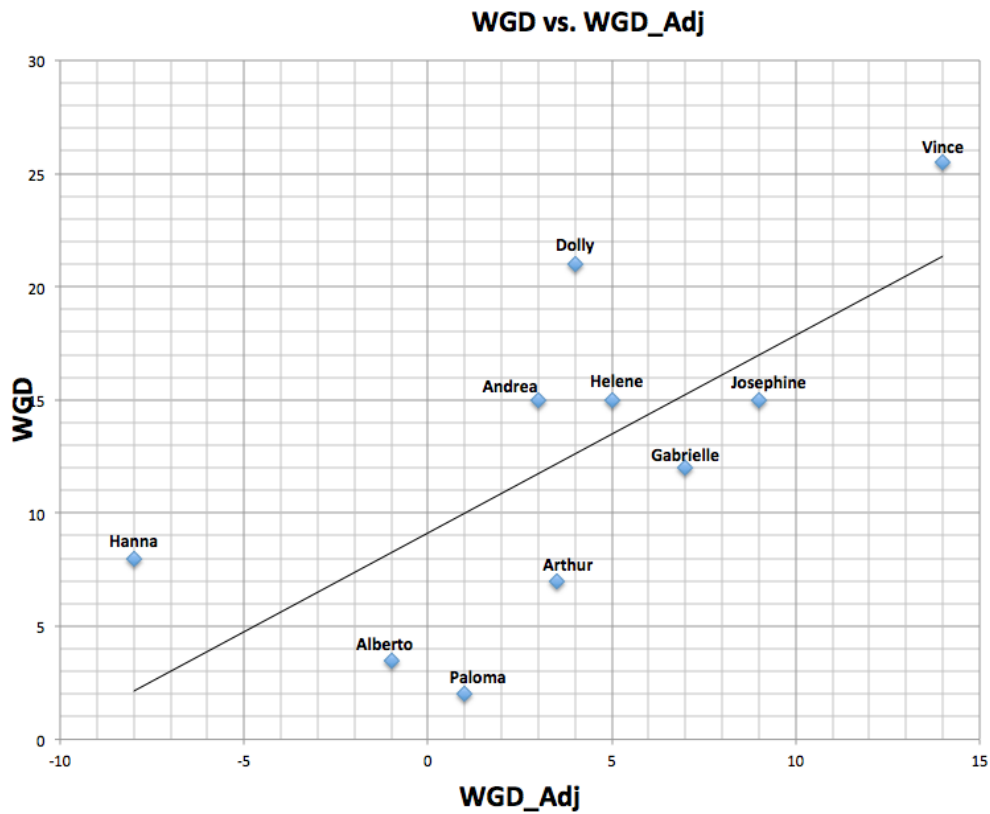


Figure 4.2: Whole-Group Discussion score (WGD) versus Adjusted Whole-Group Discussion score (WGD_Adj)

²³ Further rationale for focusing on the adjusted score can be found in the relationship between the whole-group discussion scores and the average percent of each teacher’s lesson that was whole-group discussion (%WGD). The correlation between the WGD score and %WGD is about twice as high ($r = .77$) as the correlation between the WGD_Adj score and the %WGD ($r = .39$). In other words, the WGD is skewed more simply by the amount of discussion. WGD_Adj score eliminates a significant portion of that noise.

The adjusted whole-group discussion score is constructed based on the theory and previous research discussed in Chapters 2 and 3: Instead of being located in one single moment (i.e., individual questions), opportunities to learn are stretched across the field of social interaction (i.e., across initial and follow-up moves). Theory suggests that a question's utility is largely dependent on how it is used and positioned within the field of social interaction.

4.1.3 Comparing teachers' VAM estimates with their adjusted whole-group discussion scores

Figure 4.3 illustrates the relationship between VAM estimates and adjusted whole-group discussion scores. The two sets of scores have a negligible relationship ($r = .07$), thereby confirming the assertion that it is rather dubious to make inferences about classroom instruction from value-added estimates. As can be seen in the *Figure*, there are two outliers: Helene and Hanna. Helene has the lowest value-added estimate in this sample of teachers yet has the fourth highest whole-group discussion score. Hanna has the highest value-added estimate in this sample of teachers, yet has the lowest whole-group discussion score. With the exception of these outliers, the two sets of scores converge rather well. That is, with the exception of Helene and Hanna, the teachers with the high VAM estimates (i.e., 80 or greater) have higher whole-group discussion scores than the teachers with the low VAM estimates (i.e., less than 50).²⁴

²⁴ With Helene and Hanna's scores removed, the correlation between the two sets of scores correlate at $r = .58$.

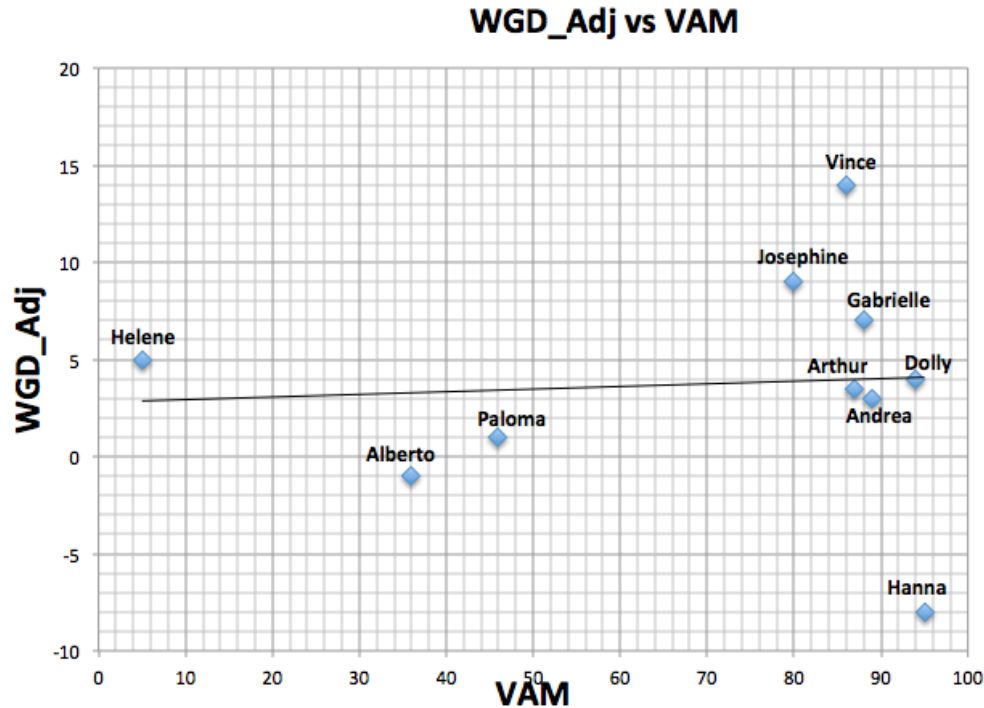


Figure 4.3: Adjusted whole-group discussion score (WGD_Adj) versus value-added estimate percentile rank (VAM)

4.1.4 Summary: Research question 1

In this section, we see that theoretically constructed whole-group discussion scores vary among teachers with similar value-added estimates. This suggests that even teachers who are considered equally “effective” through VAM conduct whole-group discussions in various ways.

We also see that theoretically constructed whole-group discussion scores converge with value-added estimates insofar that, with two exceptions, teachers with high value-added estimates had higher whole-group discussion scores when compared to their less effective peers (as measured by VAM). However, the relationship between these measures is weak. The weak relationship between theoretically constructed whole-group discussion scores and value-added estimates suggest that teachers being labeled “effective” through value-added measures in this

sample may not necessarily be conducting whole-group discussions in a way that provides students the opportunity to participate in the full range of mathematical practices.

In order to examine if these claims are accurate, a description of actual classroom practice is needed. Section 4.2 is intended to describe various ways teachers conduct whole-group discussions, and find out what it is about teachers' classroom practice that can help explain why some of the measures do not converge.

4.2 USING DESCRIPTIONS OF CLASSROOM PRACTICE TO HELP EXPLAIN WHY SOME OF THE PARTICIPANTS' MEASURES DO NOT CONVERGE

The results presented in this section pertain to the study's second research question: *What is it about the participants' classroom practice that can help explain why their theoretically constructed whole-group discussion scores did or did not converge with their value-added estimates?*

In the previous section we saw that a theoretically constructed whole-group discussion score for the teachers used in this study converge with value-added estimates. However, the relationship is weak and there are some interesting outliers. This section revisits the hypotheses put forth in Section 2.7.1 and explains why some of the measures did not converge. In so doing, this section goes inside the classrooms of teachers and describes the whole-group discussions that are occurring.

A note on terminology used in this section: This section uses the term *multiple questions* to describe teacher turns that include more than one code. There are exceptions: turns that

include *Repeat*, *Terminal*, or *Lot* and only one other code are not counted as instances of *multiple questions*. Consider the following example:

S: You can classify the shape as a quadrilateral.

T: A quadrilateral. Why? Can't I also classify it as a square?

The teacher turn above would be coded as follow: "A quadrilateral" (*Repeat*). "Why?" (*Uptake*). "Can't I also classify it as a square?" (*Literal*). This teacher turn is considered an instance of *multiple questions* because it includes two codes (*Uptake* and *Literal*) in addition to the *Repeat*.

Consider a second example:

S: You can classify the shape as a quadrilateral.

T: A quadrilateral. Why?

The teacher turn above would be coded as follows: "A quadrilateral" (*Repeat*). "Why?" (*Uptake*). This teacher turn is not considered an instance of *multiple questions* because it includes only one code (*Uptake*) in addition to the *Repeat*. The notion of *multiple questions* was discussed in Section 3.3.1.

This section also often combines the percent of a teacher's questions that are *Launch* and the percent of a teacher's questions that are *Uptake*. The sum of these two is referred to as the percent of "open-ended questions." For example, on average, 6% of Josephine's questions are *Launch* and, on average, 11% of her questions are *Uptake*. Thus, on average, 17% of Josephine's questions are open-ended.

Moreover, descriptions of teachers' discussion often include the relative frequency of certain codes (e.g., "On average, 11% of Josephine's questions are *Uptake*."). Appendix C includes a list of tables that includes the relative frequency of each code for each teacher.

4.2.1 Teachers accountable to facts and procedures and authentic discourse

There were six teachers who were accountable to facts and procedures that were also accountable to authentic discourse. That is, there were six teachers that had a high value-added estimate and a high adjusted whole-group discussion score: Vince, Josephine, Gabrielle, Andrea, Dolly, and Arthur. Interestingly, however, three different types of discussions were uncovered within this group of teachers.

Ambitious Discourse

First, Vince and Josephine exhibit what I will refer to as *ambitious discourse*. When compared to their peers not in this group—that is, their peers who are not accountable to both facts and procedures and authentic discourse—on average, a larger percent of Vince and Josephine’s questions are open-ended (16% and 17% respectively, compared to an average of 9% for teachers not in this group) and a smaller percent of Vince and Josephine’s questions are *Literal* (35% and 41% respectively, compared to an average of 46.5% for teachers not in this group).

Figure 4.4 includes the pictures of Vince’s discussions. These pictures are also representative of Josephine’s discussions.²⁵ The only notable difference between Vince and Josephine is the amount of whole-group discussion that each provides their students: Vince conducted a whole-group discussion in each of his six lessons, while Josephine conducted a whole-group discussion in four of her six lessons. Other than that difference, Vince and Josephine are strikingly similar. In fact, if Josephine’s adjusted whole-group discussion score total was divided by 4 (the number of her lessons that included a whole-group discussion)

²⁵ A picture of Josephine’s discussions can be found in Appendix D.

instead of 6 (the total number of lessons), her average adjusted whole-group discussion score would have been 14; the exact same as Vince.

Figure 4.4 (and all other “pictures of discourse” presented herein) can be read as follow: The top left of the Figure indicates which class the picture is illustrating. The rows in the Figure represent the codes from the Analyzing Teaching Moves guide. The codes are listed vertically on the left side of the Figure. The columns represent the teacher turns that were in the specific whole-group discussion indicated in the top left corner. The squares in each column represent the teaching moves that were used in that specific turn. For example, in *Figure 4.6*, during Vince’s class 1, he uttered an instance of *Launch* in his first turn during the whole-group discussion (represented by the orange square in column 1, in the “*Launch* row”). Then, in his second turn, Vince uttered an instance of *Repeat* and an instance of *Uptake*. The *Repeat* is represented by the light green square in column 2, in the “*Repeat* row.” The *Uptake* is represented by the yellow square in column 2, in the “*Uptake* row.” These two squares in the same column indicate that Vince made both of these moves in the same turn, before the students had a turn. Note, from the picture you cannot determine what move was uttered first.

When there are two instances of the same move in the same turn, it is represented by a white border around the square. For example, in turn 7 of Vince’s class 1, Vince uttered a *Repeat* (represented by a light green square in column 7, in the “*Repeat* row”) and he uttered two instances of *Literal* (represented by a dark green square with a white border in column 7, in the “*Literal* row”).

