TRACING THE NATURALIZATION OF A LEARNING PROGRESSION CENTERED ASSESSMENT SYSTEM IN A TEACHER COMMUNITY

By

Min-Joung Kim

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Learning, Teaching, and Diversity

May, 2013

Nashville, Tennessee

Approved:

Professor Richard Lehrer

Professor Rogers Hall

Professor Ilana Horn

Professor Leona Schauble

Professor Norbert Ross

To God, for your love of me on the cross

&

To my husband, Jun-Young Lee, for always believing in me

and my two sons, Yea-Jun & Ha-Joon, for being in this world

ACKNOWLEDGEMENTS

The research reported was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305K060091 to Vanderbilt University. The opinions expressed are those of the author and do not represent the views of the U.S. Department of Education.

This dissertation would not have been possible without the support of many individuals. First and foremost, I offer thanks to Professor Richard Lehrer, my advisor and the chairman of my committee, for the invaluable guidance and continued support. You always had a minute or more to talk with me when I showed up at your office door. I always returned to my office space with problems solved and full of inspiration after talking with you and learned to be a more critical and passionate researcher. I would also like to thank the members of my dissertation committee, Professors Leona Schauble, Rogers Hall, Ilana Horn and Norbert Ross, for your constructive feedback and valuable insights.

I am grateful to all of the Lehrer-Schaublites for making such an awesome community and I am so proud of being part of it. To Marta Kobiela, Eve Manz, Rob Rouse, Erin Pfaff, Seth Jones, Shelly Cotterman, Mayumi Shinohara and Amy Holmes, I thank you for your friendship, your help editing my dissertation, lunch-time chats, and celebrations for special events of my life. I especially appreciate Eve Manz for engaging in productive and critical conversation on writing and for providing encouragement to keep me writing. To my Teaching and Learning colleagues, Lynsey Gibbons, and Ken Wright, I thank you for your time to read my dissertation and provide constructive feedback. To Marta Kobiela and Paul Weinburg, I thank you for being study-buddies in our office space and work-out buddies at the Commons at late night.

I thank you to Linda Jaslaw who provided infinite support – lodging, delicious food, positive thoughts, and conversation about teaching - during my two years of the data collection process. I thank you to all my participant teachers for opening your classrooms - enjoyed the time I spent getting to know you and wish you all the best!

I thank you to my best friends, Yearn-Jin Shim and Kyoung-Boon Kwon, for being there, listening to me, and encouraging me. I am grateful to all of my friends in Nashville Korean United Methodist Church, too many of whom there are to list here, for the companionship and prayers.

To my family, I am especially indebted. I love you all. To my parents and siblings, I thank you for always encouraging me and having faith in me during this long journey. To my parents-in-law, I thank you for providing continuous encouragement and prayers. To my American parents, Pat Denny and the late David Denny, I thank you for making my study-abroad experience like home and for sharing your family events with me. And finally, to my awesome husband, Jun-Young Lee, I thank you for being my counselor and being with me through the life journey.

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

"Improving education through the improvement of educational research" (Lagemann & Shulman, 1999) highlights the need for the mutual engagement of practitioners and researchers. However, differences between communities of practice and communities of research with respect to goals, job descriptions, rewards, and time constraints may limit the fruitfulness of such engagement (Hallinan, 1996; Klingner, Ahwee, Pilonieta, & Menendez, 2003). Instructional innovations designed by researchers may not fit with the goals and practices of teachers, or with the situational contexts in which they work (Bickel & Hattrup, 1995). Educational assessment is one of the areas about which researchers, teachers, and other stakeholders may have different and perhaps even incompatible goals (Darling-Hammond, 2004; Nolen, Horn, Ward, & Childers, 2011). For instance, standards-based reform urges policymakers and other managers to employ statewide assessments as tools for accountability (Darling-Hammond, 2004), but these assessments often fail to provide teachers with information that could be employed to improve instruction and learning (Black & Wiliam, 1998; Pellegrino, Chudowsky, & Glaser, 2001).

The National Research Council (Pellegrino et al., 2001; Wilson, 2009; Wilson & Bertenthal, 2005) has called for collaboration among learning researchers, psychometricians, and teachers in order to reorient assessment away from a system based solely on accountability toward one aimed at improving the quality of instruction and of student learning. As one way of doing so, the NRC recommended organizing assessment around "learning progressions," defined as "descriptions of successively more sophisticated ways of thinking about an idea that follow one another as students learn" (Wilson & Bertenthal, 2005, p.3). This study examines one such collaborative effort in the domain of statistics education, where development of an innovative assessment system was guided by a researcher-created progression of learning in the domain (Lehrer, Kim, Ayers, & Wilson, in press). Teachers and researchers came into contact through forms of professional development that introduced teachers to the assessment system and elicited teacher feedback about the intentions and content of the system. Teachers subsequently employed the assessment system in their classrooms and provided further reactions to researchers about its functioning.

To trace teachers' use of the assessment system, I frame the system as composed of "boundary objects" (Bowker & Star, 1999; Star & Griesemer, 1989; Wenger, 1998). I investigate the roles that this progression-centered assessment system, as a set of these boundary objects, played as researchers and teachers negotiated its status and meaning. Star and Griesemer (1989) suggest that boundary objects perform dual roles: (1) They serve as focal points around which multiple communities coordinate their activities, and (2) They function as tools to help each community accomplish its independent work. In addition to these functions, boundary objects may also disrupt established practices in communities. Hence, they may instigate transformation of practice in these communities (Akkerman & Bakker, 2011; Bowker & Star, 1999).

In this study, there are four elements of the assessment system that I designate as boundary objects: (1) construct maps, (2) assessment items, (3) scoring exemplars, and (4) lessons. I will expound on these components later, but briefly, construct maps are descriptions of the outcomes of learning progressions: forms of student reasoning targeted by the lessons and assessment items ordered according to a theory of learning from least to most sophisticated (Wilson, 2005). Assessment items are tasks designed to elicit the forms of reasoning described by the constructs. Scoring exemplars are interpretative frameworks relating student assessment responses to the constructs. The lessons consisted of instructional tasks and tools that were designed by the learning researchers to provide contexts where students could engage in the invention of representations, measures, and models of data, termed *data modeling* by Lehrer and Romberg (1996). Lessons and assessment items were intended to function jointly as tools for supporting the kinds of development envisioned by the learning progression, with its intended outcomes illustrated in the construct maps.

Each element of the assessment system had a different degree of locality in relation to each community of practice. Construct maps and scoring exemplars, representing a classification system of student reasoning, (i.e., the learning progression), originally resided in the researchers' world and thus were very new and unfamiliar objects in the teachers' world. In contrast, lessons and assessment items are historically the primary tools for teaching in schools and hence were more familiar objects to teachers. However, although the lessons and assessment items were forms that were familiar to teachers, they were designed with a less traditional approach to teaching and learning in mind, requiring negotiation about their meanings.

These boundary objects circulated between the worlds of teaching and research during professional development workshops (*Figure 1*). Although changes in the boundary objects required changes in researchers' practice and understanding of these objects as well (Lehrer et al., in press), I am purposefully limiting the scope of analysis in this study primarily to the teacher community.

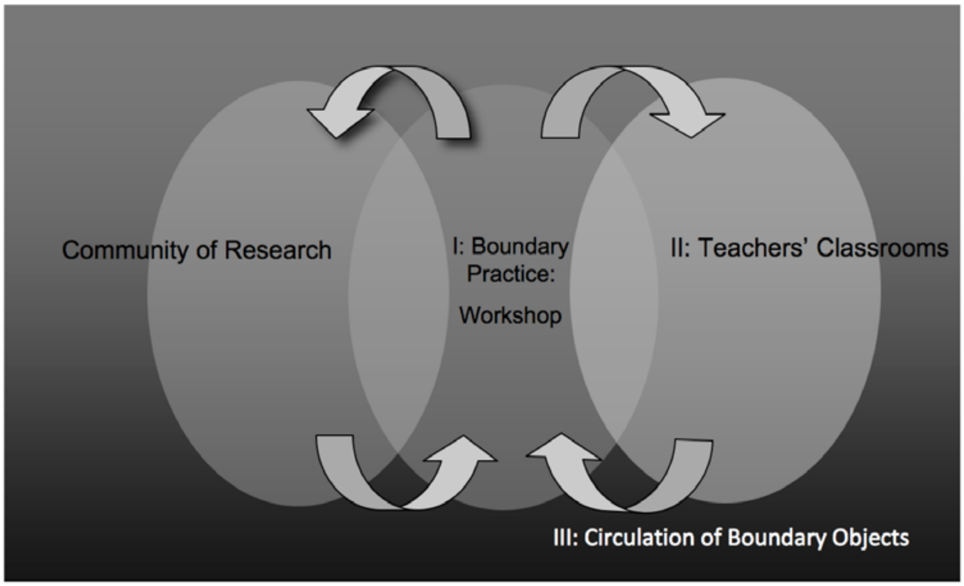


Figure 1. Configuration of the social worlds and circulation of the assessment system.

The goal of this study is to describe how the assessment system mediated the collaborative efforts between teachers and researchers in reorienting assessment toward improving the quality of instruction and supporting student learning. I trace two trajectories that were co-constituted and resulted in transformation of practice. The first describes changes in teachers' perspectives and practices of formative assessment that were mediated by the assessment system. The second describes transitions in the assessment system itself that emerged from the collaborative efforts of teachers and researchers.

To situate my investigation, I position my study within a broader context of research on formative assessment and teaching practices in mathematics education and briefly describe the theoretical entailments of boundary objects. I proceed to describe the questions that guided the conduct of my inquiry and describe methods that I employed to generate and analyze data. In the methodology section, I also describe the four components of the assessment system that served as the focal boundary objects of the study. Although it is not part of the analysis of this study, I briefly describe the structure of the professional development workshop during which the status and respective meanings of these boundary objects were negotiated and occasionally transformed. Following the presentation of results, I discuss the implications of the study.

CHAPTER II

FORMATIVE ASSESSMENT AS A SPECIALIZED FORM OF DIALOGUE

This study aims to understand how a learning-progression-centered assessment system can support teachers to enact formative assessment discussion as a specialized form of dialogue to make conceptual progress. The enactment of the specialized form of formative assessment talk requires the coordination of assessment and instruction: application of mathematical disciplinary perspectives in interpreting students' responses (Coffey, Hammer, Levin, & Grant, 2011) and enactment of particular forms of instructional moves in facilitating productive classroom discussions (Ball & Forzani, 2011; M. Stein, Engle, Smith, & Hughes, 2008). Assessment and instruction are not separable: ideally, effective teaching practice should assess student thinking constantly and make decisions about next instructional moves based on evidence of students' learning (Ball, 1993). However, assessment research and research about teaching and learning are not usually coordinated in mathematics education, with notable exceptions, such as Cognitively-Guided Instruction (CGI) (Carpenter, Fennema, & Franke, 1996; Carpenter, Fennema, Peterson, Chiang, & Loef, 1989; Fennema, Carpenter, & Franke, 1996). Because of the separation and different emphases in addressing educational issues in various research fields, some key issues in discipline-specific research are often not addressed in discipline-general research. Following the NRC's recommendation (2005), this study seeks ways to support teachers to connect assessment and instruction by conceptualizing formative assessment as a specialized form of dialogue.

First, I review formative assessment literature to identify challenges the field is trying to address. Secondly, I review empirical studies examining forms of instructional moves to support student learning during classroom discussion. Then, I will explain how I think a learning-progression-based assessment system might help teachers transform formative assessment practices to support student learning.

Formative Assessment

The term “Assessment” is closely associated with summative, high-stakes assessment in education. The function of high-stakes assessment is to evaluate overall performances of students (e.g., how much do they know?), teaching quality (e.g., how well did teachers teach, based on students' performances?), and hence the accountability of school systems. Although high-stakes assessment is important for district, state, or national policy, it is not informative enough for teachers to plan their daily instruction based on evidence of students’ understanding.

In contrast, formative assessment ideally informs instructional practices. Black and Wiliam (1998) define formative assessment as “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify teaching and learning activities in which they are engaged” (1998, pp.7-8). There is a consensus among researchers that formative assessment is very powerful for student learning (Black & Wiliam, 1998; Furtak et al., 2008), but there are differences in thinking about how to use formative assessment to get the best results and how to support teachers in using formative assessment in practice.

To date, much of the research in formative assessment has focused on developing assessment tools and tactics and strategies to implement assessment in classrooms. One approach to formative assessment focuses on expanding the traditional focus on multiple-choice items to include other forms, such as short essay questions (see, for example Treagust, Jacobowitz, Gallagher, and Parker, 2001). Some professional development programs that support formative assessment practice focus on tactics and strategies that teachers could employ in their classrooms and assume that teachers have sufficient content knowledge to use these practices productively. Wiliam (2007) stated, “The necessary changes are not changes in teacher knowledge – teachers know much of what they need to know already. The changes we need are changes in the habits and rituals of teachers’ practice that have been ingrained over many years” (p.201). For example, Black et al. (2003) suggested “longer wait time” as a way to improve questioning so that students had time to think about teacher questions and to get their responses ready. Regarding feedback, the researchers suggested providing feedback in the form of comments (rather than grades) because students did not read comments if grades were included. This line of research about formative assessment did not pay much attention to qualities of questioning and feedback in relation to discipline specific contents.