When there are three or more instances of the same move in the same turn, it is represented by a square with a white border and an “X” in the middle. For example, in turn 37 of Vince’s class 1, Vince uttered an instance of *Repeat* (represented by a light green square in

column 37) and at least three instances of *Literal* (represented by the dark green square with a white border and “X” in the middle). Note, a square with a white border and an “X” in the middle indicates that there were three or more instances of the same move in the same turn; it is not possible to determine exactly how many instances there were simply by looking at the picture.²⁶

As can be seen in *Figure 4.4*, Vince and Josephine use a wide range of moves (illustrated by the different colored squares in the Figure). Furthermore, relatively speaking to the teachers we will encounter below, you can see the large amount of open-ended questions that Vince and Josephine ask. These are represented by the orange and yellow squares in the first two rows of the figure.

²⁶ Although patterns did emerge for each teacher, there was still variation across teacher’s lessons. This variation can be seen in the picture of discourse by different combinations and number of squares. Quantitatively, the standard deviation for each teacher’s adjusted whole-group discussion score are as follow: Paloma, 1.7; Alberto, 3.9; Arthur, 4.5; Dolly, 4.7; Gabrielle, 8.7; Helene, 11; Vince, 12.5; Andrea, 12.7; Hanna, 12.8; Josephine, 15.2.



Figure 4.4: Pictures of Vince's discussions

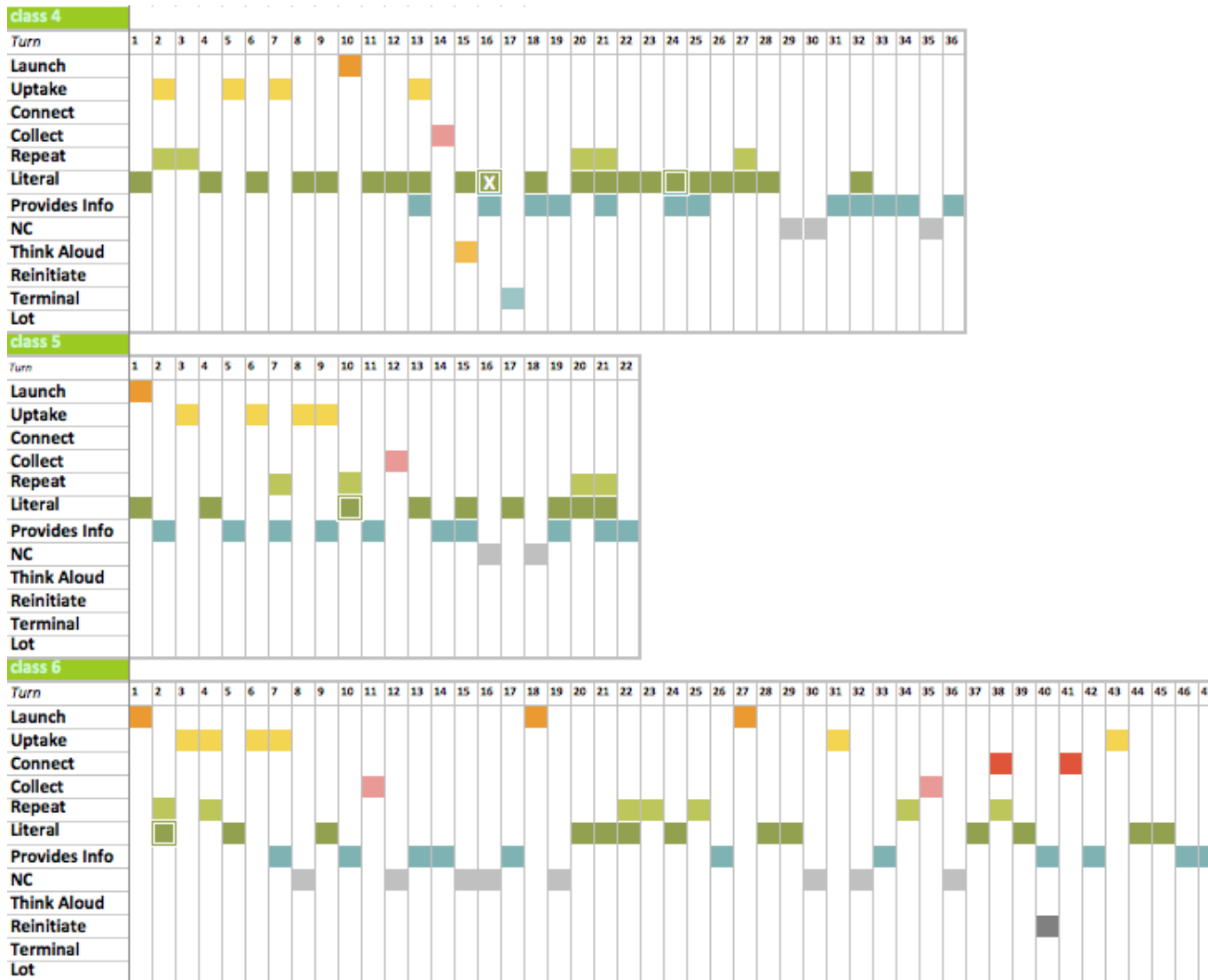


Figure 4.4: Pictures of Vince's discussions (continued)

Vince and Josephine often have their students present work; while they do, Vince and Josephine continually press their students to explain their thinking. The following piece of transcript from Josephine's class illustrates. In this example, a student is discussing the steps she took to solve a particular task:

- | | | |
|-----------|---|--|
| Student | Okay, you have to find the scale factor from the original triangle... the 3, and so you multiply | |
| Josephine | How do you know it's 3? | Uptake |
| Student | I divided by the...the square root of 9 is 3. | |
| Josephine | This is a tough one. The triangles, the triangles are a little tough. How many people had difficulties with the triangles? Yeah, they're a little tough. First of all, it's confusing because we've got three sides but we're only dealing with two dimensions, the base and the height, that's confusing for some people ... I guess my first question to you is, how did you find the area of the original? Walk us through that first. | Literal;
Provides
Information;
Uptake |
| Student | The area, I did 7 times 4, and that equals 28, and then 28 divided by 2 is 14. | |
| Josephine | Okay. So then in the problem it says "the area of the new triangle is 9 times greater than the original." 9 times greater than the original. So, what was your next step? | Uptake |
| Student | Then I found the scale factor from the original triangle to the new triangle. | |
| Josephine | How did you do that? | Uptake |

The above example illustrates how Vince and Josephine are accountable to authentic discourse. Beyond just looking for accuracy, these two teachers press their students to explain their thinking, justify their procedures, and occasionally provide students the opportunity to make connections between concepts. For example, in the following piece of transcript Vince

provides his students with the opportunity to make a connection between an equation and a t-table that both represent the same data:

Vince	Does it fit? Mentally double check. Put some X values from the table in place, replace X with it. Replace Y with the Y and make sure it's a true statement. I like it, works for me. Can I see this 2 in the equation – or, I'm sorry, in the table anywhere? Where is it, Jessica?	Literal; Provides Information; Connection
-------	---	--

Vince and Josephine are also accountable to facts and procedures; their value-added estimates are 86 and 80 respectively. This is not surprising since their MQI scores are 2.5 and 2.17, respectively. As described above, their students receive opportunities to participate in authentic discourse, which result in a relatively high whole-group discussion score when compared to their peers in this sample. Vince and Josephine's measures converge: they are accountable to facts and procedures and accountable to authentic discourse.

Diluted Ambitious Discourse

The second type of discussion uncovered while analyzing the teachers who are accountable to facts and procedures who are also accountable to authentic discourse was exhibited by Gabrielle, Andrea, and Dolly. These teachers exhibit what I will refer to as *diluted ambitious discourse*.

Although slightly less than Vince and Josephine, the portion of these three teachers' questions that are open-ended is relatively high when compared to the teachers not in this group—that is, teachers who are not accountable to facts and procedures that are also accountable to authentic discourse. Furthermore, like Vince and Josephine, Gabrielle, Andrea, and Dolly use a variety of moves with their students. The following piece of transcript from Gabrielle's class is illustrative. In this example, Gabrielle uses five different moves over the course of only six turns.

Gabrielle	So, what did you get, Susanna, for Foley?	Literal
Student	2/20ths.	
Gabrielle	2/20ths. Okay, so we broke these into 20ths, so you said this was 2/20ths. Someone have a different answer? Ben?	Repeat; Collect
Student	3/8ths.	
Gabrielle	3/8ths. Anyone have a different answer? Julian?	Repeat; Collect
Student	10/32nds.	
Gabrielle	10/32nds. Are any of these equivalent to each other?	Repeat; Connection
Student	No.	
Gabrielle	So, then someone's got to be wrong and someone's got to be right, huh?	Literal
Student	No...	
Gabrielle	You're wrong? Why are you saying you're wrong?	Uptake

What distinguishes these three teachers from Vince and Josephine is their tendency to dilute their discussions. For example, the following piece of transcript is representative of how Dolly's dilutes her whole-group discussions. In this example, notice the series of *Literal* questions and notice how Dolly repeats every student contribution:

Dolly	What do I multiply by to not change a number?	Literal
Student	1.	
Dolly	1. So my co-efficient is 1. And so what tells me whether it's increasing or decreasing, the co-efficient or the Y-intercept?	Repeat; Provides Information; Literal
Student	The co-efficient.	

- Dolly The co-efficient. So since it's a positive 1, would this be Repeat; Literal
 increasing or decreasing?
- Student Increasing.
- Dolly It would be increasing. What's that Y-intercept tell me? Repeat; Literal
- Student That that's where it's going to start.
- Dolly That's where it's going to start off at. So which of the Repeat; Literal
 graphs starts off at or crosses the Y-axis at negative 3?

When a teacher continually echoes (i.e. *Repeat*) what students say—as Dolly does in the above example—the class comes to understand that the teacher will be forthcoming with the relevant information; thus, they become cognitive economists (Hogan et al., 2000; Klaczynski, 1997) and stop expending the energy needed to listen to their classmates. In these situations, accountability to the learning community, in which participants listen to and build their contributions in response to those of others, is lost (Michaels, O'Connor, & Resnick, 2008).

Figure 4.5 includes the pictures of Dolly's discussions. Although, similar to Vince and Josephine (there is a good amount of open-ended questions—represented by the orange and yellow squares in the top two rows of the Figure), you will notice Dolly's tendency to *Repeat* student responses as represented by the frequent clusters of light green boxes across the turns. Almost a quarter (23%) of Dolly's moves are *Repeat*, and many of these instances occur across consecutive turns.



Figure 4.5: Pictures of Dolly's Discussions

Minus the preponderance of *Repeat*, the pictures of Gabrielle and Andrea's discussions look similar. However, their adjustments are a result of multiple series of *Literal* questions (pictures of Gabrielle and Andrea's discussions can be found in Appendix D). The following piece of transcript from one of Andrea's classes illustrates:

Andrea	Calvin, what's the next step?	Literal
Student	You need to divide by 2.	
Andrea	On both sides?	Literal
Student	Ah, yes.	
Andrea	Okay. So, and these two become?	Literal
Student	That becomes a 1.	
Andrea	Right. And 2 divided into 8 goes?	Literal
Student	4.	
Andrea	Great. All right. But then the next one is a little bit more tricky, right? We have this – oh, it's 5 – equals x minus 8. All right. So, Anthony, quickly tell me what to do here first.	Literal
Student	Okay, so it's 2X plus 10.	
Andrea	Right. Do I need to do anything on this side?	Literal
Student	No.	
Andrea	Okay, so I can just do this. Now what do I need to do to both sides?	Literal

These types of sequences are not the norm in Gabrielle, Dolly, and Andrea's discussions; these rare sequences do, however, differentiate these teachers from Vince and Josephine.

Gabrielle, Dolly, and Andrea are also accountable to facts and procedures; their VAM scores are 88, 94, and 89, respectively. This is no surprise since their MQI scores are 2.17, 2.25

and 2.17, respectively. Gabrielle, Dolly, and Andrea's measures converge: they are accountable to facts and procedures (high VAM) and accountable to authentic discourse (high WGD_Adj).

Student-Centered Discussion

Arthur exhibits the third type of discussion found among teachers who are accountable to facts and procedures that are also accountable to authentic discourse; I will refer to this third type as *student-centered discussion*.