Although employing diverse assessment tools beyond multiple-choice items is an important change in assessment practice and may provide richer information about students’ understanding, others suggest that these forms of change are not sufficient to support opportunities for learning. Saxe, Gearhart, Franke, Howard, and Crockett (1999) reported that forms of assessment (exercise vs. open-ended) were not the main factor in changing classroom practice. They argued that assessment tools do not support teachers’ evaluation of students' mathematical understanding if the tools do not focus on mathematical thinking. In contrast to Wiliam (2007)'s claim about the sufficiency of teacher knowledge, Borko, Mayfield, Marion, Flexer, and Cumbo (1997) reported that teachers are hindered by the lack of a discipline specific framework to interpret students' responses. Borko et al. (1997) reported that scoring guides invented by teachers showed no guidelines about mathematical concepts, but instead were composed of literacy elements such as correct spelling, grammar and so on.

There is an emerging call for the need for developing a discipline-specific theory of assessment. More generally, the work of Hill and colleagues suggest that mathematics knowledge for teaching is critical for effective instruction (Hill et al., 2008; Hill, Rowan, & Ball, 2005). Recently, Coffey et al. (2011) reanalyzed classroom interactions that were represented in influential assessment journal articles written by Black et al. (2003) and Furtak et al. (2008) and identified a lack of focus on attending to disciplinary substance. Coffey et al. state, "Assessment, we contend, should be understood and presented as genuine engagement with ideas, continuous with the disciplinary practices science teaching should be working to cultivate" (p. 1109). Coffey et al. (2011) suggested that assessment practice in classrooms should be better aligned with disciplinary practices, including mathematical ideas and forms of discussing these ideas. Building on the premises and recommendations in using formative assessment to support learning of disciplinary ideas and practices, this study examines the process of teachers' adaptation of a learning progression-centered assessment system to orchestrate productive classroom discussion around core disciplinary ideas (Sztajn, Confrey, Wilson, & Edgington, 2012).

Forms of Instructional Moves to Support Student Learning

There is an emerging consensus in mathematics education field that productive classroom discussion facilitates students' mathematical learning (Cobb, Stephan, McClain, & Gravemeijer, 2001; M. L. Franke et al., 2009). Accordingly, the teacher's role is critical in discussion-based learning environments because the teacher is in the position of constantly coordinating students' thinking and disciplinary mathematical ideas.

Some researchers identified a series of forms of teaching practices involved in supporting student learning during classroom discussion. The orchestration of different ways and levels of students' thinking involves noticing and interpreting students' thinking, sequencing and supporting the development of relations among students' diverse thinking, and responding appropriately to the substance and tone of student thought (V. Jacobs, Lamb, Philipp, & Schappelle, 2009b; V. R. Jacobs, Lamb, & Philipp, 2010; M. Stein et al., 2008). Jacobs and her colleagues implicate "professional noticing of children's mathematical thinking" as critical for achieving collective mathematical understanding in a classroom community (V. Jacobs, Lamb, Philipp, & Schappelle, 2009a; V. R. Jacobs et al., 2010). Professional noticing involves a set of three interrelated skills: attending to children's strategies, interpreting children's understandings (connecting children's strategies to mathematical ideas), and deciding how to respond on the basis of children's understandings (coming up with problems that teachers might pose next). Jacobs et al. (2010) conducted structured interviews with four groups of teachers that varied in years of teaching experience and years of CGI workshop attendance. They used a cross-sectional analysis to trace developmental paths of the three skills in relation to teaching experience and professional development (i.e., prospective teachers and experienced practicing teachers with no professional development, 2 years of professional development on children's mathematical thinking and at least 4 years of professional development). The researchers found that teaching experience seemed to support teachers to develop the skills of attending and interpreting to some extent. They also found that teachers who participated in their professional development noticed significantly more details in children's strategies than those of prospective teachers and in-service teachers with no professional development experience. In addition, the researchers found that the skill of deciding how to respond was significantly related to years of participation in professional development, suggesting that the development of this skill requires particular learning opportunities. Considering that the study was conducted in the context of structured interviews based on both students' responses and classroom video and examined teachers' conjectured instructional moves in terms of problems to pose, it can be inferred that the skill of responding during moments of interactions might be even more difficult to develop and will require particular supports.

Pushing further than the teaching practice of eliciting initial students' responses, CGI researchers attended to qualitative characteristics of interactions between a teacher and students after the initial elicitation, because effective learning opportunities are created during follow-up interactions (NCTM, 1991). Franke et al. (2010) studied what forms of follow-up questions would be most effective in supporting students to be more explicit and complete in their mathematical explanation. They selected three 3rd grade teachers who participated in an algebraic reasoning CGI workshop for more than a year, and observed two math classes within a 1-week period. Their study found that asking a series of specific questions (e.g., composed with a series of more than two related questions about something specific about students' responses and composed of multiple exchanges of teacher questions and student responses) that probed mathematical ideas in students' responses led more frequently to complete and detailed explanations about mathematical ideas, in contrast to using one specific question (e.g., asking students to elaborate specific parts of their initial explanations) or a general question (e.g., asking students to repeat their explanations). A single turn of questioning suggested that a teacher did not unpack mathematical ideas hidden in students' strategies, and students did not have enough opportunities to understand either other students' strategies or relevant mathematical ideas.

The studies suggest that it is important to support teachers in developing instructional skills that orchestrate dialogue in integration with understanding of students' mathematical ideas. However, the studies used content-general criteria to analyze instructional moves. For example, Jacobs et al. (2010) used "more details of children's strategies and few details of children's strategies" to measure attending to children's strategies, "robust, limited, and lack of interpretation of children's understanding" to analyze teachers' interpreting children's understandings, and "robust, limited, and lack of use of children's understandings" to measure deciding how to respond. Analyzing several chains of interactions between a teacher and students based on content-specific criteria is expected to inform the effectiveness of instructional moves in making particular conceptual progress in students' understanding.

Learning Progression as a Framework to Support Teachers in Transforming Formative Assessment Talk as a Dialogue

Ball (1993), as an expert teacher, describes aims of her teaching practices, which center on supporting student learning of the mathematical discipline based on the students' own ways of thinking about mathematics. Ball (1993) stated:

> Among my aims is that of developing a practice that respects the integrity both of mathematics as a discipline and of children as mathematical thinkers … I seek to draw on the discipline of mathematics at its best. In so doing, I necessarily make choices about where and how to build which links and on what aspects of mathematics to rest my practice as teacher. With my ears to the ground, listening to my students, my eyes are focused on the mathematical horizon. (p. 376)

Her instructional decisions on "where and how to build which links and on what aspects of mathematics" were made based on where her students were in terms of mathematical understanding. She emphasized that her knowledge about mathematics was a key to identify mathematical seeds that she could nurture in her instruction. She suggested that teachers should notice mathematical substance in students' thinking and make instructional moves to connect student ideas to mathematical disciplinary content.

Ball's reflection is, at heart, a theory of learning-progression centered instruction, in that she focused on leveraging current students' understanding based on her knowledge of disciplinary mathematics and likely trajectories of conceptual development. While these ideas, therefore, have been previously explored, the concrete materials illustrating learning progressions have been created only recently in diverse strands in mathematics and science (Mohan, Chen, & Anderson, 2009; Songer, Kelcey, & Gotwals, 2009). A learning progression as a classification system, illustrating developmental pathways of disciplinary content, has been proposed as a practical means for supporting better integration of assessment and instruction (Wilson & Bertenthal, 2005). Assessment associated with a learning progression may provide teachers knowledge of the variability of student thinking and of prospective pathways of development of disciplinary knowledge (what can be built up toward what) so that teachers make instructional moves to connect the mathematical discipline and students' development.

Research about basing instruction on learning progressions is sparse but recently emergent because of the increasing promise of learning progressions. Researchers have started to conceptualize "learning trajectory[1] based instruction" (Sztajn et al., 2012, p.147). In this suggested research framework, separate areas of teaching (e.g., teacher knowledge, discourse tools, formative assessment, and task analysis) are organized around research on learning progressions. The researchers propose the need for empirical studies to test their conceptualization of instruction based on learning progressions. There are some early studies that explored how developmental frameworks in mathematical ideas supported changes in teaching practice. CGI (Cognitively Guided Instruction) is a representative content specific classification system, illustrating development of sophisticated forms of students' strategies and conceptual understanding of solving arithmetic in word problem solving contexts. According to the early studies of implementation of CGI, the researchers found that the CGI teachers tended to elicit multiple students' strategies and listen to problem solving processes rather than only answers more often than teachers in the control group (Carpenter et al., 1989). Similar to the findings of the early studies of implementing CGI framework, Wilson (2009) found that a learning progression used by K-2 teachers to teach equi-partitioning supported

[1] Sztajn et al. (2012) used "learning trajectory" to describe the conjectured pathways of understanding mathematical concepts. In the mathematics education, learning progression and learning trajectory are used as synonyms.

teachers at the level of assessing (e.g., eliciting and listening). However, he did not find evidence of teachers using it to support students' conceptual change based on the path outlined in the learning progression. He found that teachers mostly used the learning progression to select and sequence students' ideas but irregularly connected different students' ideas. He stated, "For a few teachers, knowledge of the learning trajectory provided a means by which teachers could sequence students' ideas to refine students' understandings of equi-partitioning. Largely, however, the results of teachers' selection and lack of sequencing tended to yield a lack of coherence and resolution" (p.187-188). This suggests that a taxonomy of states of student reasoning is perhaps necessary but not sufficient for supporting student learning. Teachers' pedagogical practices in orchestrating classroom discussion should be integrated in formative assessment. In sum, both understanding a taxonomy of states of student reasoning and developing pedagogical practices are necessary for teachers to orchestrate formative assessment as a specialized form of dialogue.

The research on supporting teachers' use of developmental frameworks suggests that the developmental frameworks support improved orchestration of classroom discussions, but they need further work on supporting teachers to develop effective "responding" skills in the moments of interaction to support student learning. This study is expected to contribute to the field, as an early study exploring the naturalization of the researcher-created assessment system in formative assessment talk and providing empirical evidence about how teachers adapted the learning progression-based assessment system to transform formative assessment talk and how the process of adaptation was supported.

CHAPTER III

METHODOLOGY

Theoretical Framework

As noted previously, I consider the assessment system as constituted by a set of boundary objects situated within and between each community of practice (Lave & Wenger, 1991; Wenger, 1998). Studying learning trajectory-based instruction involves at least two different communities: communities of teachers and researchers, which requires consideration about ways to mediate differences in the perspectives and practices of these distinct communities. Depicting collaboration between different professional communities as mediated by boundary objects acknowledges the inevitable differences among communities of practice, yet provides a venue for thinking about ways to "overcome discontinuities in actions or interactions that can emerge from sociocultural difference" (Akkerman & Bakker, 2011, p.136). It focuses on the process of "naturalization" of objects that become part of participants' daily practices (Bowker & Star, 1999, p. 299).

Roles of Boundary Objects in Communities of Practice

Communities of practice refer to the network of social relationships that are configured when people participate together in activities with shared goals. Simultaneously, people engage in "the process of giving form to our experience by producing objects that congeal this experience into 'thingness' " (Wenger, 1998, p.58), which Wenger termed “reification.” When these objects circulate among multiple communities of practice, they are called “boundary objects”:

> Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. (Star and Griesemer, 1989, p.393)

Boundary objects meet each community’s informational needs for performing their own jobs, yet have the potential to coordinate the process of developing modes of communication or routines to get things done smoothly across different practices (Akkerman & Bakker, 2011; Bowker & Star, 1999; Star & Griesemer, 1989). For example, a patient record in a hospital is a boundary object between doctors and nurses because it provides information on patients’ statuses. Doctors give orders based on the information in these records, and nurses give medicine in accordance with these orders.

One particular form of boundary object is a classification system, “a set of boxes (metaphorical or literal) into which things can be put to then do some kind of work-bureaucratic or knowledge production” (Bowker & Star, 1999, p.12). This type of boundary object is ubiquitous, perhaps because acts of classifying occur routinely in everyday life. Importantly, classification systems reflect “consistent, unique classificatory principles” (Bowker & Star, 1999, p.12). I consider the construct maps and scoring exemplars as a classification system because they create a taxonomy of forms of student reasoning.

Classification system-as-boundary-object enables people in different communities access to information so that its use can be coordinated across communities (Bowker & Star, 1999). However, a classification system is typically reified as a static artifact, such as a text document, and in doing so, the classification system strips away the processes that brought it into being. Wenger (1998) describes this aspect by using an iceberg analogy:

> What is important about all these objects is that they are only the tip of an iceberg, which indicates larger contexts of significance realized in human practices. Their character as reification is not only in their form but also in the processes by which they are integrated into these practices. (p.61)

A challenge for education reform is to consider the kinds of practices that provide support for people from different worlds to make a classification system become part of their unique practices without having to go through the same practices and reification processes that its inventors went through. Bowker and Star (1999) call this as "a trajectory of naturalization" (p. 299). Trajectories of naturalization are not pre-determined and generally develop over sustained periods of time. In this study, the classification system (e.g., construct maps and scoring exemplars) was designed to track student progress along 7 dimensions of conceptual development. Each dimension, or "construct" reified conceptual change as a series of transitions in the form and function of knowledge about statistics and data, and each was originally intended as a means for coordinating collaboration between the psychometric specialists at the Berkeley Evaluation and Research Center (BEAR) and the learning researchers at Vanderbilt University. For these two communities, the construct maps[2] were reifications of their participation in deciding what was worth assessing about data and statistics. They functioned to guide the development of items and the scoring exemplars. For teachers to make the classification system inform their unique practices of teaching, they would need to engage in the process of naturalizing this classification system. The theoretical framework provides a

[2] The 7 construct maps illustrate 7 strands of data modeling: Theory of Measurement, Data Display, Meta-representational competence, Conceptions of Statistics, Chance, Modeling Variability, and Informal Inference.

venue for thinking about ways to support teachers to make the classification system become part of their practice.

Although the role of boundary objects as coordinators of multiple communities of practice has received the most attention (Bowker & Star, 1999; Star & Griesemer, 1989), Akkerman and Bakker (2011) identify additional learning mechanisms that boundary objects support: reflection and transformation. Coordination among communities refers to the process of cooperating effectively to accomplish distributed work by adapting shared objects without necessarily establishing consensus about interpretations of the shared objects. Reflection refers to the process of interpreting the knowledge created in other communities and, as a result, taking and making perspectives that will specify what people do in future practice. Transformation describes emerging new practices that result from rigorous efforts to negotiate different perspectives, often with the support of deliberate intervention. Transformation is the most difficult learning mechanism to promote, and it involves several steps: People have to confront problems in their own practices when they interact with people from different communities, people from these different communities must share the identified problem, and then they generate solutions in the forms of new tools and models (hybridization in Akkerman & Bakker's term). However, transformation cannot end here. These new solutions must crystallize, or be integrated into daily practices. Further, transformation requires people from different worlds to engage in the process of negotiation of meaning for a long period of time.

Boundary Objects in an Educational System

Researchers in education (Cobb & McClain, 2006; Cobb, McClain, Lamberg, & Dean, 2003; Nolen et al., 2011; M. Stein & Coburn, 2008; M. Stein et al., 2008) employ boundary objects as an analytical framework to understand negotiation of the meanings of shared objects among different communities in education, such as administrators and teachers.

Different forms and substances of boundary objects can influence the nature of interaction and forms of practice supported. Through comparative analysis of two school districts, Stein and Coburn (2008) found that the two districts differed in participation structure and nature of interaction around different boundary objects, which they argued was partly due to the design of the different boundary objects. For example, one district adopted a curriculum that specified pre-determined steps for teaching mathematics and did not provide enough room to negotiate meanings of mathematical concepts. When coaches and teachers met for professional development, they focused on discussing logistics of implementation instead of attending to students' reasoning about mathematical ideas. Another district adopted a curriculum that was focused on students' mathematical reasoning, and researchers observed that teachers and district level leaders organized discourse and practice around mathematical thinking. The differences in substances of negotiation afforded by the boundary objects provided different kinds of learning opportunities to the participants.

Boundary objects can be interpreted differently according to the adaptor's perspective toward practices. For example, Cobb et al. (2003) found that a curriculum pacing guide that was intended by its designers to assist instructional planning was instead used by school leaders to judge whether or not teachers were on pace to cover state standards. The emphasis on accountability to standards in turn tended to promote teaching mathematical procedures in the teacher community.

In sum, the form of boundary objects and the goals of those using the boundary objects influence their educational utility and vitality. When boundary objects do not provide learning opportunities for people to construct knowledge about students' reasoning, they serve to align classroom practices with standards and accountability metrics (Cobb et al., 2003). One implication is that the introduction of boundary objects should provide opportunities to identify and disrupt different perspectives, so that people from different communities engage in developing new perspectives (Akkerman & Bakker, 2011; Hall, Stevens, & Torralba, 2002).

Research Questions

This study is guided by two sets of research questions. The first set is related to tracing a naturalization process of the learning progression-centered assessment system within the community of teachers. By naturalization, I refer to the appropriation and adaptation of the elements of the assessment system for the practical purpose of improving instruction. The second set probes relations between changes in teaching practices and changes in the assessment system. I separate my questions into the two groups for convenience of presentation, but I will examine the trajectories of change in practice and objects jointly to investigate how they influenced one another.

Trajectories of Changes in Practice

For the components of the assessment system to be considered as "boundary objects," they must satisfy two requirements. First, the component must be a focal point for researchers and teachers to communicate with each other during the course of their interactions, and second, it should meet teachers' needs in the classroom. The first set of research questions involves investigating how teachers naturalized the elements of the assessment system into their instructional practice and considers the extent to which teachers used the system as intended. The assessment system was designed to provide an interpretive framework for students' reasoning in data and statistics, organized as "learning progressions" (Pellegrino et al., 2001; Wilson & Bertenthal, 2005). Learning progressions are a promising assessment mechanism in that they provide a better sense of the development of students' understanding about conceptually important big ideas in math and science (Songer et al., 2009; Steedle & Shavelson, 2009). However, for this promise to become a standard of instructional practice, we should see evidence of this framing as teachers deploy the assessment system. Thus, I selected four case teachers who were situated in different school contexts (e.g., supportive in reform mathematics practice vs. strict on aligning with state standards) and who demonstrated different degrees of change in their teaching practices during their participation in the study. With these cases, I asked:

1. When teachers conduct classroom conversations about the results of an assessment, what are the forms of in-the-moment interactions among students and the teacher?
2. What kinds of changes in interactional structures are evident over time?

3. What are trajectories of change of the four case teachers over time?
4. How does the classification system contribute to changes in assessment practice?

Change in the Assessment System as a Consequence of Circulation

Another objective of this study is to document how the assessment system was modified through collaboration to accommodate the naturalization process, particularly for teachers. My questions focus on understanding how changes in practice were related to changes in the assessment system. In relation to change in the assessment system as a result of being shared by multiple communities, I ask the following question:

1. What changes in the form of the assessment system were required for it to support teachers to make changes in their practice?

Background Information of the Study

The Assessment System

The assessment system we shared with teachers is based on a learning progression that specifies cognitive milestones of learning to reason about data, chance and statistics. The assessment system was originally created to indicate students’ development of statistical reasoning, and the measurement model employed to interpret student responses to items served as one way to test researchers’ conjectures about forms and transitions in student reasoning. Because the assessment system was designed to be informative about student thinking, it also had the potential to be an effective teaching tool.

As mentioned, the assessment system consisted of four components. The four elements of the assessment system were all researcher-created objects, but each had different meanings and intended functions. Lessons and assessment items are commonly employed instructional tools for students' learning, although as I suggested previously, the emphasis on identifying and leveraging student thinking was unusual for the teacher participants. Construct maps and scoring exemplars are components of a classification system (Bowker & Star, 1999) that reflect the outcomes of the learning progression. These were unfamiliar forms to teachers.

Seven instructional units (lessons) were designed to support student learning about data and statistics. The lessons instantiated an approach to statistics education based on the conjecture that engaging students in the invention and revision of models would support learning about data and chance (Lehrer & Romberg, 1996; Lehrer & Schauble, 2000b; Petrosino, Lehrer, & Schauble, 2003). The invention and revision of data models consists of a set of interdependent practices, which include posing questions about phenomena, identifying attributes to measure, collecting data, structuring and displaying data, and making inferences. Moreover, data modeling integrates two strands of mathematics, data and chance, which are traditionally separated in most school instruction (C. Moore, Pure, & Furrow, 1990; D. S. Moore, 1990). In Wenger's terms, the lessons were reifications of the researcher's practice in design study classrooms: these were originally informal notes that described prospective relations between elements of the classroom learning ecology and student learning, but were later translated into curricular material more familiar to teachers, albeit with greater emphasis on revealing the intentions of the instructional activities than is typical of most curriculum (Davis &

Krajcik, 2005; Schneider & Krajcik, 2002). The instructional units were included in the assessment system to ensure that instruction and assessment were aligned. Alignment is one of the cornerstones of valid assessment (Wilson & Bertenthal, 2005).

Construct maps (Wilson, 2005) delineated progressive levels of understanding about data and statistics along seven related dimensions of learning about data modeling: theory of measure, data display, meta-representational competence, conceptions of statistics, chance, modeling variability and inference. Each construct map specified cognitive milestones in developing understanding, according to results obtained during a series of instructional design studies (Lehrer & Kim, 2009; Lehrer, Kim, & Schauble, 2007; Lehrer & Schauble, 2000a, 2004; Petrosino et al., 2003). Each construct map depicted cognitive milestones as learning performances-statements of the forms of cognitive activity consistent with a particular form of reasoning. One or more examples of each learning performance were included in the construct map. Appendix III includes the Conceptions of Statistics construct map for purposes of illustration. In addition to paper version construct maps, we created video annotated construct maps. Each performance on the construct maps were exemplified with edited video clips from the design studies, so that teachers could become familiar with learning performances situated in the familiar context of classrooms.

Items were designed to assess students' levels of understanding along these seven constructs. Multiple items were designed and tested to indicate the state of student knowledge about the cognitive milestones associated with each construct. The assessment items were essential research tools, in that they represented conjectures about encapsulating forms of student knowledge that were originally framed within contexts of classroom interaction and clinical interviews. At the same time, they were common objects, typical of schooling.

For each item, a scoring exemplar specified relations between prospective student responses on an assessment item and the levels of each construct map. These too were often revised during the course of development of the assessment system.

Workshop

The teacher and researcher communities came into contact through a teacher professional development workshop. The participants in this study were teachers from a southern state in the US who enrolled in a Data Modeling workshop. Classroom teachers and district coaches represented the teaching community and agreed to attend the workshop for one day every month during the school year. The Data Modeling workshop consisted of 13 sessions over two years, seven one-day sessions from October 2008 to May 2009 and six one-day sessions from September 2009 to March 2010. The workshops were conducted at a local educational cooperative. Rich Lehrer (the principal investigator) led the workshops. I was responsible for two workshop sessions during the school year 2008 - 2009. Thirty-four teachers attended the workshop in the first year. Twenty-nine teachers participated the second year, and seventeen of these teachers were continuing participants (See Table 1). The participants consisted of math specialists, math coaches, and math and science teachers. The schools served heterogeneous populations of students, including a large population of Southeast Asians and Hispanics.

Table 1. Participants of Data Modeling Workshop in 2008-2010

	2008-2009	*2009-2010*
# of Participants	34 teachers	29 teachers 17 previous participants 12 new participants

During the initial workshop, researchers introduced the goals and intentions of the collaboration. There were two main goals of the collaboration with the participating teachers. The first was to develop psychometrically valid measures of students' reasoning about data and statistics. The second was to develop an assessment system that could provide teachers with useful information for guiding instruction. Researchers asked teachers for help in making the assessment system more intelligible so that other teachers could use it.

The workshop sessions generally followed a consistent activity structure (See Appendix I for an example of a workshop agenda) that provided teachers opportunities to examine the assessment system from their perspective and to negotiate its meanings and functions with researchers, who were developing the assessment system with specific visions of educational reform. First, teachers participated in the same forms of data modeling that were the targets for instruction. Occasionally, these experiences were modified to problematize otherwise familiar content to teachers, such as how to calculate statistics of center. The professional development sought to augment calculation with conceptual foundations of statistics-as-measures of distribution characteristics. Also, researchers and teachers explored mathematical concepts of data and statistics (e.g., measures of spread, forms of statistical inference anchored to sampling distributions) that had been requested by the teachers. Second, teachers read the lessons with an eye toward understanding how particular instructional activities were designed to support the development of student reasoning. In this sense, the curriculum materials were educative (Davis & Krajcik, 2005; Schneider & Krajcik, 2002). Third, teachers examined the development of student reasoning illuminated by a construct map. Fourth, teachers reviewed items designed by the researchers to elicit particular milestones of reasoning and tried to anticipate student responses. They often looked at samples of student responses and located student responses via scoring exemplars to construct maps. As the workshops progressed, teachers brought their students' responses to items with them and looked at those.

In addition to the activities of reviewing the elements of the assessment system, the workshop was designed to facilitate the bidirectional negotiation of meanings and functions of the assessment system in the workshop sessions and teachers' classrooms. For example, after a workshop session, teachers implemented the assessment system in classrooms based on the functions and meanings they constructed during the workshop. In a subsequent workshop, they discussed their experiences with the assessment system and these experiences often resulted in clarifying differences in communal perspectives. For example, teachers often scored students' responses as either right or wrong, but researchers intended that student responses be more differentiated indicators of states of knowledge. On some occasions, teachers challenged the ordering implied by a construct map by referring to examples of how their students thought about an item or how they engaged in an instructional activity.

Researchers facilitated the negotiation between teachers and the assessment system by (1) asking for feedback on the assessment system, (2) linking teachers' experiences to big ideas regarding mathematics and the intentions of the assessment system, (3) responding to teachers' questions, and (4) providing some guiding questions in order to highlight the important perspectives on the assessment system (See Table 2). Guiding questions to get feedback on the intelligibility of the assessment system included: (1) What feedback do you have about the intelligibility of the lesson? (2) What did you think about the items? and (3) Do you have suggestions for revisions to items and scoring exemplars? We kept this activity structure for most of the workshop sessions and covered all construct maps, lesson sequences, items, and scoring exemplars except those regarding the *Informal Inference* progress variable.

Table 2. Structure of the Workshop & Guiding Question

Activity	*Guiding Questions*
Reflection on Classroom Activity	What did you learn about students' thinking on statistics and chance by trying out Lessons and/or Quizzes?
Looking at the Construct Map	What would progress look like when thinking about statistics as summarizing distribution?
Looking at Lessons	How can we support students to think about statistics as properties of distribution, not only as calculations? What feedback do you have about the intelligibility of the lesson?
Looking at Items and Exemplars	How is each scoring exemplar intelligible? What did you think about items? Do you have suggestions for new items or revisions to items?