One difference between Arthur and the rest of the teachers who are accountable to facts and procedures and authentic discourse is the amount of whole-group discussion provided. On average, Arthur spends 14% of his mathematics lesson conducting whole-group discussions. Whereas, on average, the rest of the teachers in this group spend 24% of their mathematics lessons conducting whole-group discussions. But there is another important difference that distinguishes Arthur from the rest of the teachers in this group: During whole-group discussion, a teacher turn is often followed by more than one consecutive student turn. The following piece of transcript is representative of Arthur's whole-group discussions:

Arthur	How does 3^{10} compare to 10^6 ?	Launch
Student	I put 3^{10} is less than 1 million, because 3^{10} is close to 9^3 , which is a lot less than 1 million, because 10^6 is 1 million.	
Student	Yeah, I did $3 \times 3 \times 3$ and then you get 9^5 and 10^6 , so you know 10^6 is bigger.	
Arthur	Oh, so 3^{10} is 9^5 ?	Literal
Student	Yeah.	
Arthur	And Shay you said it's 9^3 . Either way it's less than 10^6 , right? How did you guys figure out that 9^5 is equal to 3^{10} ?	Provides Information; Uptake
Student	Because you do $3 \times 3 \times 3^{10}$ times, and then you group them and you get 5 nines.	

Student	Couldn't you just take half of the exponent and then multiply it by itself?	
Students	[Cross talk between multiple students].	
Arthur	Hang on. One at a time. Okay Dave, what now?	Re-Initiate
Student	To find if 9^5 is equal to 3^{10} you half the 10, the exponent, and if you do that you have to double the base, which is 3.	
Arthur	Half the exponent, double the base.	Repeat
Student	You have to multiply it by itself.	
Arthur	Oh, square the base. Half the exponent, square the base. What were you going to say James?	Collect
Student	I was wondering does it always work as though you squared the numerator and half the power?	
Arthur	Square the numerator or square the base?	Literal
Student	Square the base.	
Arthur	Square the base and then half the power?	Literal
Student	Yeah.	
Arthur	Do you think it's going to work?	Literal
Student	Did you just say numerator?	
Student	I don't know.	
Arthur	Base. Does that make sense?	Literal
Student	Yeah.	
Arthur	Because if you're pairing off the 3s, what are you doing to the exponent?	Literal
Student	You're halving the exponent.	
Arthur	You're halving the exponent, huh? So it does work.	Repeat; Provides Information

Arthur's whole-group discussions usually proceed in this manner. He uses a variety of moves, rarely asks multiple questions in the same turn and, notably, his students do a lot of talking. These are all indicators that he is accountable to authentic discourse.

Moreover, Arthur appears extremely knowledgeable (his MQI score is 2.17). He easily (and correctly) makes connections between concepts and often presents material in multiple ways. His students appear equally gifted—often finishing Arthur's sentences and making connections on their own. Thus it is no surprise he has a high VAM estimate (87) and his measures converge.

Analysis of Arthur's discussions uncovers a possible limitation to the current iteration of the scoring guide presented herein. Arthur's students are very accountable to the learning community. They build upon the ideas of others, test their own generalizations, and often directly question each other (as opposed to “going through” the teacher to do this). Simply stated, in many discussions, Arthur does not have to do much for the conversation to include opportunities to participate in various mathematical practices—the students do it themselves. It is hard to determine if this occurs on the heels of Arthur's hard work setting-up this type of learning environment, or if his students are simply gifted. If it is the former, Arthur ought to get credit. But the coding scheme presented herein has no way to capture that.²⁷

4.2.2 Teachers accountable to facts and procedures but not authentic discourse

Hanna was the only teacher found to be accountable to facts and procedures but not authentic discourse. Hanna exhibits what I will refer to as *traditional discourse*. This type of discussion is

²⁷ Pictures of Arthur's discussions can be found in Appendix D

easy to see in her pictures of discourse (see *Figure 4.6*). There are few open-ended questions or connections (represented in the figure by a lot of white space in the first three rows). There is a preponderance of *Literal*—on average, over half (53%) of her moves are *Literal* (easily seen in the figure with the many dark green rectangles). Furthermore, on average, 30% of Hanna's turns include instances of multiple questions. This can be seen in *Figure 4.6* where there are multiple squares in the same column, or one square with a white border (which, as stated above, represents multiple instances of the same move in the same turn).

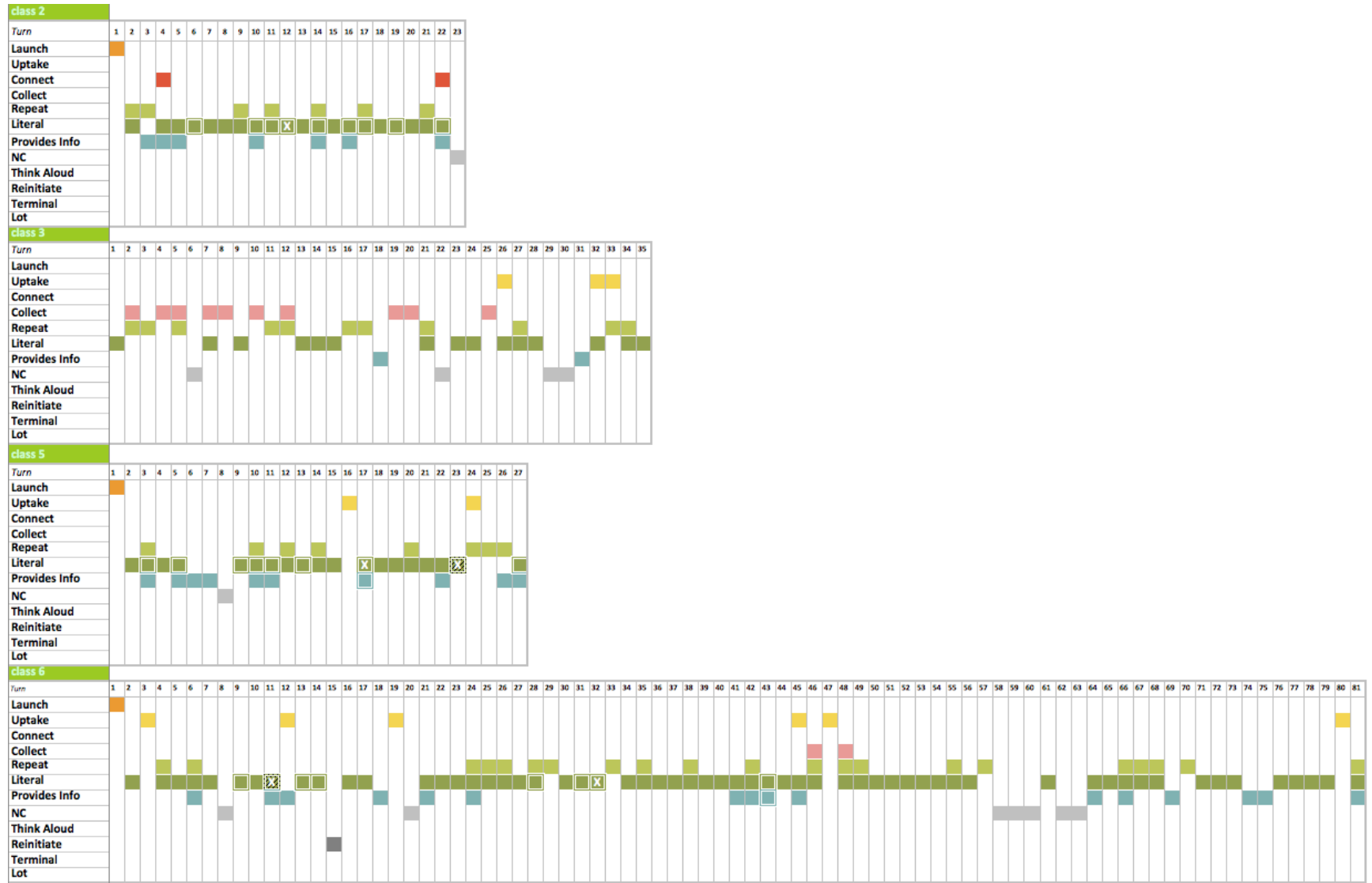


Figure 4.6: Pictures of Hanna's discussions

The following piece of transcript illustrates typical teacher-student interactions during Hanna's whole-group discussions: a preponderance of *Literal* questions and instances of multiple questions:

Hanna	These are three different points, okay? You've got to stick to it. So our run is... Which two points did you use to get 30,000? You've got to do that for your x. Put your fraction bar between them and you can divide them and find the unit rate. Okay. So now what's this going to be? It's 30,000, isn't it...divided by 150.	Provides Information; Literal; Provides Information; Literal; Literal
Student	It's just 2 – 200.	
Hanna	Which did you put in first?	Literal
Student	30,000.	
Hanna	Okay, 30,000 by what number?	Literal
Student	200.	
Hanna	200. All right. How's everybody doing?	Repeat; Literal
Student	Beautiful.	
Hanna	Did you get similar to her?	Literal
Student	I don't know.	
Hanna	Maybe she made a mistake and you can help her? You can help her, Andrew. Okay, so it's \$200 per foot. A little more expensive than the other problems. So now what's – okay, so Y equals – what about that B, that pesky B? It's hard when you don't have it on graph paper trying to guess what that is. His is a little more correct. So this is what you did, 25,000, 50,000, 75,000. You see, so how far up do you think that is?	Provides Information; Literal

Hanna's adjusted whole-group discussion score is relatively low because her discussions do not provide students opportunities to participate in all of the various mathematical practices. However, her discussions are clear and mostly error-free (Hanna's MQI score is 2.17). Thus, it is

not surprising that she has a high value-added estimate of 95. As a result, her measures do not converge: she is accountable to facts and procedures (high VAM) but not authentic discourse (low WGD_Adj).

Analysis of Hanna's discussion presents a couple of peculiar artifacts of the present iteration of the coding scheme presented herein: A teacher who refrains from conducting whole-group discussions (e.g., Paloma, who is discussed below) can receive a higher adjusted whole-group discussion score than her peers who conduct discussions mainly using *Literal* questions. Unless it can be theorized that no discussion is better than bad discussion, this artifact of the coding scheme seems to be a limitation.

Furthermore, while Hanna's whole-group discussions may lack opportunities to participate in various mathematical practices, her discussions are not harmful to students. Therefore, it can be argued, she ought to score higher than her peers who conduct potentially harmful whole-group discussions that contain major mathematical errors (e.g., Helene and Alberto, who are discussed below). These "limitations" of the coding scheme presented herein will be further discussed in Chapter 5.

4.2.3 Teachers accountable to authentic discourse but not facts and procedures

Helene was the only teacher found to be accountable to authentic discourse but not facts and procedures. Building on the descriptions already presented herein, it is easiest to describe Helene's discussions as *incorrect ambitious discourse*.

Helene spends more time engaging her students in whole-group discussions and asks more open-ended questions than most of her peers. On average, Helene spends about a quarter (23%) of her mathematics lessons conducting whole-group discussions. During these

discussions, on average, 13% of her questions are open-ended.²⁸ These characteristics contribute to her decent adjusted whole-group discussion score. However, Helene has a very low value-added estimate. Why do her scores not converge?

A close look at Helene's instruction unveils a host of mathematical errors and lack of clarity. The following piece of transcript is illustrative:

Helene	Do you think that the pizzas are going to be shared equally? At the larger table, it's 16 ... and at the smaller table it's 15 slices, in a sense. Are they shared equally?	Literal
Student	No.	
Helene	No.	Repeat
Student	I think there's less people at the smaller table	
Helene	There are less people at the smaller people, but are they shared equally?	Repeat; Literal
Student	Ahh...	
Helene	We now made them, we made, we now made 40 the same number, so there's 40 people at the smaller table, and there's 40 people at the bigger table. Is there going to have an equal amount of pizzas both tables?	Provides Information; Literal
Student	No	
Helene	No. So, Chase, what are you going to write for A? Because I know you haven't done it yet, so what are you going to write for A? ... This is hard, Chase, I want you to think about this. Chase, look up here. The reason why we wrote a new fraction, an equivalent fraction of 16/40ths and 15/40ths, we've changed how many people are going to be at the table, so both tables now have 40 people, but the larger table has 16 pizzas and the smaller table has 15 pizzas. Do you think they're gonna be shared equally? ... Okay, the larger table has 16 to 40 people. And the smaller table has 15 to 40 people.	Repeat; Literal; Provides Information; Literal

²⁸ Pictures of Helene's discussions appear in Appendix D.

In this brief exchange, it is hard to determine if Helene is asking about the pizzas being shared equally among those at the same table, or among everybody. The pizzas could be shared equally at each table, but that does not necessarily mean that those at one table get the same amount of pizza as the students at the second table. Further, it is not clear if the pizzas are the same size. Moreover, in the last turn, it appears that Helene asserts that $\frac{16}{40}$ and $\frac{15}{40}$ are equivalent fractions. In the same turn, she states, “the larger table has 16 to 40 people. And the smaller table has 15 to 40 people.” One can assume she means to say “16 pizzas to 40 people” and “15 pizzas to 40 people.” Although, we are not actually sure if the numbers “15” and “16” refer to pizzas, because in the beginning Helene refers to these numbers as the number of slices. Helene’s instruction is filled with these types of confusing interactions, where the mathematical point is muddled.

Also contributing to the divergence of Helene’s score might be that her mathematics lessons are filled with discipline issues that disrupt the lesson and flow of discussion. This behavior is not detected using the Analyzing Teaching Moves guide. The following turn illustrates:

Helene Okay. Shhh. So let’s move on. We’ve got a couple of things to talk about with Julia. Red is negative; black is positive. Shhh. Katlin, will you read, “the Julia puts chips on the board”, the second paragraph. Ebony, Martine, can we stop, please?

Student You want me to put it on the board?

Helene There’s blacks in these little one. Are we done talking to Katlin? Come back in. Read ... No, I said ‘read’. There should in one in that package.

Most of Helene’s discussions are filled with turns like these, and it is difficult to follow the direction of the lesson.

Analysis of Helene’s whole-group discussions exposes another limitation of the scoring scheme proposed herein—that is, the mathematical accuracy of a teacher’s lesson is not reflected in the whole-group discussion score. Helene has a relatively high adjusted whole-group discussion score because she engages her students in a lot of whole-group discussion and she asks a lot of open-ended questions during her discussions. However, her instruction includes frequent errors and imprecision (Helene’s MQI score is 1.33), and lacks clarity of mathematical content. Thus, her low value-added estimate (5) is not surprising.

Helen’s measures do not converge: she is accountable to authentic discourse but not facts and procedures.

4.2.4 Teachers not accountable to facts and procedures or authentic discourse

There were two teachers who were not accountable to facts and procedures or authentic discourse: Alberto and Paloma. While both Alberto and Paloma made a lot of mathematical errors (they both have an MQI score of 1.33), they differ in how they conduct whole-group discussions.

Alberto engages his students in quite a bit of whole-group discussion. On average, Alberto spends about a quarter (24%) of his mathematics lesson conducting whole-group discussions. However, on average, only 2% of Alberto’s questions are open-ended. Alberto spends a large portion of his moves telling his students a lot—on average, 27% of his moves are *Provides Information*. *Figure 4.7* includes picture of Alberto’s discussions. In the figure, it is easy to see the preponderance of *Provides Information* (represented by the teal squares across the turns) and very little use of open-ended questions (represented by large amounts of white space in the first three rows of each lesson).



Figure 4.7: Pictures of Alberto's discussions

The following piece of transcript is illustrative of Alberto's whole-group discussions. It starts with Alberto suggesting a solution path to a student who is having difficulty:

Alberto	I would go 20 over 48. But still, it's about 1 over 4.	Provides Information
Student	It's 45.5, right? Is that what you said?	
Alberto	It's 47.5.	Provides Information
Student	Oh, I thought it was 45	
Alberto	But you could round them, so it's about 2 over 5... And the Donner, what's that one? Ah... Lucas?	Provides Information; Literal
Student	The Donner is... 1 over 4.	
Alberto	Good. So which one is the steepest?	Literal
Student	Wait. What'd we get for the Alpine?	
Alberto	It has a rise of 1 and a run of 4.	Provides Information
Student	4?	
Alberto	So which one is the steepest? It'll have the sharpest rise over the shortest run. The sharpest rise over the shortest run.	Literal; Provides Information

This example is representative of how Alberto gives “answers” to student questions and student confusions instead of asking questions that might help them discover the “answer” on their own. It is clear that Alberto is the source of mathematical knowledge: the students look to him to validate their responses and settle disputes. This is problematic, however, because Alberto's instruction is filled with imprecision and errors (Alberto's MQI score is 1.33). As a result, it is not surprising that Alberto's has a low value-added estimate of 36. Alberto's measures converge: he is not accountable to facts and procedures and he is not accountable to authentic discourse.

Unlike Alberto, Paloma spends very little time conducting whole-group discussions. On average, Paloma spends 3% of her mathematics lesson conducting whole-group discussion.²⁹

Instead of whole-group discussion, the students in Paloma's class spend a lot of time doing seatwork. On average, her students are doing seatwork during 44% of the mathematics lesson (on average, 28% of the time engaged in individual seatwork, 16% of the time engaged in group seatwork). This seatwork is generally on the heels of Paloma engaging the students in guided practice. Thus, her mathematics lessons resemble the traditional American cultural script: the teacher models a procedure; the students practice the procedure; class proceeds with very little discussion during the lesson.

Moreover, Paloma's instruction is filled with significant mathematical errors (her MQI score is 1.33). Hence, it is not surprising that her value-added estimate is low (46). Paloma's measures converge: she is not accountable to facts and procedures and she is not accountable to authentic discourse.

4.2.5 Summary: Research question 2

In this section, we took a close look at teachers' whole-group discussion practices. Findings illustrate that teachers with similar value-added estimates can have different teaching practices and that high value-added estimates do not necessarily reflect a full range of classroom teaching practices. This section also uncovered some possible limitations of the coding scheme presented herein. These limitations will be further discussed in Chapter 5.

²⁹ Pictures of Paloma's discussions appear in Appendix D.

With the exception of two teachers (Hanna and Helene), the subjects in this sample had adjusted whole-group discussion scores that converge with their VAM scores. That is, teachers who had high VAM scores ended up having higher adjusted whole-group discussion scores when compared to their peers who had low VAM scores.

Why did Hanna's and Helene's scores not converge?

The situative theory of learning suggests that students will learn what they have the opportunity to learn. Thus, for example, to do well on assessments that measure learning through a set of behavioral, skill-based questions, students only need to have the opportunity to practice the skills related to those questions.

Traditional standardized assessments include a set of behavioral, skill-based questions, and not questions related to the other more cognitively demanding mathematical practices. Hence, students do not necessarily need the opportunity to participate in various mathematical practices to perform well on such assessments.

Hanna provides us with an example. Hanna's adjusted whole-group discussion score does not converge with her high VAM estimate. Although her whole-group discussions are mathematically accurate, they are filled with a preponderance of low-level *Literal* questions. Hanna very rarely begins her whole-group discussions with an open-ended question that invites student thinking. Furthermore, Hanna does not provide her students with many opportunities to explain their thinking, to build on the thinking of others, or to challenge solutions that have been put forward. If these opportunities are indeed valued, then it is important to highlight that teachers do not have to provide them in order to get high value-added estimates (as is the case with Hanna).

Helene, on the other hand, gave her students the opportunity to participate in many whole-group discussions that included opportunities to explain their thinking and build off the thinking of others. But why did her adjusted whole-group discussion score not converge with her VAM estimate?

While Helene's whole-group discussions may appear "rich" on the surface, a close analysis reveals they are filled with errors. Hence, her students do not get the opportunity to practice many basic-procedures correctly. Since traditional standardized assessments are full of questions related to basic procedures, it is no surprise that Helene's students perform poorly on such exams (resulting in a low value-added estimate for Helene).

5.0 DISCUSSION

Learning has been traditionally viewed as the process of accumulating and strengthening skills (i.e., the behaviorist perspective). Thus, teaching consisted of direct instruction (i.e., the teacher transmits knowledge, the student absorbs knowledge). This view of learning and instruction, however, is limited: It says nothing about how one's surrounding environment affects learning. In other perspectives, such as the situative, learning and understanding are viewed as inherently social. Instead of viewing learning as an increased accumulation of skills, social constructivist perspectives view learning as the increased ability to use skills to participate in the practices of a particular community. In this view, teaching takes the role of a good apprenticeship; teachers guide their students through content and provide opportunities to engage in various mathematical practices. "Effective" teaching then consists of more than just the direct teaching of basic skills and procedures. As such, measures of "effective" teaching should assess more than just basic facts and procedures.

Current measures of teacher effectiveness, such as value-added estimates, that rely solely on student performance of basic facts and procedures are thus incomplete: they do not tell us about actual classroom teaching. If we want teachers to improve their practice, the conversation needs to focus more on what effective teachers do (that distinguish them from ineffective teachers) and less on what they produce on standardized assessments. If we have the wrong

metrics, we will strive for the wrong things. The field of education needs to develop measures that will shed light on specific teaching practices. This study attempted to do just that.

This study is one of the first to go inside the classrooms of teachers being labeled effective through value-added modeling. It uncovers how these teachers are interacting with their students during whole-group discussions of a mathematical task. Results suggest that teachers have different lesson structures. For example, one teacher spends, on average, 34% of his lessons conducting whole-group discussions, while another teacher spends, on average, only 3% of her lessons conducting whole-group discussions.³⁰

Interestingly, the study finds great variation even among teachers with similar value-added estimates. For example, one teacher with a value-added estimate of 87 spends on average 15% of his lesson conducting whole-group discussion. And during his discussions, on average, 6% of his questions are open-ended. Another teacher with a value-added estimate of 86 spends on average 34% of his lessons conducting whole-group discussions. And during his discussions, on average, 16% of his questions are open-ended. This variation illustrates how limited VAM estimates are in isolation—that is, VAM estimates tell us nothing about teaching practices.

This study also proposed a theoretically constructed measurement of whole-group discussion: the adjusted whole-group discussion score. This score provides a unique measure of how discourse moves are used in the surrounding context of discussion (moving the field beyond a singular focus on the “type” of question). The adjusted whole-group discussion score helps us better understand if a teacher is conducting whole-group discussions in a way that aligns with the authentic discourse of the underlying community.

³⁰ By focusing on the teachers in Hill et al.’s study (2011) that were at the extremes, this study may have missed important whole-group discussion patterns of “average” teachers (as determined by value-added estimates).

The adjusted whole-group discussion score converged with teachers' value-added estimates; insofar that, with a few exceptions, teachers with high value-added estimates had higher adjusted whole-group discussion scores when compared to their peers who had low value-added estimates.

In one of the cases where the measures did not converge, we found a teacher that was accountable to facts and procedures (i.e., she had a high value-added estimate), but was not accountable to authentic discourse (i.e., she had a relatively low whole-group discussion score). This particular subject provides an example of a teacher who has been labeled “effective,” even though she does not provide her students opportunities to participate in various mathematical practices.

If we want students to increase their ability to participate in the practices of mathematics, they must be provided opportunities to do so. The results discussed herein suggest that teachers vary in both the amount and types of opportunities to participate they provide their students during whole-group discussion of a mathematical task.

5.1 POLICY IMPLICATIONS

Although the teachers in this study who were labeled “effective” had similar VAM scores, their actual classroom practice of conducting whole-group discussions varied. A couple examples include, (1) the average percent of mathematics lesson dedicated to whole-group discussion varied: Vince spends, on average, 34% of his mathematics lessons conducting whole-group discussions, while Josephine only spends, on average, 14% of her lessons engaged in whole-group discussions. (2) The percent of moves that were open-ended questions varied: On average,

17% of Josephine's moves were open-ended questions, while, on average, 6% of Hanna's moves were open-ended questions. With all of this variance, a value-added estimate alone is not very informative.

The main policy implication of these findings is that teachers need targeted intervention. With all of the talk in the field of education about the need to *individualize instruction* for our students, it is a bit surprising this same maxim is not used when discussing what our teachers need—*individualized intervention*.

For example, Andrea might benefit from an intervention that highlighted the importance of pressing students to explain their thinking. While Dolly, for instance, might benefit from an intervention that focused on how to get students accountable to the learning community. And, of course, a few of the teachers would benefit from an intervention that was designed to improve mathematical content knowledge.

Although the present study only investigated whole-group discussions, it is hypothesized the same type of *individualized intervention* would be needed for other teaching practices.

A second policy implication of this study is the important finding that a high VAM estimate is no guarantee of ambitious instruction. As we saw with Hanna, a teacher can be accountable to facts and procedures without being accountable to authentic discourse. This finding is important to highlight in the new Common Core State Standards era. If being mathematically literate includes the standards for mathematical practice as outlined in the Common Core State Standards (e.g., construct viable arguments and critique the reasoning of others), then it will take a lot more than traditional standardized assessments to determine if students are mathematically literate.

5.2 LIMITATIONS

This study is not without its limitations. This section will split the limitations into two categories: (1) limitations of the proposed measure of whole-group discussions, and (2) limitations of the actual study.

5.2.1 Limitations of the adjusted whole-group discussion score

Viewing knowledge as socially constructed has important implications for mathematics education; it means that learning involves being initiated into mathematical ways of knowing (Driver, Asoko, Leach, Mortimer, & Scott, 1994). The teacher plays a critical role in this type of learning environment by enabling or constraining students' opportunities to participate in various mathematical practices. Currently our field lacks transparent measures of such classroom teaching. Developing the adjusted whole-group discussion score is an attempt to begin filling this hole.

This study exposed some limitations to the scoring scheme's present iteration. First, the Analyzing Teaching Moves guide (which serves as the basis of the scoring scheme) does not account for teacher errors. Hence, as we saw with Helene, a teacher whose instruction is filled with mathematical errors can receive a whole-group discussion score that is higher than the whole-group discussion score of a teacher who has mathematically correct discourse. Thus, in its

current iteration, the Analyzing Teaching Moves guide cannot be used to determine if a teacher's teaching is accountable to facts and procedures.³¹

Second, the current scoring scheme does not give credit to teachers who have set up an environment where students are accountable to the learning community. As students participate in repeated social events, they gradually assume more control and responsibility over the activity. The coding scheme mentioned herein is silent on what the students are saying. One could imagine a very effective teacher reaching the point where she has to say very little during a classroom discussion (as we saw with Arthur). The students press one another and ask for clarification. The coding scheme used herein does not capture these behaviors.