Data Collection & Methods of Analysis

This study employs qualitative research methods (Denzin & Lincoln, 2000). I collected data from multiple sources including video and audio recordings of participants' teaching practices in classrooms, their interactions in the workshop and their responses to interviews. I also collected documents such as samples of student work and workshop materials. I conducted modified "teaching sets" (Simon & Tzur, 1999) to triangulate observed teachers' assessment practices with teachers' accounts. A teaching set consists of an observation in a classroom and a follow-up interview with the teacher about his/her intentions regarding specific instructional moves and about his/her rationale for the organization of classroom interactions. The video and audio recordings of classroom lessons and teacher interviews were analyzed by using discourse analysis (Gee, 1999), which allowed me to identify structures and patterns in discourse mediated by the assessment system. It also allowed me to track changes in practices in the classroom.

I focused on changes in one particular element of the assessment system: the video-annotated construct maps. The text versions of the construct maps were enough for researchers to conduct psychometric analyses, but we developed the video-annotated construct maps for teachers. The original version of the video-annotated construct maps was meant to exemplify each level of performance with excerpted video clips from the design study classrooms (Lehrer & Kim, 2009; Lehrer et al., 2007) to help teachers use the assessment items for instructional purposes. But, as I later describe more completely, the video exemplars were further elaborated to include episodes of formative assessment practice, initially drawn from the design study classrooms and later including episodes from participants' classrooms. The focus of the analysis is to examine how the trajectories of change in teacher practice and the trajectories of change in the assessment system co-evolved.

In the following section, I describe under each theme the process of collecting data and analyzing the data to answer the research questions. Consistent with the research questions, the data collection and analysis is organized by two themes: (1) trajectories of changes in practice and 2) change in the assessment system as a consequence of circulation between workshop and classroom.

Trajectories of Changes in Practice

Data From Classroom Observations. I recruited some of the workshop participants to conduct further study of their adaptation of the assessment system. To generate the sample, I categorized workshop participants into three groups based on their relative level of participation during the workshops. High-level participants were those who actively engaged in trying out the assessment system in their classrooms and frequently provided feedback on the assessment system at workshop sessions. Teachers who were rated at a medium level of participation were those who provided feedback on the assessment system only occasionally. Teachers who were rated at a low level of participation attended the workshop regularly but were relatively quiet during the workshop. By consulting with a local math specialist (the workshop coordinator at the regional district office) who worked with many of the workshop participants, I recruited teachers from each category. The sample selected served dual purposes. One was to see variations in assessment practice, and the other was to see changes in teachers' practices during the conduct of classroom discussions around assessment items. I did not recruit as many teachers at the low level as teachers at the high or medium levels, because I wanted to learn about teachers' use of the system when they were at least moderately engaged in its implementation.

In the first year (2008-2009), I recruited ten teachers for classroom study: five teachers at the high level, three teachers at the medium level, and two teachers at the low level. During the semester, the participating teachers invited colleagues whom they thought would benefit from the workshop as they did. For example, two teachers (Nancy and Sally) joined the workshop in January 2009. Although the teachers joined the workshop later, they very quickly fell into the “high” participation group and were added to the classroom observation list in March 2009. The teachers agreed to participate in four observations and interviews. Initially I audiorecorded lessons because I was unsure of whether teachers would feel comfortable being videotaped. As trust was established through our collaborative relationship, I started to video-record classroom interactions. As a result, the first two or three observations were audiotaped and the last one or two were videotaped in the first year. Recordings of classroom interaction in Year 1 consisted of teachers using either lessons or assessment items. During the first year, most teachers taught lessons during my classroom observations.

In the second year (2009-2010), five teachers stopped participating in the study after the first year for various reasons, including school constraints, overwork, promotion, and health problems. However, five teachers continued to participate, and I recruited two new teachers (Catherine & Maggie). Catherine’s level of participation was at the medium level and Maggie’s at the high level. Four of the teachers (Carla, Laura, Maggie & Nancy) were at the high level, two teachers (Catherine & Rana) at the medium level, and one teacher (Theresa) at the low level. Among the four teachers at the high level, three teachers who were at the high level in Year 1 continued to participate in the study and one teacher (Maggie) was a new participant in the second year. Rana, who was at the low level in Year 1, engaged in discussions more actively during the sessions that she attended in Year 2 and was classified as a medium level. In reverse, Theresa who was at the medium level in Year 1 became less engaged in Year 2.

The participating teachers and I planned to conduct five observations and interviews during the second year. All observations were videotaped in the second year. All teachers except Theresa and Rana were observed and interviewed five times. Theresa was promoted to a coach in the second year, so she felt a lot of pressure to figure out her role as a coach. She was not able to use the assessment system as much as she did in Year 1, and seemed to use it only when I visited her. I observed her four times in Year 2. Rana's school was under a school improvement program in the second year of the collaboration, and this program's requirements forced her to cancel the classroom observations that were scheduled in the middle of the school year. I was only able to observe her in the beginning of the school year and then after state testing was completed. The corpus of the observation data consisted of assessment item classroom conversation except for Theresa, whom I observed conducting one classroom conversation about assessment during the second year.

At each observation, I made notes on moments that I had questions about or that I thought interesting, and made sure I asked follow-up questions at the end of the observation. The data that I collected from classrooms also included students' work and photos of the whiteboard. Finally, teachers provided me with students' work that they had collected when they had tried the assessment system on their own.

For this dissertation, I selected four teachers (Theresa, Rana, Catherine, and Nancy) from four different schools considering (1) the degree of support from their schools, (2) their level of participation during the workshop, and (3) variations in trajectories of adapting the assessment system in their teaching practices. Particularly, three teachers (Theresa, Rana and Nancy) participated in the study for two years, thus providing opportunity for longitudinal analysis. Two teachers, Theresa and Rana, were subject to institutional pressures in the forms of pacing guides and accountability assessments. The other two teachers, Catherine and Nancy, reported experiencing less institutional pressure and more institutional support. For example, Catherine and Nancy described their principals as very supportive of reform-oriented mathematics instruction (e.g., incorporating student thinking), and they worked closely with their district math specialists.

Table 3. Cases selected from classroom study from 2008-2010

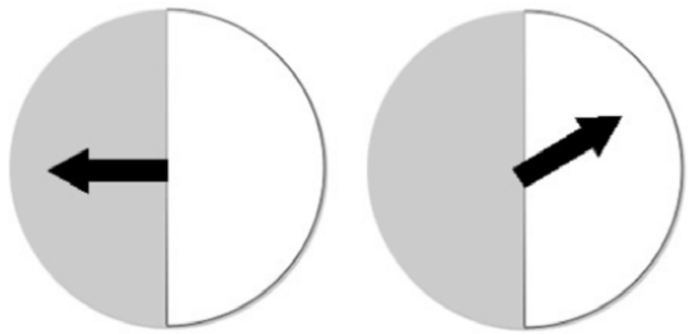
Name	Grade	Years of Teaching Experience	Institutional Context	Level of Participation	
				Year 1	Year 2
Theresa	6th	5	Standard test accountability focused/ Traditional school pedagogy centered	Medium	Low
Rana	7th	1	Standard test accountability focused/ Under school improvement program governed by the state	Low	Medium
Catherine	5th	1	Supportive leadership/ Reform oriented pedagogy encouraged	N/A	Medium
Nancy	5th	15	Supportive leadership/ Reform oriented pedagogy encouraged	High	High

Analysis of Classroom Observations. All classroom audio and video was transcribed. The transcripts were imported into InqScribe, a computer transcription tool, with classroom videos for the further analysis related to teacher or student gestures and inscriptions. As I elaborated the transcripts, I identified four distinct forms of formative assessment practice and used them to select episodes for further in-depth retrospective analysis. The categories were: (1) The teacher employing an I-R-E (Initiate-Respond-Evaluate) discourse pattern to communicate correctness of students' performances with students (Right vs. Wrong), (2) The teacher employing a turn-taking structure to share different students' responses but without any obvious regard to the states of knowledge described by the construct (Sharing student thinking), (3) The teacher eliciting students' responses that represented in the classification system following the order of sophistication (Eliciting particular learning performances), and (4) The teacher making connections (e.g., contrasting and comparing) among elicited students' responses (Making intentional connections among students' responses).

By broadly characterizing classroom interactions with the four categories, I selected samples of classroom observations for each teacher for further in-depth retrospective analysis. I particularly paid close attention to early observations and final observations to identify changes in how teachers orchestrated talk about assessments. To facilitate analysis of changing assessment practices within individual teachers and across teachers, I selected episodes across time where teachers used identical items or used items related to the same construct. For example, three case teachers (Rana, Catherine, and Nancy) used an assessment item, *Two Spinners* (Figure 2), and had instructional conversations with their students. This facilitated the comparison of the three teachers' particular instructional moves. The results of the analysis appear in Chapter IV (ANALYSIS OF CASES).

Then, retrospective analysis of classroom observations was conducted to identify moments that teachers orchestrated productive construct-centered assessment talk. I developed transcripts of interactions that filtered classroom talk as evidence of particular levels (forms) of reasoning according to the construct most closely related to the discussion. I also sought evidence of teaching moves consistent with intentions to support student learning. As an example, I present an episode from Nancy's classroom observations that I initially identified as "Making intentional connections among students' responses" (See Table 4). Nancy was orchestrating assessment talk about *Two Spinners* (Figure 2[3]), assessing students' understanding of the probability of a compound event.

[3] A gray box contains an elaborated description of an assessment item and related levels of performances. You can skip the box if you want to continue reading about analysis.



These two spinners are part of a carnival game. A player spins both spinners. To win a prize, **both** spinners must land in the **gray** section.

1. What is the probability of winning a prize?

   a. 1/2
   b. 1/3
   c. 1/4
   d. 3/4

2. Explain why you chose this answer.

The assessment item can elicit four different levels of student thinking from the Chance construct (Cha). The first is typical of students who think about the structure of each spinner without consideration of their joint action (NL ii); these students choose ½ because there are two spinners, and only one lands on the gray section. Other students focus on the instance displayed in the item, without considering repeated trials, and so respond that the probability would be ¼ (1 shaded region of the four regions of the two spinners). They think about the four parts of the two spinners as the total possible outcomes and the current particular outcome as a target outcome, choosing an answer of ¼, which is categorized as Cha 1B. Another possibility is treating the two spinners as a simple event. Students either think that total possible outcomes are four and target outcomes are two because there are two gray sections or the probability is going to be ½ by just looking at one spinner (Cha 3C). Finally, students may consider the combinations that can be generated by spinning the two spinners simultaneously, a response scored as Cha 6A.

Figure 2. Description of Two Spinners & related levels of performances.

Each turn of students' and a teacher's talk was coded according to levels of learning performances on the construct maps. For example, in Table 4, "S Cha 6A" indicates an inference about a student's level of mathematical understanding evident in talk, and "T Cha 6A" specifies a target performance that a teacher appears to support by particular instructional moves, such as juxtaposing. Then, the interactions of the levels of mathematical ideas between a teacher and students were inspected for two purposes: 1) to identify how well dynamics of levels of mathematical ideas in talk were aligned with the learning progression and 2) to examine forms of teachers' coordination of levels of students' mathematical understanding to the learning progression. The next step was to characterize instructional moves that were employed to foster conceptual changes.

Table 4. An example of transcript of interactions

#	Speaker	Transcript	Performances in Talks	Instructional Moves
1	Don:	I did change my answer to one fourth cause after what Eric said I realize that there's only one fourth of chance cause there's four outcomes that you can get.	S Cha 6A	
2	T:	Okay. You don't think there's fifty fifty chance of winning anymore half chance of winning. Okay. Um. Baylee what you are gonna say?		Contrasting Don's previous thinking vs. current Eliciting
3	Baylee:	Um. I chose one fourth because there are there two spinners but there's four there's four parts there's two parts on one spinner and then two parts (I just realize that) four parts and that's why I chose the four and that's how I took out anything that didn't have four in it. And then I got the one because there's only one gray part on each one.	S Cha 3C	
4	T:	Okay. I guess my question is why is it four. Is it because there's four spaces on there that we're looking at one two three four ((pointing at each section)) or is it because there's four different outcomes. I heard two different answers. I heard several people say well I think it's a fourth because there's four spaces and then I heard someone else say well no it's a fourth because there's four different outcomes.	T Cha 6A	Asking a level-specific question Juxtaposing Don's Cha 6A thinking vs. Baylee's Cha 3C thinking

Data From Teacher Interviews. Semi-structured post-observation interviews were conducted after each observation (See Appendix II). The interviews were directed toward understanding (1) what teachers noticed about student thinking during the course of their classroom conversation about one or more items, (2) teachers' perceptions of the intelligibility and utility of the assessment system, and (3) teachers' perceptions of teaching and learning mathematics in their classroom. I wanted to learn how teachers' perspectives on the functions of the assessment system changed as they engaged in the workshop and used the system in their classrooms. Hence, I asked questions regarding the intelligibility and utility of the assessment system both at the beginning and at the end of the study.

Questions about what teachers noticed about student thinking included: (1) What did you learn as you scored students' responses based on the scoring exemplars? (2) What would a student have to know about the relevant mathematical construct to correctly answer this item? (3) What did you notice about students' thinking regarding this item? (4) What difficulties did you notice students having when they solved the problem? (5) How did you help the students? and (6) Have you seen any changes in students' thinking today?

To address the intelligibility and utility of the assessment system, I asked teachers to rate their agreement regarding simple statements about each component of the system and to elaborate on their ratings based on their classroom experiences. The scale ranged from strongly disagree (1) to strongly agree (5). Simple statements included: "The lessons suggest productive ways of engaging students in learning," "The construct maps help me see the nature of progress," and "The progression outlined in paper version construct maps (or video-annotated construct maps, and exemplars) influences my teaching."