Third, teachers who do not conduct whole-group discussions can receive higher whole-group discussion scores than those who do conduct whole-group discussions. For example, Paloma's adjusted whole-group discussion score was 1 and Hanna's adjusted whole-group discussion score was -8. Yet, on average, Paloma only spent 3% of her lessons conducting whole-group discussions whereas Hanna spent, on average, 15% of her lessons conducting whole-group discussions. While quality and quantity should not be confused, children at least need the opportunity to engage in whole-group discussion.

³¹ As previously stated, the authors of the Analyzing Teaching Moves guide acknowledged this limitation while crafting the coding scheme. To overcome this limitation, the authors suggest that an expert in the field of mathematics do a second coding of teachers' discussions to determine the mathematical accuracy of instruction. In this study, teacher's MQI scores served as the second, "expert" coding of instruction.

5.2.2 Limitations of the study

There is much more to discourse than just words. For example, gestures can “say” a lot during a discussion (Rogoff, 1990; Shein, 2012). By just focusing on what teachers say, the present analysis might have missed important non-verbal cues.

Furthermore, this study did not track which students were speaking. Students who are not speaking can still be learning as legitimate peripheral participants (Lave & Wenger, 1991); however, over time, we would hope all students have opportunities to speak. Other than distinguishing the teacher from the student, this study did not keep track of who was talking.

Moreover, there are “pronounced changes” in student use of language and logic when they are provided with sufficient wait time—the amount of time between two turns (Rowe, 1986; Tobin, 1986). This includes wait time after a teacher question as well as wait time after a student response. This study did not capture this important dimension.

Finally, the actual interactions between students and their teacher are nested within the task at hand. Not all tasks are created equal—different tasks have different cognitive demands (Stein & Lane, 1996; Stein & Smith, 1998). It can be argued that the task at hand moderates the types of questions that teachers can ask. For example, Lee and Kinzie (2012) found teachers asked twice as many open-ended questions during “experiments” when compared to “book reading activities” (57% versus 29% respectively). Sloan & Pate (1966) reported teachers implementing the “new math” curricula asked significantly fewer recall questions and significantly more comprehension and analysis questions when compared to teachers who taught the traditional mathematics curricula.

A limitation, therefore, of this study is that it did not categorize the types of tasks that students were working on. A relatively low whole-group discussion score could be the result of

poor task selection, and not bad discussion practices per se. Although not formally investigated, I do not believe this to be the case in the present study.

5.3 NEXT STEPS

Productive classroom discussions in mathematics are widely agreed to be a major component of reform-oriented mathematics (e.g., Chapin, O'Connor, & Anderson, 2003; Dantonio & Beisenherz, 2001; Kazemi & Stipek, 2001; Lampert, 2001; Stein, Engle, Smith, & Hughes; 2008). Yet measures of this practice are underdeveloped. This study attempted to test a transparent, debatable measure of classroom discussion. One next step to this study is to iterate the scoring scheme presented herein to address some of the limitations that surfaced.

Another “next step” is to develop a measure of learning that is in line with the situative perspective of learning. As students negotiate and renegotiate the meaning of mathematics, they will naturally change the way they participate in it. This change in participation, from a situative perspective, is what signifies learning. Thus, the field of education needs to develop a metric that measures students’ participation in various mathematical practices (and a change in participation over time). The present study simply measured opportunities to participate in various mathematical practices. It did not investigate how those opportunities are related to learning, because a robust measure of learning as defined by the situative perspective is not available. (It is the belief of this author that measuring learning through standardized testing is limited, as shown in the study discussed herein).

A third step is developing a way to support teachers as they create environments that are conducive to ambitious classroom discourse. There are ground rules for discussion, and these

ground rules vary depending on where (and with whom) the discussion is taking place. For instance, each family may have their own “ground rules” for discussion around the dinner table. These rules may be considerably different to the ground rules during the family’s church service, for example. It does not take long for individuals to learn the ground rules and adapt to different situations. Classroom talk is no different: there are ground rules. Traditionally, the rules include the teacher controlling the conversation. Students are to talk only when called upon. At first, conducting ambitious classroom discussions may be difficult, not because students cannot engage in such talk, but rather, the ground rules are different (see Mercer & Hodgkinson, 2008 and Yackel, 2001 for multiple ground rule setting approaches).

Finally, while designing the Analyzing Teaching Moves guide, the authors—Kevin Ashley, Richard Correnti, Moddy McKeown, Peg Smith, Mary Kay Stein, James Chisholm, and Jimmy Scherrer—envisioned the coding scheme being used in an intelligent tutoring environment for teachers.³² In this environment, a group of teachers will code each other’s classroom discussions, which are stored in an online database. A picture of discourse—similar to the examples provided in Section 4.2—will be automatically generated based on the codes. This picture serves as a focal point of discussion for the teachers.³³ Using a computer mouse, or touchscreen on a tablet, teachers can hover over certain sections of the picture of discourse and the transcript from that specific moment of the lesson will appear in a box beneath the picture of discourse. Teachers can select aspects of their classroom discourse that they would like to work

³² The Analyzing Teaching Moves guide was developed at the University of Pittsburgh, Learning Research and Development Center.

³³ Similar to the pictures presented herein, the pictures generated in the Intelligent Tutoring environment would not be able to distinguish mathematically correct instruction from mathematically incorrect instruction. The authors of the Analyzing Teaching Moves guide are currently wrestling with this limitation.

on and track changes in their pictures of discourse over time. A fourth next step is to continue developing this unique piece of the original project.

5.4 CONCLUSION

The field of education is currently in a precarious situation: the use of new measures, such as value-added estimates, outpaces the understanding of how to correctly process the measures. Value-added estimates are useful, but extremely limited. The field of education, however, is using value-added estimates without correctly processing their limitations. This is leading an entire field to shift their time and energy (to increasing standardized test scores) based on the faulty assumption that value-added estimates are a good proxy for robust teaching.

To bring the measurement of teaching into clear contact with the situative perspective on learning, I proposed one theoretically constructed score that quantifies various types of teacher-student interactions during whole-group discussion. The hope is that this metric will be one of many that are developed to measure actual teaching practices through a situative lens.

As with any theoretically constructed metric or framework, the utility of the adjusted whole-group discussion score will be tested by “whether it stimulates sufficient work to lead to its revision, abandonment, or enrichment through an increasingly well-targeted set of research results from the learning and educational sciences” (Koedinger, Corbett, & Perfetti, 2012, p.791).

APPENDIX A

COVARIATE ADJUSTED MODELS USED TO CALCULATE TEACHER VALUE- ADDED ESTIMATES (HILL ET AL., 2011, PP.806-807)

$$y_{ijkl} = \beta_0 + \beta \mathbf{x}_{ijkl} + \tau_{kl} + \xi_{jkl} + \varepsilon_{ijkl},$$

where y_{ijkl} is SMA scale score for the i th student in the j th classroom of the k th teacher at the l th school in 2007–2008 and \mathbf{x}_{ijkl} is a vector consisting of the $ijkl$'s student's previous-year SMA scale score (SMA06), SMA06 squared, SMA06 cubed, grade indicator variables, previous year's standardized reading scale score (SRA06), previous year's standardized science scale score (SSA06) and SMA06, SRA06, and SSA06 by grade indicators interaction terms. β is a parameter vector. The $\tau_{kl} \sim N(0, \sigma_\tau^2)$ are teacher random effects, and ξ_{jkl} and ε_{ijkl} are classroom- and student-level error terms assumed to be independent where,

$$\begin{aligned} \varepsilon_{ijkl} &\sim N(0, \sigma_\varepsilon^2) \\ \xi_{jkl} &\sim N(0, \sigma_\xi^2) \end{aligned}$$

Model 2: School fixed effects model. Adjusting for prior student scores, teacher random effect (as with the simple model), plus a school fixed effect, the school fixed effects model is,

$$y_{ijkl} = \beta_0 + \beta \mathbf{x}_{ijkl} + \tau_{kl} + \phi \mathbf{c}_l + \xi_{jkl} + \varepsilon_{ijkl},$$

where \mathbf{c}_l is a vector of indicator variables for each school and the ϕ are the parameters indicating the school fixed effects.

Model 3: Student background adjusted model. Adjusting for prior student scores and a teacher random effect (as with the simple model) plus student background variables, the student background adjusted model is,

$$y_{ijkl} = \beta_0 + \beta \mathbf{x}_{ijkl} + \tau_{kl} + \lambda \mathbf{s}_{ijkl} + \xi_{jkl} + \varepsilon_{ijkl},$$

where \mathbf{s}_{ijkl} is a vector of student covariates including indicator variables for accelerated or enriched, algebra, FRL, English language learner (ELL), special education (SPED), Spanish test language, indicator variables for each ethnicity and a SPED by SMA06 interaction, and λ is a parameter vector.

APPENDIX B

LESSON SEGMENTS

Whole-Class Discussion

The teacher and students are engaged in a whole-class discussion about a mathematical task that the students have had an opportunity (either individually and/or in groups) to work on. This task/problem might have been done in class or could have been completed for homework.

Individual Seatwork

Students are at their seats working on a problem, and the teacher did not specifically ask the students to work in groups. Although students may actually be talking with other classmates during individual seatwork, it is usually not about the mathematical task they are working on. Students often do individual seatwork when completing a warm-up, working on problems in a text, or getting a start on the night's homework.

Group Seatwork

The teacher has asked the students to work on a mathematical task/problem in groups (i.e., two or more).

Review of Work

Teacher is giving answers (e.g., telling, or displaying on an overhead) to an assignment. *Reviewing work* could also include a teacher asking the students (often in a round-robin manner) to give the answers, but does not include a discussion. That is, any dialogue between the teacher and student typically does not go beyond “the answer.”

Guided Practice

The teacher is solving tasks/problems with the class that the students have not yet had the opportunity to solve either individually or with others. Instead, they “go through” the problem(s) while the teacher is guiding them. Guided practice often involves a lot of modeling and will frequently involve a discussion. However, it differs from *whole-class discussion* in that the students did not receive an opportunity (either independently or with others) to solve the task/problem without the teacher’s guidance.

Direct Instruction

The teacher is telling/giving information to the students. *Direct instruction* resembles a traditional “lecture.” At times, *direct instruction* may appear like *guided practice*, but lack the student involvement that is typical of *guided practice*.

Transition/Set-Up

Transition/Set-Up are times in the lesson where the class is not engaged in a mathematical task/problem and is not being lectured (i.e. *Direct Instruction*). During these segments, the

teacher is often taking roll, giving directions/instructions, accessing prior knowledge, reading the intro paragraphs in a text, and distributing materials. *Transition/Set-Up* differ from *guided practice* in that the class is not actually solving mathematical problems.

APPENDIX C

BREAKDOWNS OF TEACHERS' WHOLE-GROUP DISCUSSIONS

Relative frequency of moves during Alberto's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
ALBERTO	1	17	3	0	35	0	9	0	0	3	29	0	0	21
	2	31	0	0	32	0	16	0	0	11	14	0	0	27
	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	42	1	0	42	1	1	0	0	8	32	0	0	14
	5	54	4	0	49	0	0	0	0	4	32	0	0	12
	6	0	0	0	0	0	0	0	0	0	0	0	0	0
	<i>Mean</i>		24	2	0	40	<1	7	0	0	7	27	0	0

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Andrea's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
ANDREA	1	17	5	0	40	2	2	2	0	15	10	0	0	26
	2	16	0	0	43	14	0	0	0	11	8	0	0	24
	3	37	8	1	43	14	1	0	0	11	13	0	1	6
	4	13	0	0	53	8	0	0	0	18	13	0	3	5
	5	25	3	1	44	12	0	0	0	16	15	0	0	9
	6	25	1	0	56	0	0	0	0	14	14	0	1	13
	<i>Mean</i>		22	3	<1	47	8	<1	<1	0	14	12	0	<1

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Arthur's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
ARTHUR	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	46	3	1	38	7	1	0	1	14	23	0	0	11
	4	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	34	0	0	45	9	0	1	0	14	18	0	0	13
	6	5	0	0	43	0	0	7	0	14	14	0	0	21
	<i>Mean</i>	14	1	<1	42	5	<1	3	<1	14	18	0	0	15

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Dolly's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
DOLLY	1	47	5	1	44	5	1	0	0	23	10	0	0	10
	2	23	0	0	38	17	0	0	0	23	13	0	0	9
	3	37	7	1	42	10	3	0	0	24	10	0	0	2
	4	28	3	0	46	5	3	3	0	23	11	0	0	5
	5	25	1	2	48	8	1	1	0	25	10	0	1	2
	6	30	2	1	35	2	1	1	0	20	20	0	0	17
	<i>Mean</i>	32	3	<1	42	8	2	<1	0	23	12	0	<1	8

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Gabrielle's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
GABRIELLE	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	5	0	9	45	18	0	0	0	9	18	0	0	0
	4	36	1	1	38	5	5	2	0	11	16	0	1	20
	5	20	5	2	52	2	4	0	0	9	16	2	4	5
	6	34	1	0	39	11	3	0	1	12	24	0	0	10
	<i>Mean</i>	16	2	3	44	9	3	<1	<1	10	19	<1	1	9

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Hanna's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
HANNA	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	12	2	0	63	0	0	4	0	15	15	0	0	2
	3	18	0	0	33	7	22	0	0	24	4	0	0	9
	4	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	22	2	0	58	7	0	0	0	15	20	0	0	2
	6	37	<1	<1	56	5	2	0	0	17	13	0	0	6
	<i>Mean</i>	15	1	<1	53	5	6	1	0	18	13	0	0	5

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Helen's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
HELEN	1	16	4	0	48	7	2	0	0	11	24	0	2	2
	2	35	4	1	33	18	8	0	0	10	15	0	1	10
	3	12	0	3	53	9	3	0	0	15	3	0	0	15
	4	36	0	2	49	8	0	0	0	17	19	0	4	2
	5	6	16	26	32	0	0	0	0	5	0	0	0	21
	6	35	0	0	37	11	4	0	1	13	11	0	3	21
	<i>Mean</i>	23	4	5	42	9	3	0	<1	12	12	0	2	12

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Josephine's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
JOSEPHINE	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	42	3	3	38	18	1	0	0	9	19	0	1	8
	3	15	7	0	52	3	3	0	0	31	3	0	0	0
	4	15	5	5	38	15	3	0	0	10	18	0	0	5
	5	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	10	7	0	36	7	9	0	0	14	20	0	0	7
	<i>Mean</i>	14	6	2	41	11	4	0	0	16	15	0	<1	5

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

Relative frequency of moves during Paloma's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC	
PALOMA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
	2	13	3	0	63	9	0	0	0	0	9	0	0	16	
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	
	5	4	8	0	38	8	0	0	0	8	8	0	0	31	
	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	<i>Mean</i>	3	6	0	51	9	0	0	0	4	9	0	0	24	

Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

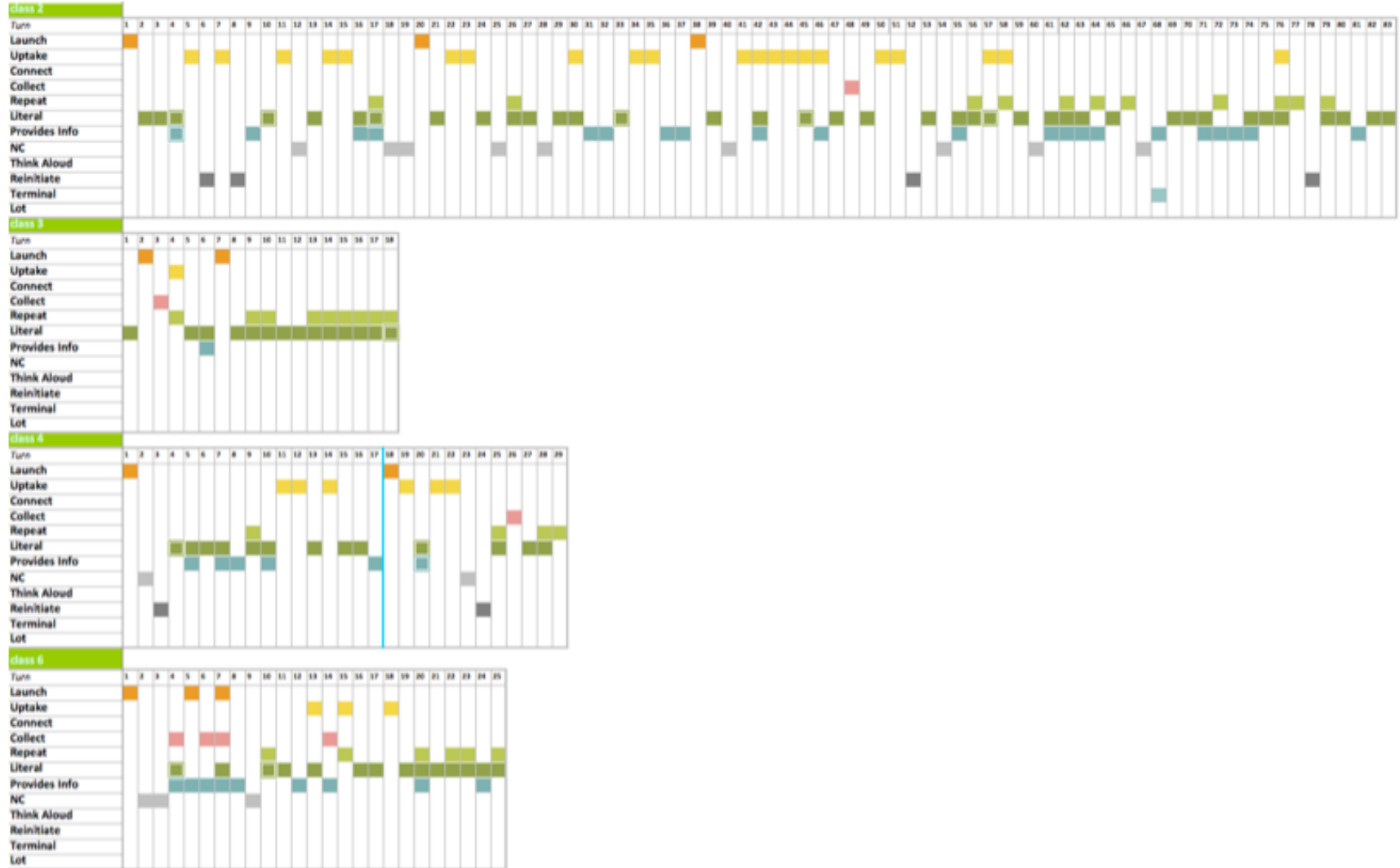
Relative frequency of moves during Vince's whole-group discussions (%)

	Lesson	% WGD	LA	RI	LI	UP	CL	CN	LO	RP	PI	TH	TR	NC
VINCE	1	69	5	0	37	13	0	2	0	13	26	0	0	5
	2	36	4	2	28	15	0	4	0	21	18	0	0	8
	3	8	7	0	43	7	0	0	0	16	20	0	0	7
	4	52	2	0	46	8	1	0	0	10	23	2	2	2
	5	11	3	0	31	13	3	0	0	13	31	0	0	6
	6	27	6	2	26	11	4	4	0	13	21	0	0	13
	<i>Mean</i>	34	5	<1	35	11	1	2	0	14	23	<1	<1	7

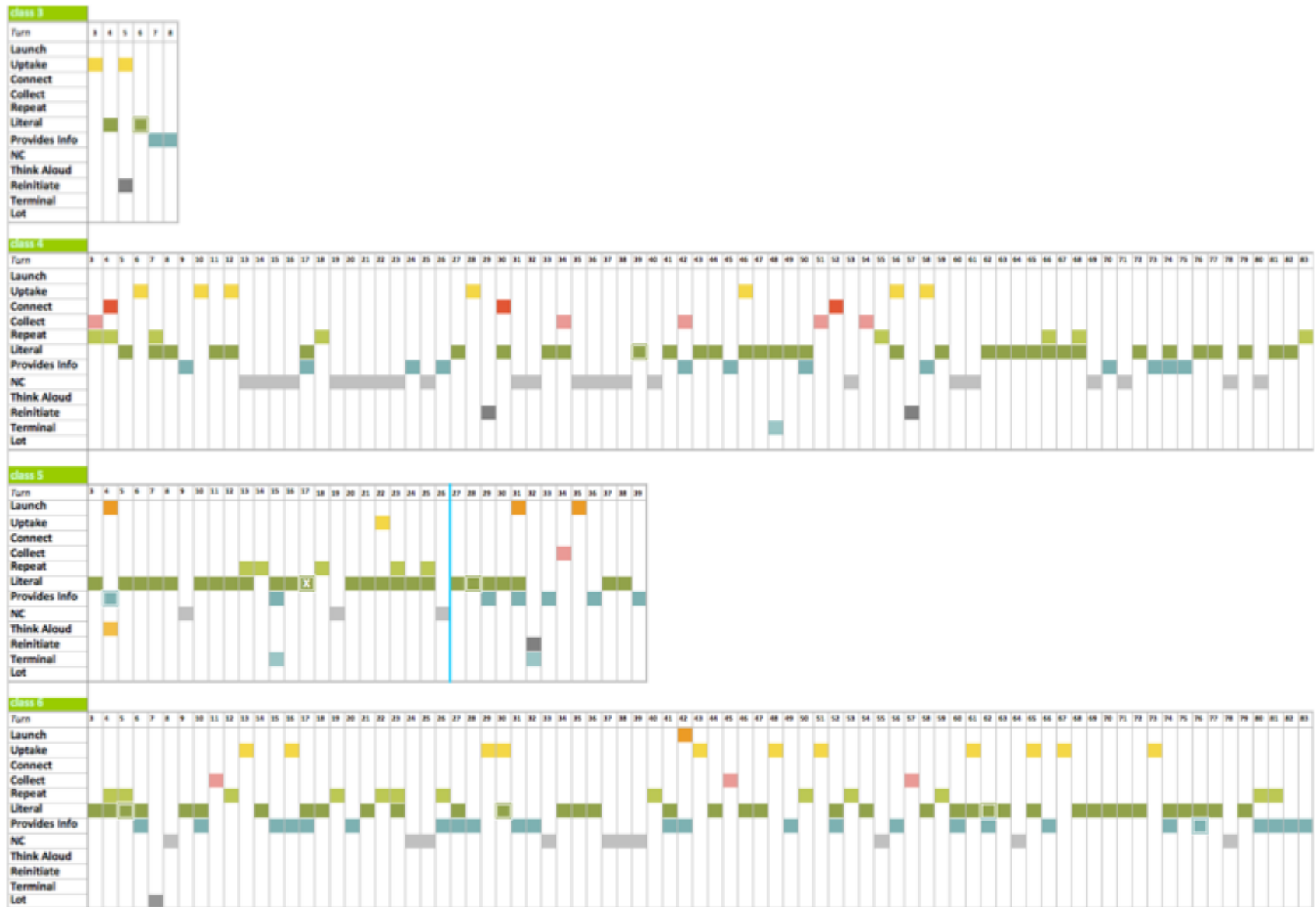
Note: % WGD = percent of turns that was whole-group discussion; LA = Launch; RI = Re-initiate; LI = Literal; UP = Uptake; CL = Collect; CN = Connect; LO = Lot; RP = Repeat; PI = Provides Information; TH = Think Aloud; TR = Terminate; NC = No Code