Questions about teachers' awareness of changes in their mathematical knowledge and their perception of math included: Has your participation in the partnership between Vanderbilt and teachers in the state changed your knowledge of, or the way you think about, math? and Have you experienced changes in what you know about how students think about data and statistics as you participated in the workshops?

Analysis of Teacher Interviews. The analysis of *teacher reflection* and *perceptions of teaching & learning mathematics* was intended to illustrate teachers' intentions behind their instructional moves and their organization of classroom interactions in relation to mathematical ideas illustrated in the assessment system. Transcripts of teacher reflection and perceptions of teaching & learning mathematics were divided by learning activities, assessment items, or the strands of constructs. As I analyzed classroom interactions and conjectured about teachers' intentions behind instructional moves, I read the transcripts of teacher reflection and perceptions of teaching & learning mathematics with an eye toward confirming or dismissing my conjectures by finding supporting or disconfirming evidences of them. Also, I paid close attention to teachers' attribution of their instructional moves to particular elements of the assessment system.

Teachers' Likert scales were put into an Excel sheet chronologically and were examined for significant changes in their ratings. Teachers' elaborations on their ratings were imported in NVIVO 9 (a qualitative analysis tool). Each teacher's elaborations were also arranged chronologically to facilitate the identification of significant changes in their perceptions of the elements of the assessment system in relation to their teaching practices. Teachers' elaborations on their perceptions of the assessment system were examined in relation to any significant changes in instructional moves identified by the analysis of classroom interactions.

Change in the Assessment System

Teachers' Written- and Verbal- Feedback & Think-aloud Protocol. Teachers provided feedback on the video-annotated construct maps through various channels. The data sources for the analysis include workshop video recordings, teacher interviews regarding the video annotated construct maps, video recordings of a talk-aloud protocol (Ericsson & Simon, 1984), and teacher notes. We asked teachers at the workshop and during interviews what they thought about the video annotated construct maps, how they were helpful for teaching practice, and how we could improve them so that they would be more helpful and useful for teachers. In addition, I conducted a think-aloud protocol (See Appendix II) to observe teachers' interpretations of the video-annotated construct maps. I asked teachers to say whatever came to mind as they interacted with these artifacts. Teacher explorations of the video-annotated construct maps were recorded by a screen capture program (IShowU ©). Some teachers provided me with notes that they took when they watched the videos by themselves, and these notes were also included as a data source.

Data Analysis. The focus of the analysis was what about teachers' practice motivated transformations in the video annotated construct maps and how these transformations influenced teachers' practices. The focus of the think-aloud protocol analysis is to examine what teachers noticed or what they looked for in the video annotated construct maps. I will describe how the analysis of the data was incorporated into changes in the video annotated construct maps, and how teachers both thought about the revised video annotated construct maps and took advantage of the revised construct maps.

CHAPTER IV

ANALYSIS OF CASES

In this section, I illustrate changes in the four cases of the teachers’ perspective and practice as mediated by the elements of the assessment system. The first case, Theresa, represents a case of making a little progress in formulating a new perspective on assessment or changing her teaching practice. As I mentioned previously, her institutional context was one of accountability to statewide assessments, and direct instruction seemed to have been a main model of instruction. Theresa did not conduct any assessment conversations during my visits in Year 1 and only demonstrated enactment of instructional activities from lessons. However, as a surrogate for the assessment conversations, I will illustrate her enactment of portions of lessons designed to provide opportunities for in situ formative assessment. During my visits in Year 2, Theresa conducted enactments of portions of the same lessons that she had used in Year 1. She also conducted discussions of formative assessment items. I will illustrate Theresa’s enactment of portions of the same lessons to compare her perspectives and practices over the two years. In addition, I will illustrate her enactment of a formative assessment item to illustrate how she orchestrated formative assessment talk as a form of dialogue.

Rana, as the second case, illustrates change in perspective about the interpretation of students’ responses and a shift in practice that represented a hybrid of her existing practice (e.g., eliciting procedural steps to get a right answer) and some new elements of practice that I later characterize as highlighting (Goodwin, 1994) and juxtaposition. Rana also worked in an institutional context of accountability. Rana’s school was directed to participate in a school improvement program initiated by the department of education of the state, which meant that teachers adhered to a particular curriculum and pacing guide aimed at enhancing students' test scores on a statewide examination. She mentioned that her school leadership encouraged strict adherence to state mandated accountability policies.

The third case, Catherine, exemplifies a pre-existing interest in student thinking that was augmented by the classification system. Her initial dichotomous perspective (right and wrong) on assessment appeared to change toward using the classification system to make distinctions among forms of student thinking. In addition, in her assessment practice, her questions changed from those that were more generic, content-general to those that probed more nuanced aspects of student thinking. Her institutional context was one in which the school principal supported efforts to re-orient mathematics education away from mere calculation toward meaning and dialogue.

The last case, Nancy, engaged in an earnest negotiation process with disruption incumbent to using the assessment system. Most of all, she demonstrated an instructional trajectory that incorporated the learning progression. Nancy also worked at a school that provided institutional support to teachers in adapting reform oriented instructional approaches to support students' learning.

Theresa: Developing a Rough Categorization of Mathematical Ideas

Theresa had been teaching for five years, and was in her second year of teaching sixth grade math when she started participating in the study. She had a bachelor's degree in educational sciences. Her original certificate was for preschool through fourth grade.

She then completed additional coursework for an endorsement to teach fifth and sixth grades.

Theresa’s school had historically embraced traditional forms of mathematics teaching but had recently switched to a more reform-oriented approach. Her school had used a very traditional textbook (i.e., Saxon math) that focused on teaching procedures but recently had changed its textbook, according to her colleague’s description, to one more oriented toward engaging students in doing mathematics (i.e., Glencoe). Theresa and her colleague, a fellow teacher, often mentioned that reform oriented strategies and tactics such as “hands-on activities” and “discussion based class” were their instructional foci, but they did not explicitly address mathematical ideas in relation to these reform oriented strategies. Theresa’s math coach also attended the workshops, but the coach did not seem to actively collaborate with her teachers to explore the assessment system. When the coach visited Theresa’s class one time, she sat in the back of the class and did not participate in teaching the class.

Theresa’s participation in the second year of the study was limited by her shift in roles within her school, as she was promoted to the position of math coach when her former coach left the school. Although Theresa arranged team-teaching with her colleague, she ended up working in her colleague’s classroom only when I visited her school. Theresa described her colleague’s instruction as traditional lecture and rare discussion. In addition to having limited access to a classroom to try the assessment system, Theresa lost opportunities to learn and negotiate meanings of the elements of the assessment system with other teachers and researchers. Theresa only attended three professional development sessions out of six in Year 2 because of conflicts with her school schedule (e.g., target testing). Theresa explained that her new job required a lot of administrative duties (e.g., preparing teachers and students for the state standardized test and benchmark tests, making mock-up tests, and attending district meetings). As an instructional leader, she was a resource person who located curriculum materials (e.g., classroom activities, manipulatives) to support other teachers. These expectations seemed to be distant from that of supporting teachers with mathematical ideas and student thinking.

Theresa's Practice in Year 1

Centering Classroom Discussion on Mathematical Substance & Helping Students Experience Mathematics as a Form of Sense-Making. Theresa expressed strong interest in the new approach to data, chance, and statistics illustrated in the assessment system and actively participated in the workshops in Year 1. The approach taken in the assessment system is to orient teachers toward the kinds of reasoning about data display and statistics that typically guide the practice of the discipline. Theresa demonstrated an ability to center classroom discussions around the big ideas of data, chance, and statistics. For example, during the last lesson in the Inventing Displays sequence, "Describing and Comparing Displays," Theresa appeared to look for particular forms of student reasoning that are illustrated in the constructs describing landmarks in student reasoning about representational competencies and about meta-representational competencies. In the lesson, students produce a set of data having measurement errors, identify patterns in the class's measurements and invent displays that show the identified patterns. The activity can elicit all levels of performances in the Data Display and Meta-representational Competence constructs. The Data Display construct (DaD) largely characterizes the development of students’ understanding about displays from a case specific perspective (e.g., focusing on specific data points such as minimum and maximum) to an aggregate perspective (e.g., center clump and shape of distribution). The Meta-representational Competence construct (MRC) outlines the progression of understanding about forms and functions of displays. The important conceptual achievement outlined in the construct is to select displays that best support arguments based on understanding what displays show and hide about patterns in data.

On this day (November 2008), each student measured the circumference of Theresa’s head and invented displays working in small groups. The invented displays varied both in type and in the interval on the X-axis. For example, two groups created frequency graphs with intervals of 2s (See Figure 3) and 5s (See Figure 4). Another two groups created stem-and-leaf plots (See Figure 4).

In the excerpt that follows, the class was looking at the frequency graph with the interval of two (Figure 3) as they engaged in sharing their noticing about the display (Excerpt 1). This class was audio recorded, limiting detailed transcription of gestures.

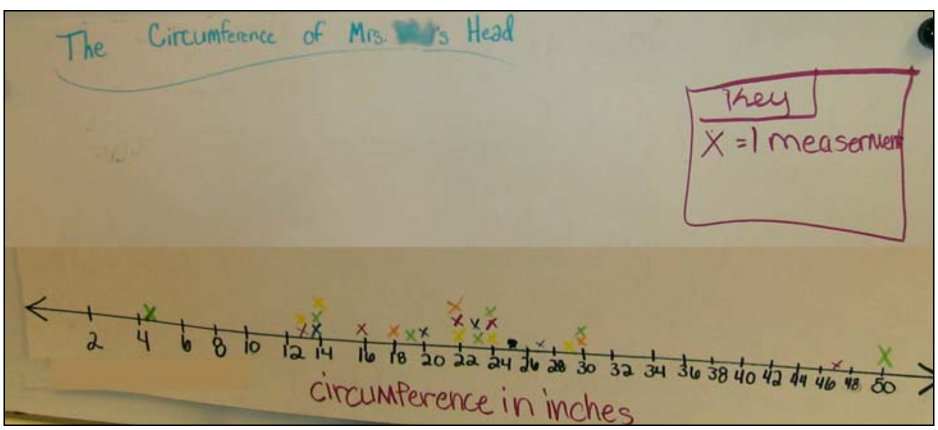


Figure 3. A frequency graph with the interval of 2s created by a group of student in Theresa’s class.

In Excerpt 1, Theresa appeared to support students' sense-making of big ideas of data displays that are illustrated in the constructs. She elicited mathematical ideas by using the Thought-Revealing-Questions in lesson 1 and by pressing students to explain their reasoning behind their answers.

Excerpt 1

1	S:	((inaudible)) counting by two.
2	T:	Okay. Counting by two's so we have an interval of two? Anybody else? Notice anything about that one? What do you see in the data here? What stands out to you when you look at our data display? Sydney?
3	Sydney:	Key (can) help…it has a big gap it says from measer.
4	T:	Hmm hmm. The keys say, but let's look at the data. Right here ((invisible but it is conjectured that Theresa is referring to the number line)).
5	S:	Colors, they're different colors.
6	T:	Kelly?
7	Kelly:	The majority of the x's are like in the center.
8	T:	The majo::rity of the x's are in the CENTER of our number line, our little piece of the number line there. So what do you think that shows?
9	Kelly:	What uh that most people uh got about the same.
10	T:	Very good. And most people got about the same measurement. Most people not all but most. Okay.
		….
11	T:	Sydney?
12	Sydney:	Notice it has outliers.
13	T:	You can notice the outliers really easily because they're a long ways away from everything else aren't they?

The first piece of evidence of aligning instruction with mathematical substance from the construct map is that Theresa intentionally drew students' attention to patterns in data when students focused on characteristics of displays not relevant to the data structure (e.g., having a key in a graph). It was invisible in the audio-recording, but I noted that she said phrases such as "here" and "right here" during the discussion to redirect students’ attention to patterns in data. Several students shared their noticing of characteristics of the frequency graph that were not related to data structure (DaD 1A: Interpret data displays without relating to the goals of the inquiry and MRC 2B: List observed characteristics of displays without explicit reference to data structure or purpose of data collection). For example, Sydney pointed out that the display had a key (line 3, “Key (can) help”), and another student mentioned that the data points were in different colors (line 5, “Colors, they’re different colors”). In response to this type of noticing, Theresa redirected students’ focus toward the data (line 2, “What do you see in the data here?” and line 4, “let’s look at the data. Right here”), pushing students toward MRC 3 (Articulate how features of display reveal something about the structure of the data).

Another example of Theresa teaching toward learning performances on the construct map can be seen in lines 7-10. Kelly observed that the majority of marks were in the center (MRC 2A: List and compare observed characteristics of displays without explicit reference to data structure or purpose of data collection). Theresa repeated Kelly’s noticing as she prolonged pronunciation (e.g., majo::rity) and used volume (e.g., CENTER) to emphasize important mathematical ideas in Kelly’s noticing. Following this emphasis, Theresa asked a question to help Kelly connect his initial observation to the purpose of data collection (line 8, “So what do you think that shows?”), targeting a higher level of representational competence (MRC 3: Articulate how features of display reveal something about the structure of the data), where the display is viewed as constructed with the purpose of the data collection process firmly in mind.