APPENDIX D

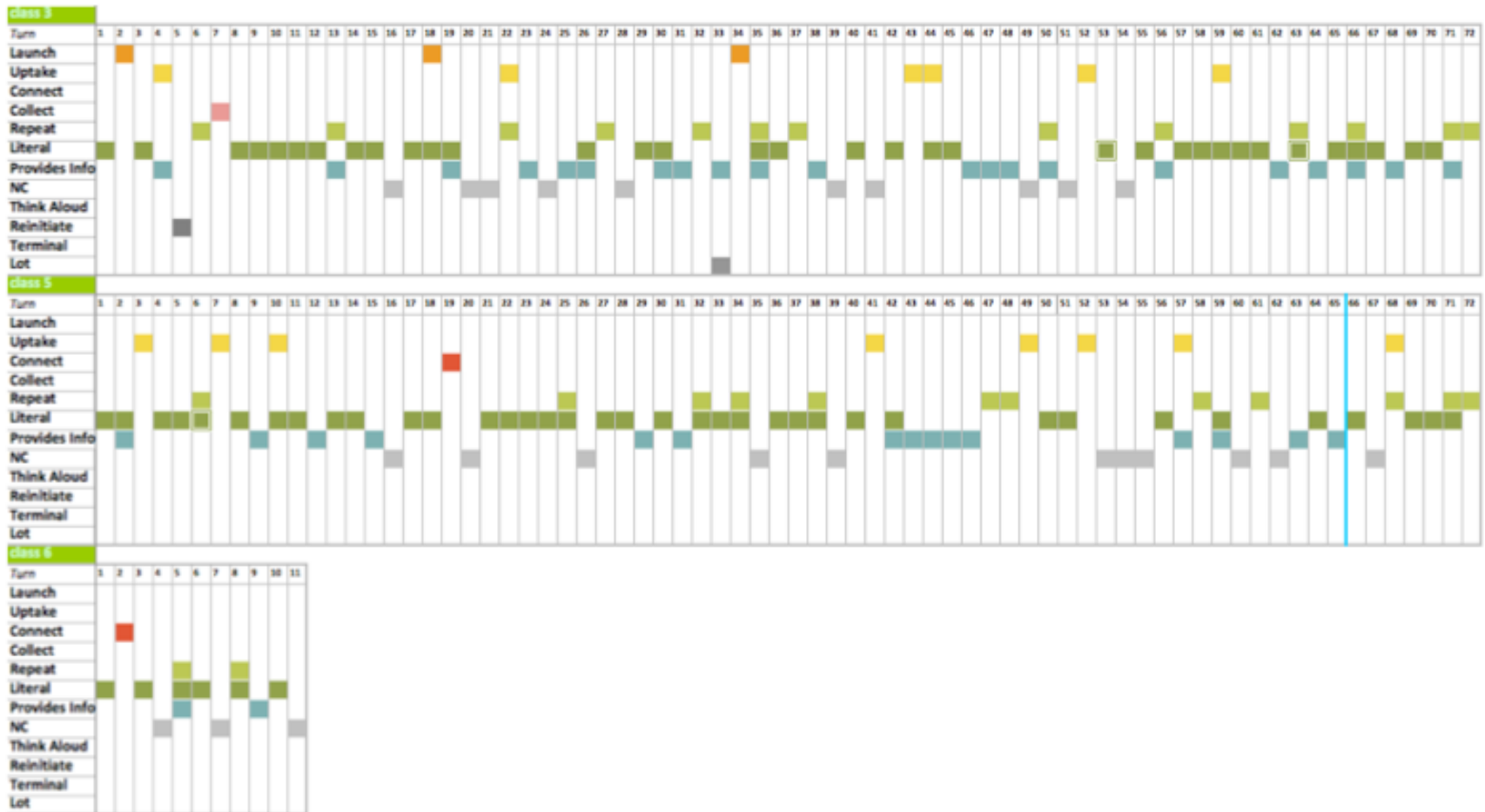
PICTURES OF TEACHERS' DISCUSSIONS



Pictures of Josephine's Discussion



Pictures of Gabrielle's Discussions: Note: For space considerations, the picture of Gabrielle's class 4 is missing the last few turns



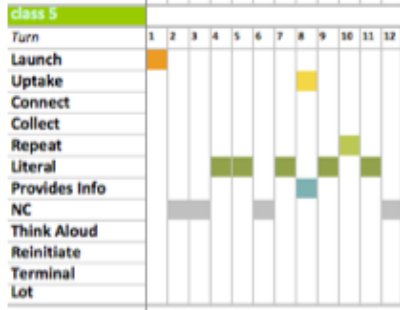
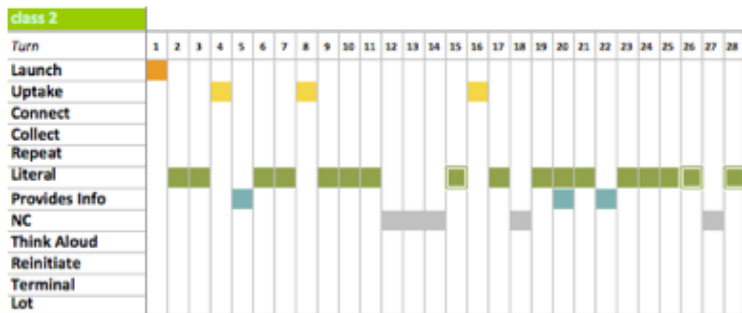
Pictures of Arthur's Discussions



Pictures of Helene's Discussions



Pictures of Helene's Discussions (continued)



Pictures of Paloma's Discussions

BIBLIOGRAPHY

- Abd-Kadir, J., & Hardman, F. (2007). The discourse of whole class teaching: A comparative study of Kenyan and Nigerian primary English lessons. *Language and Education*, 21(1), 1-15.
- Adams, S. J., Heywood, J. S., & Rothstein, R. (2009). *Teachers, performance pay, and accountability: What education should learn from other sectors*. Washington, D.C.: Economic Policy Institute.
- Adey, P., Shayer, M., & Yates, C. (2001). *Thinking science: the materials of the CASE project* (Third ed.). Nelson Thornes.
- Aleksandrov, A. D., Kolmogórov, A. N., & Lavrent'ev, M. A. (Eds.). (1999). *Mathematics: its content, methods and meaning*. New York, NY: Courier Dover Publications.
- Alexander, R. J. (2000). *Culture and pedagogy: International comparisons in primary education*. Wiley-Blackwell.
- Alexander, R. (2003). *Talk for Learning: the first year*. Northallerton: North Yorkshire Council.
- Alexander, R. (2004). *Towards dialogic teaching: rethinking classroom talk*. Cambridge: Dialogos.
- Askew, M., Brown, M., Rhodes, V., Johnson, D., & Wiliam, D. (1997). *Effective teachers of numeracy: final report*. London: King's College, University of London.
- Barnes, D. R. (1976). *From communication to curriculum*. Harmondsworth: Penguin Education.
- Barnes, D. R., & Todd, F. (1995). *Communication and learning revisited: Making meaning through talk*. Portsmouth, NH: Heinemann.
- Baumann, J. F. (1988). Direct instruction reconsidered. *Journal of Reading*, 31(8), 712-718.
- Bereiter, C., & Scardamalia, M. (1989). Intentional learning as a goal of instruction. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 361-392). Hillsdale, NJ: Erlbaum.

- Bishop, J. P. (2012). "She's Always Been the Smart One. I've Always Been the Dumb One": Identities in the Mathematics Classroom. *Journal for Research in Mathematics Education*, 43(1), 34-74.
- Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*. Mahwah, NJ: L. Erlbaum.
- Boaler, J. & Brodie, K. (2004). The importance, nature and impact of teacher questions. In D. McDougall & J. Ross (Eds.), *Proceedings of the 26th annual meeting of the North American chapter of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 773-781). Toronto: Ontario Institute of Studies in Education/University of Toronto.
- Britton, B. K., Van Dusen, L., Glynn, S. M., & Hemphill, D. (1990). The impact of inferences on instructional text. *Psychology of learning and motivation*, 25, 53-70.
- Brodie, K. (2008). Describing teacher change: interactions between teacher moves and learner contributions. In J. Matos, P. Valero, & K. Yasukawa (Eds.), *Fifth International Mathematics Education and Society Conference*. Lisbon:
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, 18(1), 32-42.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Bruner, J. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.
- Bryk, A. S., & Schneider, B. (2003). Trust in schools: A core resource for school reform. *Educational Leadership*, 60(6), 40-45.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, J., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.
- Burkhardt, H. (1988). Teaching problem solving. In H. Burkhardt, S. Groves, A. Schoenfeld, & K. Stacey (Eds.), *Problem solving--A world view (Proceedings of the problem solving theme group, ICME 5)* (pp. 17-42). Nottingham: Shell Centre.
- Chang, K. E., Sung, Y. T., & Chen, S. F. (2001). Learning through computer-based concept mapping with scaffolding aid. *Journal of Computer Assisted Learning*, 17(1), 21-33.
- Chapin, S., O'Connor, C., & Anderson, N. (2003). *Classroom discussions: Using math talk to help students learn*. Sausalito, CA: Math Solutions.
- Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13(1989), 145-182.

- Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. *Journal for Research in Mathematics Education*, 258-277.
- Cochran-Smith, M. & Zeichner, K. (Eds.) (2005). Studying teacher education: the report of the AERA Panel on Research and Teacher Education. Mahwah, NJ: Lawrence Erlbaum Associates.
- Collins, J. (1982). Discourse style, classroom interaction and differential treatment. *Journal of Reading Behavior*.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (p. 487). Hillsdale, NJ: Lawrence Erlbaum.
- Coltman, P., Petyaeva, D., & Anghileri, J. (2002). Scaffolding learning through meaningful tasks and adult interaction. *Early Years*, 22(1), 39-49.
- Confrey, J. (1990). A review of the research on student conceptions in mathematics, science, and programming. In C. Cazden (Ed.), *Review of research in education* (pp. 3-56). Washington, DC: American Educational Research Association.
- Contu, A., & Willmott, H. (2003). Re-embedding situatedness: The importance of power relations in learning theory. *Organization science*, 14(3), 283-296.
- Corey, S. M. (1940). The teachers out-talk the pupils. *The School Review*, 48(10), 745-752.
- D'Andrade, R. G. (1981). The cultural part of cognition. *Cognitive Science*, 5(3), 179-195.
- Daniels, H. (2001). *Vygotsky and pedagogy*. London: Routledge.
- Dantonio, M., & Beisenherz, P. (2001). *Learning to question, questioning to learn: Developing effective teacher questioning practices*. Needham Heights, MA: Allyn & Bacon.
- Dennen, V. P. (2004). Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 813-828). Mahwah, NJ: Lawrence Erlbaum Associates.
- Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational researcher*, 23(7), 5-12.
- Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J. et al. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, 237-252.
- Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. London: Routledge.

- Ellis, S., & Rogoff, B. (1986). Problem solving in children's management of instruction. In E. Mueller & C. Cooper (Eds.), *Process and outcome in peer relationships* (pp. 301-325). Orlando, FL: Academic Press.
- Fisher, R., & Larkin, S. (2008). Pedagogy or ideological struggle? An examination of pupils' and teachers' expectations for talk in the classroom. *Language and Education*, 22(1), 1-16.
- Floyd, W. D. (1960). *An Analysis of the Oral Questioning Activity in Selected Colorado Primary Classrooms: Unpublished doctoral dissertation*. Colorado State University, Fort Collins.
- Franke, M. L., Webb, N. M., & Chan, A. G. (2009). Teacher Questioning to Elicit Students' Mathematical Thinking in Elementary School Classrooms. *Journal of Teacher Education*, 60(4), 380.
- Gadamer, H. G. (1998). *Truth and method* (J. Weinsheimer & D. Marshall, Trans. 2). New York: Continuum.
- Gall, M. D. (1970). The use of questions in teaching. *Review of Educational Research*, 40(5), 707-721.
- Gee, J. P. (2008). A sociocultural perspective on opportunity to learn. In P. Moss, D. Pullen, J. P. Gee, E. Haertel, & L. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 76-108). New York, NY: Cambridge University Press.
- Geertz, C. (1983). *Local Knowledge: Further Essays in Interpretive Anthropology*. New York, NY: Basic Books.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology*, 15(1), 1-38.
- Gitomer, D. H., & Bell, C. A. (2013). Evaluating teachers and teaching. In K. F. Geisinger (Ed.), *APA Handbook of testing and assessment in psychology*. Washington, DC: American Psychological Association.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104.
- Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational researcher*, 26(1), 5-17.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist*, 53(1), 5-26.
- Greeno, J. G. (2011). A situative perspective on cognition and learning in interaction. In T. Koschmann (Ed.), *Theories of Learning and Studies of Instructional Practice* (pp. 41-71). New York, NY: Springer.

- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Macmillan.
- Halliday, M. (1993). Towards a language-based theory of learning. *Linguistics and Education*, 5(2), 93-116.
- Halmos, P. R. (1980). The heart of mathematics. *The American Mathematical Monthly*, 87(7), 519-524.
- Hanushek, E. (2002). Teacher quality. In W. Evers & L. Izumi (eds.) *Teacher quality* (pp.1-12). Palo Alto, CA: Hoover Institution.
- Hanushek, E. (2009). Teacher deselection. In D. Goldhaber & J. Hannaway (Eds.), *Creating a new teaching profession* (pp. 165-180). Washington, D.C.: Urban Institute Press.
- Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. *Handbook of the Economics of Education*, 2, 1051-1078.
- Harris, D. N. (2009). Teacher value-added: Don't end the search before it starts. *Journal of Policy Analysis and Management*, 28(4), 693-699.
- Harré, R., & Van Lagenhove, L. (1999). *Positioning theory: moral contexts of international action*. Malden, MA: Blackwell Publishers.
- Haynes, H. C. (1935). *Relation of teacher intelligence, teacher experience, and type of school to types of questions: Unpublished doctoral dissertation*. George Peabody College for Teachers, Nashville.
- Hendricks, C. C. (2001). Teaching causal reasoning through cognitive apprenticeship: What are results from situated learning? *The Journal of Educational Research*, 94(5), 302-311.
- Hicks, D. (1995). Discourse, learning, and teaching. *Review of Research in Education*, 21, 49-95.
- Hiebert, J., & Morris, A. K. (2012). Teaching, rather than teachers, as a path toward improving classroom instruction. *Journal of Teacher Education*, 63(2), 92-102.
- Hiebert, J., & Wearne, D. (1993). Instructional tasks, classroom discourse, and students' learning in second-grade arithmetic. *American Educational Research Journal*, 30(2), 393-425.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., 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.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.