As illustrated in the previous paragraphs, Theresa noticed different levels of mathematical substance in students’ responses and directed students toward thinking about important patterns in data. However, she did not yet make connections visible among different students' noticing about data as a way to support student learning and often wrapped up interactions with students with strong feedback (e.g., "very good") on their thinking. For example, in Excerpt 1, there were three students who shared important mathematical ideas about the display. First, in line 1, a student shared that she noticed an important form (i.e., interval) of the frequency graph ("counting by two"). Then in line 3, Sydney noticed an important structure of the data (i.e., "a big gap"). In line 12, Sydney shared that she noticed another structure of the data (i.e., "it has outliers"). All these ideas are pieces of a big idea (DaD 4A: Display data in ways that use its continuous scale to see holes and clumps in the data). Because the frequency graph used a continuous scale with the interval of two, it showed outliers and gaps. Theresa accepted students' noticing by repeating them (line 2, "Okay. Counting by two's so we have an interval of two?" and line 13, "You can notice the outliers easily…"). However, Theresa did not build on the noticing to help students make close connection between forms (e.g., interval) and functions (e.g., showing a big gap and outlier) and learn a higher level of thinking (DaD 4A).

As students shared their common noticings (e.g., outliers, center clump), Theresa acknowledged their ideas, supported their sense-making of data display and ensured that the noticings were made public for the class. She often used "very" to express her strong agreement with students' responses. For example, in Excerpt 1, Theresa asked Kelly what the majority of Xs in the center showed, and Kelly replied that it meant that most people got about the same measurement. Then, Theresa indicated that Kelly's response was right by saying, “Very good.” Theresa seemed to be moving past the traditional I-R-E discourse pattern and eliciting student thinking.

Theresa continued supporting her students to make connections between features of the display and the structure of the data in the following excerpt, in which the class discussed two other displays: a stem and leaf plot and a frequency graph with the interval of 5 (Figure 4).

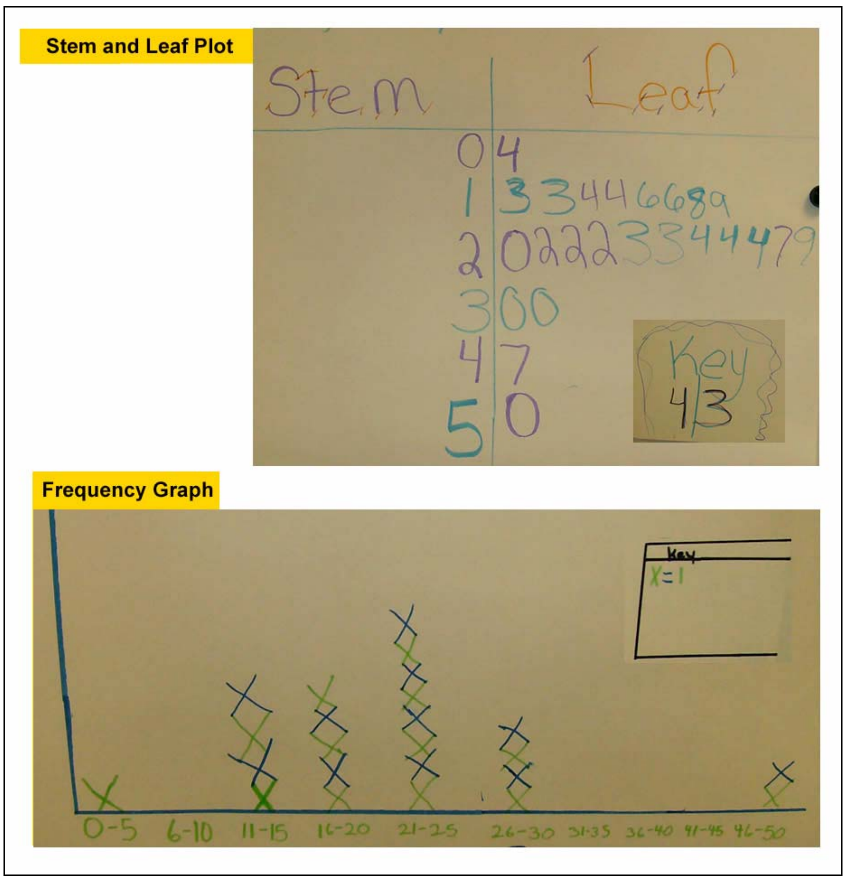


Figure 4. A stem and leaf plot & a frequency graph with the interval of 5s.

Excerpt 2 also illustrates Theresa noticing and acknowledging bits and pieces of mathematical ideas about data displays provided by students but not integrating these ideas toward a higher level of performance.

Excerpt 2

1	T:	Stem and leaf. And what does the stem and leaf show? Angeline.
2	Angeline:	It shows that most of the measurements were bet- were either in the tens and the twenties.
3	T:	Okay it shows that most of the measurements were in the tens and twenties. Very good thought.
4	S:	It does show what's the outliers.
5	T:	It also shows the outliers, doesn't it? Because there aren't very many up there where the stem is 0, where you have a 0 and a 10 place, there's just one isn't there?
		((transition to the frequency graph))
6	T:	Okay the third one, what does it hide and show? Angeline.
7	Angeline:	It makes it; it makes it look like there's not really a big outlier.
8	T:	Okay. You don't notice the outliers as much. Why do you think that is on that one?
9	Angeline:	Because they put the numbers, the intervals were bigger.
10	T:	The intervals were bigger. Very, very good. Tammie what were you going to say?
11	Tammie:	Um.
12	T:	Don't remember okay. Kelly?
13	Kelly:	By what they have, you can tell that mostly they're in 21 through 25.
14	T:	Okay. So you can tell really quickly that most people were between 21 and 25 or there was a bigger majority between, in, in that interval or in that bin, okay.

Theresa employed a mix of a transformed I-R-E discourse pattern and a turn-taking structure in sharing students' thinking about the stem-and-leaf plot. Theresa initiated the discussion about the stem-and-leaf plot by a Thought-Revealing Question in the lesson and called on Angeline in line 1 (Initiate). Angeline reported her noticing of the center clump in line 2 (Respond). Then Theresa repeated Angeline's answer and provided her evaluation in line 3 (Evaluate: “Very good thought”). Following Angeline’s noticing, a student voluntarily shared her noticing that the stem-and-leaf plot showed outliers (Respond: line 4, “It does show what’s the outliers”). Theresa strongly agreed that the stem-and-leaf plot showed the outliers (Evaluate: line 5, “It also shows the outliers, doesn’t it?”). Then Theresa provided her justification that an outlier existed (line 5, “Because there aren’t very many up there where the stem is 0, where you have a 0 and a 10 place, there’s just one isn’t there?”). Theresa’s justification of the outlier only described the frequency of data in a stem and did not consider the distance from the clump.

In this interaction, Theresa seemed to be satisfied that students noticed a clump and mentioned outliers from the stem-and-leaf plot (DaD 2A and DaD 3A) but did not examine students’ noticings in relation to the distribution of data. As evidence, the stem-and-leaf plot hid outliers. Theresa did not problematize the student’s reasoning and did not ask follow-up questions to understand why the student thought that the stem-and-leaf plot showed outliers.

The class moved on to discuss another graph, a frequency graph with the interval of 5s. Theresa again initiated the conversation with a Thought-Revealing-Question from the lesson (line 6, “what does it hide and show?”). Angeline noticed that the frequency display did not make an outlier look as distant as she thought it might (line 7). In return, Theresa asked Angeline her rationale for a pattern in the data, making a connection between forms and functions of the display (line 8, “Why do you think that is on that one?”). Angeline responded that the size of interval mattered in making outliers less visible, coordinating forms and functions of the display (line 9, “Because they put the numbers, the intervals were bigger”). This is supported in the construct map - exploring effects of “bin” size on the shape of the data (DaD 4B: Recognize the effects of changing bin size on the shape of the distribution). Theresa wrapped up her conversation with Angeline by providing her evaluation (line 10, “Very, very good.”) and called on another student to elicit the student’s noticing.

The post observational interview with Theresa supported the interpretation that her instructional moves were intended to support her students to think beyond traditionally emphasized features of data displays. Theresa said:

> We’re pushing them to extend their thinking instead of just being satisfied with yes they can make a graph and they can put the title on it and they can put a key on it, you know. They need to go beyond that and for so long we’ve been so stuck on, oh you don’t have a title, you know, it’s a bad graph … they do have to have those things [a title and a key] on there but that’s not what’s most important. What’s most important is the data. [Post Interview, November 2008]

Theresa acknowledged that her previous instruction on data display had been more focused on teaching how to make conventional graphs correctly without relating forms to data structure. She identified renewed instructional goals (“they do have to have those things [a title and a key] on there but that’s not what’s most important. What’s most important is the data, you know”) that were aligned with the assessment system, consistent with her focus during classroom discussion. However, Theresa did not talk explicitly about the categories of the constructs and did not explicitly represent her attempts to have students relate design choices, such as the width of the interval, to the shape of the data, suggesting that she had a rough categorization of mathematical ideas in students’ responses (e.g., making sense of the data).

Theresa’s classroom discourse pattern illustrated above can be interpreted as an intermediate step toward coordinating classroom discourse to affect students’ learning.

Her interview indicates that not only did Theresa make an effort to change the mathematical substance of her class, but she also tried to improve her discourse practice. She shared her difficulties with teaching the data display lesson when she attempted it the first time:

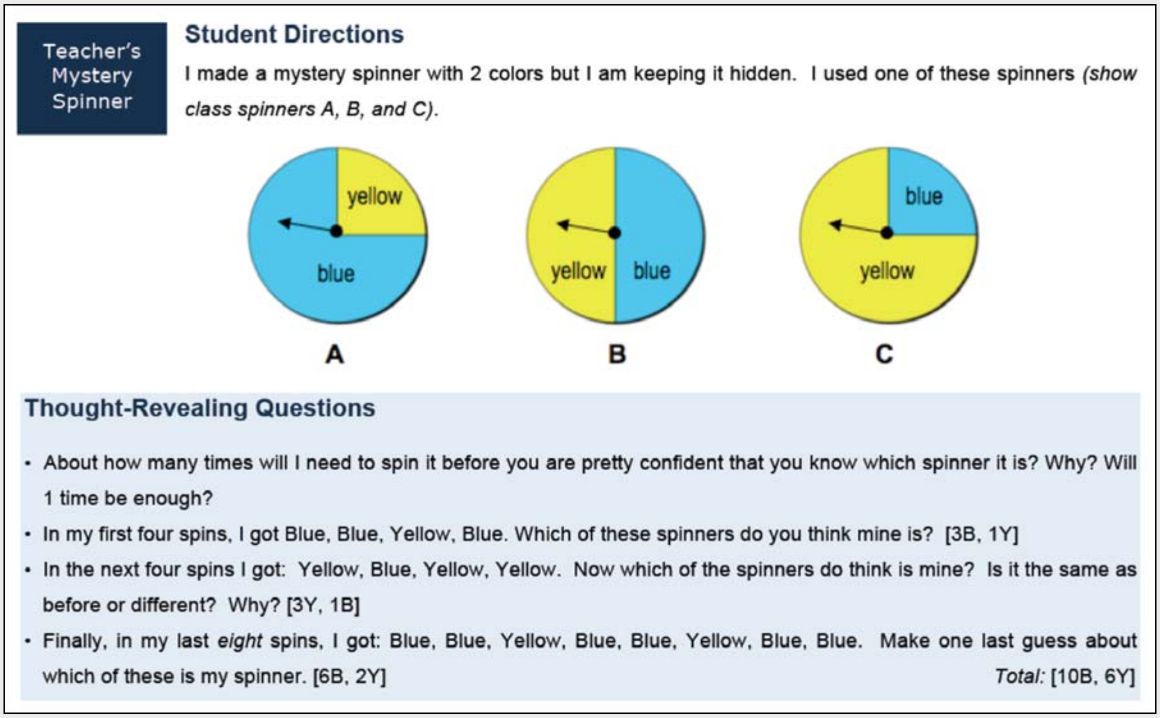
> Because even though I can look at those graphs and I can see, well you know this one shows this and this one shows this, I really have a hard time questioning the kids without just giving them the answer. … one time I just said, you know, this is what I see and this is what I don't see and just told them, you know, everything and they didn't have any part of the discussion. [Post Interview, November 2008]

Theresa believed that her role as a teacher was to ask good questions so that students constructed their own knowledge instead of giving answers. However, she told students important mathematical ideas because she had a hard time generating effective questions she could use. She seemed to have explored the provided lesson plan more to find productive questions.

> And so that's something that when I went back and looked all the way through that lesson plan cause I didn't look all the way through it the first time and I realized, oh there's all those examples in there, and I really like having those photographs of the actual graphs and then, you know, the little descriptions about, you know, what this was or what the kids said about it, things like that. It really helps me develop better questions or even just steal those questions. I'm, I'm, you know, I'm shameless. I don't mind to use them. [Post Interview, November 2008]

As Theresa iterated the instructional activity, she tried to improve her instructional strategies to teach the big ideas of data display. She explained that Thought-Revealing-Questions and exemplary student work in the lesson helped her have better discussions with her students. She was eager to appropriate some of the questions suggested as aides for revealing student thinking and appeared to embrace the intention of helping students experience mathematics as a form of sense-making.

Developing a Rough Categorization of Mathematical Ideas & Approximate Instructional Intention of Learning Activities. Later in the year, Theresa taught a portion of the curriculum intended to support the development of conceptions of chance (Figure 5). Variability in chance is rarely taught in school mathematics (Shaughnessy, 1997), and the participant teachers said that one of the big ideas of chance, the law of large numbers, was very new to them. It was innovative for the teachers to introduce these ideas to their students, as they mainly taught calculation of theoretical probability as emphasized in school mathematics.