- Hogan, K., Nastasi, B. K., & Pressley, M. (2000). Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions. *Cognition and instruction*, 17(4), 379-432.
- Howe, C., Tolmie, A., Thurston, A., Topping, K., Christie, D., Livingston, K. et al. (2007). Group work in elementary science: Towards organisational principles for supporting pupil learning. *Learning and Instruction*, 17(5), 549-563.
- Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26(1), 101-136.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379-424.
- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, 38(6), 523-550.
- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45-83.
- Kawanaka, T., & Stigler, J. W. (1999). Teachers' use of questions in eighth-grade mathematics classrooms in Germany, Japan, and the United States. *Mathematical Thinking and Learning*, 1(4), 255-278.
- Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal*, 102(1), 59-80.
- King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. *Contemporary Educational Psychology*, 14(4), 366-381.
- King, A. (1994). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American educational research journal*, 31(2), 338-368.
- King, A. (1999). Discourse patterns for mediating peer learning. In A. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 87-115). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kitcher, P. (1984). *The nature of mathematical knowledge*. New York, NY: Oxford University Press.
- Klaczynski, P. A. (1997). Bias in Adolescents' Everyday Reasoning and Its Relationship With Intellectual Ability, Personal Theories, and Self-Serving Motivation. *Developmental Psychology*, 33(2), 273-283.
- Knight, S. L., Edmondson, J., Lloyd, G. M., Arbaugh, F., Nolan, J., Whitney, A. et al. (2012). Examining the Complexity of Assessment and Accountability in Teacher Education. *Journal of Teacher Education*, 63(5), 301-303.

- Koedel, C., & Betts, J. R. (2005). Re-examining the role of teacher quality in the educational production function. *University of Missouri Department of Economics Working Paper*, 07-08.
- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757-798.
- Koester, L. S., & Bueche, N. A. (1980). Preschoolers as teachers: Where children are seen but not heard. *Child Study Journal*, 10, 107-118.
- Kumpulainen, K., & Lipponen, L. (2010). Productive interaction as agentic participation in dialogic enquiry. In K. Littleton & C. Howe (Eds.), *Educational dialogues: understanding and promoting productive interaction* (pp. 48-63). New York: Routledge.
- Lam, S., Law, Y., Shum, M. (2009). Classroom discourse analysis and educational outcomes in the era of education reform. *British Journal of Educational Psychology*, 79, 617-641.
- Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven, CT: Yale University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lee, Y., & Kinzie, M. B. (2012). Teacher question and student response with regard to cognition and language use. *Instructional Science*, 40, 857-874.
- Lefstein, A. (2010). More helpful as problem than solution: some implications of situating dialogue in classrooms. In K. Littleton & C. Howe (Eds.), *Educational dialogues: understanding and promoting productive interaction* (pp. 170-191). New York: Routledge.
- Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, 23(1), 87-163.
- Lemke, J. L. (1985). *Using language in the classroom*. Victoria, Australia: Deakin University Press.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V. N., & Martinez, J. F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement*, 44(1), 47-67.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572-606.

- McLane, J. B. (1987). Interaction, context, and the zone of proximal development. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 267-285). San Diego, CA: Academic Press.
- Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Harvard University Press Cambridge, MA.
- Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, 80(1), 1-14.
- Mercer, N., & Hodgkinson, S. (Eds.), (2008). *Exploring classroom talk*. London: Sage.
- Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: a sociocultural approach*. New York: Routledge.
- Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: Accountable Talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27(4), 283-297.
- Mirza, N. M., & Perret-Clermont, A. N. (2009). *Argumentation and education: Theoretical foundations and practices*. New York: Springer.
- Moll, L. (Ed.). (1990). *Vygotsky and education*. Cambridge: Cambridge University Press.
- Moyles, J. R., Hargreaves, L., Merry, R., Paterson, F., & Esarte-Sarries, V. (2003). *Interactive teaching in the primary school: digging deeper into meanings*. Berkshire: Open University Press.
- Mugny, G., Perret-Clermont, A., & Doise, W. (1981). Interpersonal coordinations and sociological differences in the construction of the intellect. In G. Stevenson & G. Davis (Eds.), *Applied social psychology (vol. 1)*. Chichester: Wiley.
- Nathan, M. J., & Knuth, E. J. (2003). A study of whole classroom mathematical discourse and teacher change. *Cognition and Instruction*, 21(2), 175-207.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington, DC: Author.
- Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, 35(2), 135-198.
- O'Connor, C. (2001). "Can any fraction be turned into a decimal?" A case study of a mathematical group discussion. *Educational Studies in Mathematics*, 46(1), 143-185.
- Palincsar, A. S. (1986). The role of dialogue in providing scaffolded instruction. *Educational psychologist*, 21(1-2), 73-98.

- Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. In J. Spence, J. Darley, & D. Foss (Eds.), *Annual Review of Psychology* (pp. 345-375). Palo Alto, CA: Annual Reviews.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction*, 1(2), 117-175.
- Penuel, W., Sun, M., Frank, K., and Gallagher, A. (2012). Using Social Network Analysis to Study How Collegial Interactions Can Augment Teacher Learning from External Professional Development. *American Journal of Education*, 119(1), 103-136.
- Pepper, S. (1942). *World hypotheses: a study in evidence*. Berkeley, CA: University of California Press.
- Peterson, B., & Leatham, K. (2009). Learning to use students' mathematical thinking to orchestrate a class discussion. In L. Knott (Ed.), *The role of mathematics discourse in producing leaders of discourse* (pp. 99-128). Charlotte, NC: Information Age Publishing.
- Peterson, P., & Walberg, H. (Eds.). (1979). *Research on teaching: Concepts, findings and implications*. Berkeley, CA: McCutchan.
- Polya, G. (1945). *How to solve it: a new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
- Polya, G. (1954). *Mathematics and plausible reasoning, Volume 1: Induction and analogy in mathematics*. Princeton, NJ: Princeton University Press.
- Resnick, L. B. (1987). Learning in school and out. *Educational researcher*, 16(9), 13-20+54.
- Resnick, L. (1989). Treating mathematics as an ill-structured discipline. In R. Charles & E. Silver (Eds.), *The teaching and assessing of mathematical problem solving* (pp. 32-60). Reston, VA: National Council of Teachers of Mathematics.
- Resnick, L., Saljo, R., Pontecorvo, C., & Burge, B. (Eds.). (1991). *Discourse, tools, and reasoning: essays on situated cognition*. Berlin, Germany: Springer.
- Resnick, L. B., & Scherrer, J. (2012). Social Networks in "Nested Learning Organizations". *American Journal of Education*, 119(1), 183-192.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York: Oxford University Press.
- Rogoff, B. (1991). Social interaction as apprenticeship in thinking: Guided participation in spatial planning. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 349-364). Washington, DC: American Psychological Association.

- Rogoff, B., & Gardner, W. (1984). Adult guidance of cognitive development. In B. Rogoff & J. Lave (Eds.), *Everyday Cognition: development in social context* (pp. 95-116). Cambridge, MA: Harvard University Press.
- Rojas-Drummond, S., & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research*, 39(1-2), 99-111.
- Rojas-Drummond, S., Mercer, N., & Dabrowski, E. (2001). Collaboration, scaffolding and the promotion of problem solving strategies in Mexican pre-schoolers. *European Journal of Psychology of Education*, 16(2), 179-196.
- Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). Benchmarking and alignment of standards and testing. *CSE Technical Report 566*.
- Rothstein, R. (2004). *Class and Schools: Using Social, Economic, and Educational Reform to Close the Black-White Achievement Gap*. New York: Teachers College Press.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.
- Rowe, M. B. (1986). Wait time: slowing down may be a way of speeding up! *Journal of teacher education*, 37(1), 43-50.
- Sass, T., Hannaway, J., Xu, Z., Figlio, D., & Feng, L. (2010). Value Added of Teachers in High-Poverty Schools and Lower-Poverty Schools. *CALDER Working Paper 52*.
- Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin*, 95(2), 122-140.
- Scherrer, J., & Stein, M. K. (2013). Effects of a coding intervention on what teachers learn to notice during whole-group discussion. *Journal of Mathematics Teacher Education*, 16(2), 105-124.
- Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, D.C.: U.S. Department of Education.
- Schoenfeld, A. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-371). New York: Macmillan.
- Schoenfeld, A. H., Minstrell, J., & van Zee, E. (1999). The detailed analysis of an established teacher's non-traditional lesson. *The Journal of Mathematical Behavior*, 18(3), 281-325.
- Scott, P., Ametller, J., Mortimer, E., & Emberton, J. (2010). Teaching and learning disciplinary knowledge: developing the dialogic space for an answer when there isn't even a question.

- In K. Littleton & C. Howe (Eds.), *Educational dialogues: understanding and promoting productive interaction* (pp. 289-303). New York: Routledge.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational researcher*, 29(7), 4-14.
- Shein, P. (2012). Seeing with two eyes: a teacher's use of gestures in questioning and revoicing to engage English Language Learners in the repair of mathematical errors. *Journal for Research in Mathematics Education*, 43(2), 182-222.
- Sinclair, J. M. H., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press Oxford.
- Sloan, F. A., & Pate, R. T. (1966). Teacher-pupil interaction in two approaches to mathematics. *The Elementary School Journal*, 67(3), 161-167.
- Smith, F., Hardman, F., Wall, K., & Mroz, M. (2004). Interactive whole class teaching in the National Literacy and Numeracy Strategies. *British Educational Research Journal*, 30(3), 395-411.
- Snow, C. (1979). The role of social interaction in language acquisition. In A. Collins (Ed.), *Children's language and communication* (pp. 157-182). Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Snow, C. E. (1982). Are parents language teachers. In K. Borman (Ed.), *Social life of children in a changing society* (pp. 81-95). Hillsdale, N.J.: Erlbaum.
- Solomon, Y., & Black, L. (2008). Talking to learn and learning to talk in the mathematics classroom. In N. Mercer & S. Hodgkinson (Eds.), *Exploring talk in school* (pp. 73-90). London: Sage.
- Staples, M. (2007). Supporting whole-class collaborative inquiry in a secondary mathematics classroom. *Cognition and Instruction*, 25(2), 161-217.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical Thinking and Learning*, 10(4), 313-340.
- Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation*, 2(1), 50-80.
- Stein, M. K., & Smith, M. S. (1998). Mathematical Tasks as a Framework for Reflection: From Research To Practice. *Mathematics Teaching in the Middle School*, 3(4), 268-275.
- Stevens, R. (1912). *The question as a measure of efficiency in instruction: a critical study of classroom practice*. New York: Teachers College Contributions to Education.

- Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.
- Stronge, J. H., Ward, T. J., & Grant, L. W. (2011). What Makes Good Teachers Good? A Cross-Case Analysis of the Connection Between Teacher Effectiveness and Student Achievement. *Journal of Teacher Education*, 62(4), 339-355.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4), 295-312.
- Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning, and schooling in social context*. Cambridge, UK: Cambridge University Press.
- Thorndike, E. L. (1906). *Principles of teaching based on psychology*. New York: AG Seiler.
- Tobin, K. (1986). Effects of teacher wait time on discourse characteristics in mathematics and language arts classes. *American Educational Research Journal*, 23(2), 191-200.
- van Kerkhove, B., & van Bendegem, J. P. (Eds.). (2010). *Perspectives on mathematical practices: bringing together philosophy of mathematics, sociology of mathematics, and mathematics education*. Dordrecht, The Netherlands: Springer.
- van Merriënboer, J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educational Psychology Review*, 17(1), 147-177.
- Voloshinov, V. N. (1986). *Marxism and the philosophy of language*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1987). Thinking and speech (N. Minick, Trans.). In R. Rieber & A. Carton (Eds.), *The collected works of L. S. Vygotsky*. New York, NY: Plenum Press.
- Webb, N. M., Nemer, K. M., & Ing, M. (2006). Small-group reflections: Parallels between teacher discourse and student behavior in peer-directed groups. *Journal of the Learning Sciences*, 15(1), 63-119.
- Webb, N. M., Troper, J. D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of educational psychology*, 87(3), 406-423.
- Wegerif, R. (2007). *Dialogic education and technology: expanding the space of learning*. New York: Springer.

- Wells, G. (1993). Reevaluating the IRF sequence: A proposal for the articulation of theories of activity and discourse for the analysis of teaching and learning in the classroom. *Linguistics and Education*, 5(1), 1-37.
- Wells, G. (1994). The complementary contributions of Halliday and Vygotsky to a “Language-based theory of learning”. *Linguistics and Education*, 6(1), 41-90.
- Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge: Cambridge University Press.
- Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. New York, NY: Cambridge University Press.
- Wood, D., Wood, H., & Middleton, D. (1978). An experimental evaluation of four face-to-face teaching strategies. *International journal of behavioral development*, 2, 131-147.
- Yackel, E. (2001). *Explanation, justification and argumentation in mathematics classrooms*. Proceedings from 25th Psychology of Mathematics Education Conference, Utrecht, Netherlands.