Teacher's Mystery Spinner

Student Directions

I made a mystery spinner with 2 colors but I am keeping it hidden. I used one of these spinners *(show class spinners A, B, and C)*.

Thought-Revealing Questions

- About how many times will I need to spin it before you are pretty confident that you know which spinner it is? Why? Will 1 time be enough?
- In my first four spins, I got Blue, Blue, Yellow, Blue. Which of these spinners do you think mine is? [3B, 1Y]
- In the next four spins I got: Yellow, Blue, Yellow, Yellow. Now which of the spinners do think is mine? Is it the same as before or different? Why? [3Y, 1B]
- Finally, in my last *eight* spins, I got: Blue, Blue, Yellow, Blue, Blue, Yellow, Blue, Blue. Make one last guess about which of these is my spinner. [6B, 2Y] *Total:* [10B, 6Y]

The "Teacher's Mystery Spinner" activity Theresa used in March 2009 focuses on a central idea in the study of chance: the experimental probability of an outcome approaches its theoretical probability in the long run. The task is to predict the structure of the teacher's mystery spinner based solely on its outcomes. There are three spinners that students can choose: Spinner A has ¾ colored in blue and ¼ in yellow. Spinner B has ½ of yellow and ½ of blue. Spinner C has ¾ of yellow and ¼ of blue. Students are asked to make a guess about the structure when the results of 4 spins are given, 8 spins, and then 16 spins.

The performances on the Chance construct that the activity can elicit are as follows: from complete absence of structure regarding chance (Cha 1: Hold an informal view of chance), to quantifying theoretical probability or frequency (Cha 3: Quantify chance as probability and relate it to the structure of a simple event), to understanding the relationship between theoretical probability and empirical probability in many repetitions of an event (Cha 4: Empirically examine the relationship between observations and all possible outcomes of repeated simple events). The top performance that we would expect a teacher to elicit and support in this lesson is Cha 4D (Recognize that, with enough repetitions of an event, the relative frequency of an outcome will approach its theoretical probability).

Figure 5. Activity: Teacher's Mystery Spinner.

During a classroom discussion in Year 1, Theresa tried to make the uncertainty of experimental results visible by echoing and further elaborating on a student's idea. When Theresa asked her students if they could figure out the teacher's mystery spinner based on experimental outcomes, a student said, "You really couldn't figure out with the experimental because like the 50 and 50 that we did while ago they weren't even. And so it could be any of `em." The student was arguing he would not know what the spinner looked like based on the experimental outcomes. In return, Theresa reminded students of their past experience with unlikely strings of outcomes ("How we have, you know, a spinner or we flip a coin, and sometimes it's just a >long<ways away from our theoretical probability. So it's possible that we couldn't figure out at all").

A further illustration of Theresa emphasizing the uncertainty of experimental results can be seen in Excerpt 3. During this conversation, Theresa announced the result of the first four spins and asked students what they thought the teacher's mystery spinner would be. Several students said the teacher's spinner could be A or B. Nobody said it could be C, which had ¾ of yellow and ¼ of blue. So Theresa asked if students considered spinner C as a possible option.

Excerpt 3

1	T:	Could be A or B, but you don't think it could be C?
2	Hope:	No.
3	Lee:	It could.
4	T:	It could?
5	Lee:	It's possible.
6	T:	Why is it possible?
7	Lee:	Because there's like a little bit of red [4] [yellow] and there's still some C [blue], but you never know what's gonna land on.
8	T:	That's right. With experimental, we don't ever know. So and that was only four spins.

Here, Theresa made an instructional move that led students to consider the nature of experimental probability. In line 1, Theresa asked a question that explored students' thinking about Spinner C as possibility a possible option based on the four outcomes. Her question ("Could be A or B, but you don't think it could be C?") played an important role, making the class remain uncertain about the design of the teacher's mystery spinner with the short numbers of trials. For example, students like Hope who thought Spinner C was not possible might have had a second thought based on Lee's claim (line 5 and 7, "It's possible. Because there's like a little bit of red, and there's still some C [blue], but you never know what's gonna land on."). Theresa concluded students' guess of the mystery spinner as undecided based on the first four outcomes, keeping the uncertainty of experimental results alive ("That's right. With experimental, we don't ever know. So and that was only four spins").

The illustrated examples suggest that Theresa took the mathematical discipline oriented perspective shared through the workshop as well as the assessment system into

[4] The original colors of the spinner were yellow and blue. Black and white copies of the page made it hard to differentiate the original colors. Theresa changed the colors to red and blue.

account as she attempted to incorporate big ideas of chance during classroom discussion in Year 1. It also shows that Theresa did not have a firm understanding about the law of large numbers or hold firmly in mind the instructional intention of the learning activity, which was making connections between numbers of trials and trial-to-trial variability. For example, in Excerpt 3, Theresa showed that she did not consider more repetitions of a process as a better basis for an estimate. At Theresa's request to justify Lee's claim about Spinner C (line 6), Lee did not attribute "it's possible" to numbers of trials, rather he pointed at the structure of the spinner (line 7, "Because there's like a little bit of red, and there's still some C [blue]"). If he had mentioned anything about the number of trials, he would have been placed at Cha 4C (Recognize that an unlikely string of outcomes is possible and even expected over many repetitions of the event) because he would have connected unlikely string of outcomes to short runs, instead of at Cha 1C (View chance as indicating complete absence of structure). Theresa seemed not to notice this difference. Instead, Theresa confirmed that Lee was right (line 8, "That's right. With experimental, we don't ever know."), without further questioning about Lee's reasoning. Theresa mentioned shortly about the number of trials right after her confirmation of Lee's idea ("So and that was only four spins") without further linking the number of trials to Lee's argument. This suggests that Theresa did not hold firmly in mind the instructional intention of the learning activity.

The post observation interview confirmed that Theresa did not strongly grasp the connection between variability and numbers of trials. Theresa explained why she reinforced Lee's idea, indicating that she thought that his idea was at a higher level of thinking than other students. Theresa said,

> …especially Lee in first period. He kept saying well, it could happen. You know he I would say was I think his thinking was probably on a little bit higher level than most of the kids … This space is huge and this space is tiny. But it still could happen. [Post Interview, March 2009]

She may have been right about her diagnosis about levels of students' responses, but she did not base her diagnosis on evidence of whether Lee had concluded anything about more certainty coming with more repetitions. This suggests that her way of interpreting the mathematical substance and levels of students' responses was not yet completely aligned with the classification system. The classification system suggests that theoretical and experimental probabilities are both estimates. Also, it suggests that although any possible outcome is uncertain, more stable estimates result from many trials of a repeated process. In contrast, Theresa seemed to have a broad goal of helping students make sense of data and chance with only occasional evidence of employing the classification system to interpret students' responses.

Summary of Theresa's Practice in Year 1. Theresa provided evidence that she developed a rough categorization of mathematical ideas. Her perspective seemed to be intuitively aligned with some portions of the constructs, especially making sense of the data and reasoning about the uncertainty of experimental results. In line with her development of understanding about disciplinary ideas in association with students' ways of expressing the disciplinary ideas, Theresa indicated her effort to use better questioning skills to support students' learning and her awareness of the need to support students' sense-making in Year 1. Curriculum materials seemed to support her initial step toward reform oriented practice: Learning activities and Thought-Revealing-Questions helped Theresa elicit big ideas of data displays, as she "stole" them to use. However, she mainly supported her students in sharing different levels of mathematical ideas and did not connect them toward higher levels of understanding.

Theresa's Practice in Year 2

In Year 2, Theresa demonstrated both her enactment of instructional activities in lessons and discussions of assessment items. Although the two forms of classroom interactions involved different elements of the assessment system, my analysis suggests some consistent patterns in Theresa's classification of students' reasoning and orchestration of classroom discussion. I illustrate both her enactment of instructional activities and an assessment item discussion in this section to examine the extent to which Theresa conducted construct-centered instruction.

Keeping a Rough Categorization of Mathematical Ideas & Approximate Instructional Intention of Learning Activities. Theresa enacted instructional activities from the lessons in a similar manner to Year 1: Theresa elicited students' mathematical ideas by using Thought-Revealing-Questions from lessons. However, she demonstrated that she did not further develop understanding of the instructional intention of the learning activities and sophisticated classification of students' reasoning. The classroom conversation about the same activity, *Teacher's Mystery Spinner (Figure 5)*, that she had enacted in Year 1 provides evidence of Theresa mainly discussing the Cha 3 level of performance (Quantify chance as probability) during the classroom discussion, rather than pushing students toward thinking about variability in outcomes in a small number of trials (Cha 4: Empirically examine the relationship between observations and all possible outcomes of repeated simple events). Although Theresa asked Thought-Revealing-Questions that were intended to promote discussion about variability, she did not discuss the idea explicitly with her students. The following episode happened right after Theresa provided the first four spins, which were blue, blue, yellow and blue.

Excerpt 4

1	T:	Why do you think it's A?
2	Kai:	Because it's got more blues particular.
3	S:	It's 75% blue.
4	T:	Okay. So right now I have I mean 75% blue.
5	T:	So are you guys confident it's A? ((looking at the lesson as she asks this question)) We can stop?
6	Ss:	No, no!
7	T:	Do we need to do some more?
8	S:	[Yeah.
9	S:	[It's just a guess.
10	T:	Okay, but that's a, that's a good prediction isn't it? Alright. So let’s do next four. Yellow. Blue. Yellow. Yellow ((writing on the whiteboard)).
11	S:	C.
12	S:	C.
13	T:	So now I have 1, 2, 3, 4 blues, 1, 2, 3, 4 yellows.
14		((Several students said B))
15	T:	So now it looks like B, So are you totally throwing A out now?
16	Ss:	No.
17	T:	No? Do you think we need some more?
18	S:	Oh?
19		((silent))
20	T:	Okay. ((writing eight more results on the whiteboard))

Theresa kept the uncertainty alive in conversation, which was the instructional intention of the activity, by following the direction of the lesson. However, she did not make explicit connections between the uncertainty and the short numbers of trials. The loose link seemed to be related to Theresa not holding firmly in mind the instructional intention of the learning activity. For example, from line 1 to 4, Theresa and the class discussed their prediction of Teacher’s mystery spinner based on calculation of probability (Cha 3C: Quantify probability as the ratio of the number of target outcomes to all possible outcomes). Then, in line 5, Theresa asked, “So are you guys confident it’s A?” which was a question that Theresa seemed to modify from a Thought-Revealing-Question in the lesson. The question was intended to support students to relate the number of repetitions to the soundness of the estimate of probability. In return, several students expressed that they were not confident with their prediction (line 6, “No, no!”) without justifying their uncertainty. Probing for students’ justifications might have supported students in connecting uncertainty and short numbers of trials. Instead, Theresa moved onto the next instructional step of providing more results. She also expressed her agreement with the prediction based on the calculation of probability in line 10 (“Okay, but that’s a, that’s a good prediction isn’t it?”). This countered her expression of uncertainty (lines 5 - 9).

Further evidence for the loose link between the uncertainty and the short numbers of trials can be related to Theresa not noticing mathematical ideas in students’ responses. An example can be seen in lines 10 to 15. When Theresa provided the next four results (line 10, “Yellow. Blue. Yellow. Yellow”), two students predicted that it would be Spinner C. It seemed that the two students only considered the four results without adding the previous four results. The two students seemed not to consider that more trials of a repeated process would help them make more stable estimates. However, Theresa seemed not to notice the mathematical significances of the students’ responses and did the mathematical work for the students to facilitate their prediction based on the calculation of probability (line 13, “So now I have 1, 2, 3, 4 blues, 1, 2, 3, 4 yellows”).

In Excerpt 5, it became more evident that Theresa did not fully understand the instructional intention of the activity or did not know how to orchestrate classroom discussion around the big idea. The episode happened right after Theresa provided 16 spins, of which 10 were blue and 6 were yellow.

Excerpt 5

1	T:	What do you think?
2	S:	A.
3	T:	A B or C?
4	Ss:	A.
5	T:	Well is there one that we can for sure throw out?
6	Kai:	Yes. C.
7	Ss:	C.
8	T:	Why can we throw C out?
9	S:	Because there're not way too many yellows.
10	T:	So you would expect if the answer were C then we would have a lot more yellows than that right?
11	Kai:	Yes.
12	S:	Yes.
13	T:	I agree I think you're right. I don't think it could be C.

The first piece of evidence of not holding the instructional intention of the activity firmly in mind is from her question. Theresa initiated a question that ignored the critical attribute of chance, variability. She asked (line 5), “Well is there one that we can for sure throw out?” Students were confident in saying C. Theresa evaluated that the students were right in line 13, “I agree I think you’re right. I don’t think it could be C.” Rather than asking students to reason about the relationship between more trials of a repeated process and more stable estimates, she proceeded to frame the chance event as definite (“for sure”). By doing so, Theresa focused on the conversation around the calculation of probability.

In the conversation that followed from Excerpt 5, some students talked about the possibility of having four yellows in the next four spins, therefore making Spinner B a possibility. Theresa could have concluded the conversation by keeping the uncertainty alive or by elaborating further on students' suggestions (next four spins would help them be more certain). However, she decided to wrap up the conversation by saying, "So, but what do you think, just based on those 16?" directing students toward calculating an empirical probability.

In sum, the analysis of Theresa's enactment of *Teacher's Mystery Spinner* provides evidence that Theresa did not deepen her understanding of mathematical ideas. It is conjectured that Theresa's institutional context limited her further development in mathematical understanding in Year 2. Theresa missed the workshop on chance in Year 2, which might have impacted how she used the chance lesson in Year 2. At this workshop, the big idea of probability as reflecting structure in repeated trials was emphasized by using Tinkerplots™ (a data analysis tool) and examining sampling distributions. The chance unit was revised to make the "law of large numbers" visible for different sizes of samples. However, Theresa did not have the opportunity to explore these ideas, and hence used the original lesson when she taught Chance in Year 2.

Providing Explanations of Mathematical Concepts by Herself. The analysis of Theresa's assessment discussion showed that Theresa supported students in making sense of a distribution of the data (DaD 3A: Notice or construct groups of similar values from distinct values and CoS 1A: Use visual qualities of the data to summarize the distribution). It suggests that Theresa incorporated some big ideas of data and display that were shared through the workshop and the assessment system. However, she was not able to come up with productive instructional moves to support students in progressing rudimentary levels of understanding (e.g., CoS 1A: Use visual qualities of the data to summarize the distribution and CoS 2B: Calculate statistics indicating variability) to the target performance of the assessment item (CoS 3F: Choose/Evaluate statistic by considering qualities of one or more samples). Theresa tried hard not to “give them [students] the answer,” as she mentioned in her interview in Year 1. However, in Year 2, she explained mathematical concepts by herself when students did not provide the right answers to her questions. Theresa demonstrated the practice of explaining mathematical concepts when she discussed an assessment item, *Range* (See Figure 6).

Range

The range measures the spread of data. It is the difference between the smallest value and the largest value. For example, if the values of the data were 12, 13, 19, 22, the range would be 10 (22-12 = 10).

Carl says that the range is **always** a good measure of spread.

Do you *agree* or *disagree* with Carl? Choose **(a)** or **(b)** and explain why.

(a) I **agree** because ______________________________

____

____

____

(b) I **disagree** because ______________________________

____

____

____

The assessment item, *Range*, asks students whether the range is always a good measure of spread and why. Theoretically, the item is designed to elicit several levels of performances including CoS 2B and CoS 3F. One anticipated response from students who are at CoS 2B level of understanding is that the range would always be a good measure of spread based on the matching between their calculation of the spread and the range given in the assessment item. Students who are at CoS 3F level of understanding would say that they would disagree by considering a sample distribution that has an outlier.

Figure 6. Description of Range & related levels of performances.

When Theresa was asked during the post instruction interview what she learned when she glanced through students' responses, she stated:

> Well, I think the main thing with, well with the range problem there, you know almost all the kids said they agree, and if they said disagree then it was kind of the answer didn't usually make sense. You know, I don't think they really, they're not thinkin' about those outliers and how that changes your data. [Post Interview, December 2009]

She summarized what her students lacked ("they're not thinkin' about those outliers and how that changes your data"). Based on her diagnosis, Theresa decided to present a distribution (Figure 7) that made visible an outlier that would affect the magnitude of range tremendously.

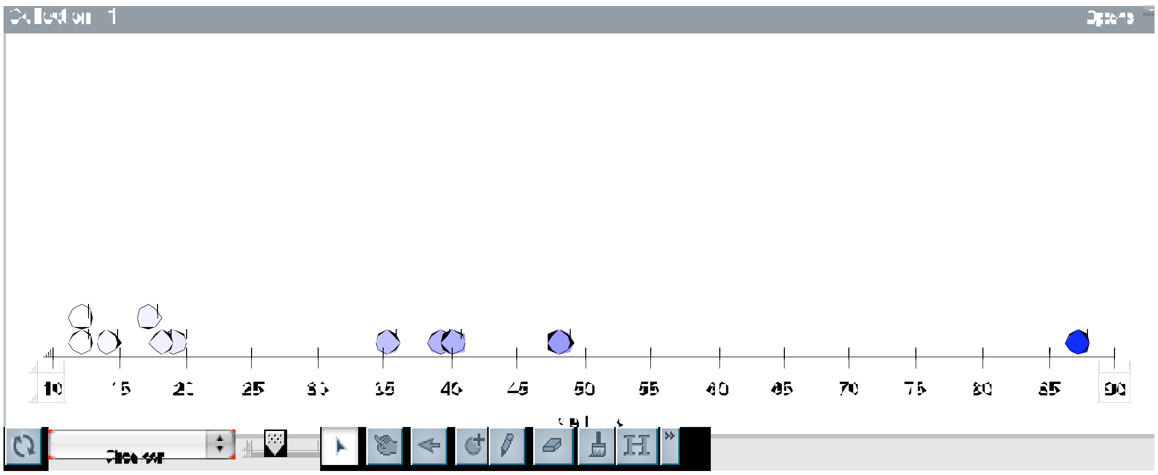


Figure 7. A distribution with a big outlier

When Theresa presented the distribution with a big outlier (Figure 7) to her students, it revealed that her students had misconceptions about the meaning of range. Students were able to calculate the range of the distribution, which was 75 (CoS 2B). When Theresa asked students to show the range on the distribution ("Can you come show us on our line plot where 75 is?), Kai pointed at 75 (the data point on the X- axis). This suggests that he considered the range to be a data point, not a distance between the lowest and highest data points. In addition, Kai and several other students seemed to consider the range as a reference point to decide a middle clump. Kai said, "most of the numbers are from here ((10)) to here ((75))." Theresa did not catch Kai's misconception about range and moved on, suggesting that she did not notice the mathematical significance of Kai pointing at a data point (75), not the distance.

Without addressing the misconceptions or helping students understand the meaning of range, Theresa talked about attributes (i.e., a clump and outlier) of the distribution (Figure 7) (CoS 1A and DaD 3A) and calculated the range (CoS 2B) of the distribution. Then she asked students about the effect of the outlier on range (aiming at CoS 3D: Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components). The instructional trajectory was not aligned well with her students' current state of understanding. Her students had only a shaky understanding about the meaning of range. Without addressing this, Theresa moved on to asking about how the outlier affected the range (CoS 3D). The students were not able to relate the effect of the outlier to the range since most of them understood range as a point. In Excerpt 6, Theresa used more traditional teaching practices, focusing on having students memorize mathematical terms and providing an answer to a question.

Excerpt 6

1	T:	Just one number. Is that 87 important? Is it having an effect on our data?
2	S:	Yeah.
3	T:	On our range?
4	S:	Yes.
5	T:	How is affecting our range?
6	S:	Because that is the biggest number that we've got isn't it? So.
7	T:	Hmm hmm, it's a big number. If we look at our data, how does 87 compare?
8	S:	An outlier.
9	T:	Thank you. It's a what?
10	S:	Outlier.
11	T:	You guys know that word. What?
12	S:	Outlier.
13	T:	It's an outlier. What's an outlier mean?
14	S:	It means it's just way off
15	T:	Way out there like a mistake, seems like. Outliers aren't always mistakes, but when we see one number that's way out there, a long ways away from the bulk of our data, where would you say the bulk of our data?
16	S:	It's between 12 and 29, 18 or 19, 12 and 19.
17	T:	Well, we have
18	S:	20.
19	T:	I would say that, that's kind of the center clump of our data, isn't it? Cause that's kinda, there's a bunch of numbers clumped right there together, but we have several out here, too.
20	S:	It might be 50 through 12.
21	T:	So, we could say 12 or 10 to, to 50, whatever that number was. I think it was like 47 or something. So, that's actually where most of our data is, isn't it? It's between those two numbers. And then the range of 75 that kind of gives us kind of a misconception about the spread doesn't it? Cause it makes it seem like the numbers are spread way apart, when are they really? No. No, it's because of that 87, isn't it? That outlier is messing everything up, isn't it?

Theresa initiated the interaction with a question targeting CoS 3D, but her follow-up questions in response to students did not build intermediate steps toward the target performance (CoS 3D). When Theresa asked how 87 affected the range, a student described the characteristic of 87 not the effect of 87 (line 6, "because that is the biggest number that we've got isn't it?"). In return, Theresa shifted the discussion to CoS 1A and DaD 3A (lines from 6 to 20). For example, Theresa asked about the highest number (line

7, “If we look at our data, how does 87 compare?”) and the center clump (line 15, “… where would you say the bulk of our data?”). This shift lasted from line 6 to line 20, comprising most part of the conversation.

Excerpt 6 also illustrates Theresa’s emphasis on making sure that students memorized a mathematical term, outlier. When a student said outlier (line 8), Theresa expressed that it was what she had been waiting to hear (line 9, “Thank you.”). Then Theresa asked students to repeat the word outlier several times (line 9, “It’s a what?” and line 11, “You guys know that word. What?”), suggesting that she wanted to make sure that students knew the mathematical term that often appeared on the state standardized test.

In contrast, conversation around CoS 3D was not elaborated further. Instead, Theresa wrapped up the conversation by telling students the answer. In line 21, Theresa said, “And then the range of 75 that kind of gives us kind of a misconception about the spread doesn’t it? Cause it makes it seem like the numbers are spread way apart, when are they really? No.”

As the discussion continued, Theresa asked a question (“Is this i- in this case with this line plot, is range a good measure of spread?”), targeting the CoS 3F level of understanding. The class discussion again was centered on CoS 1A and DaD 3A. Then, Theresa provided a long explanation to students:

> So, do you guys see how i-, when you have outliers, the range isn’t necessarily always a good measure of spread because that outlier makes it seem like your numbers are spread way way ((opening her arms very wide)) out when really they’re not, are they? They’re all about right here ((pointing at 12 to 50)). They’re kinda, kinda grouped together with our largest clump being where?

She seemed to have a hard time coming up with intermediate questions that could guide her students toward the target performance. As a result, she ended up explaining the mathematical concept.

Summary of Theresa's Practice in Year 2. The description of Theresa's classroom interactions and interview excerpts in Year 2 suggests that her development of understanding about disciplinary ideas in association with students' ways of expressing the disciplinary ideas in Year 1 was not further refined in Year 2. Although Theresa provided evidence that she promoted some sense-making of data display during the assessment talk about *Range*, she appeared to be more guided by the needs of the statewide assessment that students know particular skills and pieces (e.g., recitation of outliers and calculation of probability). Rather than building on current states of students' understanding, Theresa demonstrated a traditional form of assessment review discourse, explaining how to get a right answer.

Summary of Theresa's Naturalization Process of the Assessment System

Theresa evidenced some progress in making changes in her perspective and practice in using the elements of the assessment system during Year 1. During the course of the first year, she exhibited increasing alignment with the mathematical disciplinary perspective suggested by the constructs and lessons and enacted that perspective during classroom discussions. Observations and interviews in Year 1 indicated that she focused on learning performances in the construct maps and supported her students' progress toward higher learning performances by using the Thought-Revealing-Questions found in the lessons. She supported students' sense-making of mathematical ideas by a combination of a turn-taking and a transformed I-R-E discourse pattern that was initiated by open-ended questions to elicit student thinking.

However, the promising changes made in Year 1 were not further refined in Year 2. Instead, in Year 2, she seemed to align more with a traditional school mathematics perspective, focusing on teaching performances that were often tested in the state standardized exam and not identifying mathematical ideas from students' responses. Theresa made sure that students memorized mathematical terminologies (e.g., outliers, median) and explained mathematical concepts by herself. In addition, Theresa did not further accomplish using the curriculum materials in coordination with the classification system. In her interview conducted in year 2, Theresa said she rarely referred to the classification system as she planned her lessons or assessment discussions. She demonstrated that she kept the rough categorization of mathematical ideas in students' response and was not able to make instructional moves that were tailored to students' current states of understanding.

The declined progress in changes in Theresa's perspective and practice may be attributed to less attendance to the workshops and to the limited opportunities to explore the assessment system because her new role demanded less practice and more supervision. As a result, she did not establish a routine (e.g., teach with a lesson, assess student learning by assessment items, score students' responses by scoring exemplars, and plan assessment conversations to support student learning) through which she might be able to develop a mathematical disciplinary perspective and improve her teaching practice.

Rana: Illustrating Learning Progression in Action

Rana began participating in the study as a first-year teacher. She had an undergraduate degree in mathematics and a master's degree in teaching from a state university.

As I previously mentioned, Rana's school participated in a school improvement program initiated by the state department of education. Accordingly, the curriculum she used emphasized practicing calculations. Rana's school coordinated logistics to support the calculation-oriented curriculum: 45 minutes of instruction seemed to be too short for students to engage in deep mathematical thinking. Also, problem worksheets were designed in a way that there was no space for students to express their reasoning; fitting four pages to one, two pages on one side and two pages on the back. Rana identified the composition of her classes as another challenge. She reported that there were many students classified as ESL and special education in her classes, which she attributed as a barrier to conducting classroom discussion.

Like other teachers at the workshop, Rana was always under the pressure of the state standardized test. Every two weeks, Rana's school administered two-day tests to help students prepare for the benchmark tests, which were administered every nine weeks. Rana reported that teachers at her school were always short of time for teaching, because they lost instructional time for the practice tests and because the state test was scheduled in early April, but they still had to cover the state framework for the entire school year. Therefore, teachers had to have their entire curriculum taught in seventy five percent of the time outlined in the framework.

It seemed that Rana's school was under a heavier pressure of the state standardized test in Year 2 because the school continued participating in a school improvement program. In Year 1, Rana and her school team attended all workshop sessions together. However, in Year 2, they attended sessions that discussed mathematical strands that were directly related to the state standards, such as measurement, measures of centers, and chance. In Year 1, I was able to visit her classroom four times, but in Year 2, Rana often canceled her scheduled observations because of conflicts with school events such as benchmark tests. In the study's second year, I was only able to visit Rana's classroom twice, once in the beginning of the school year (October 2009), and once after the state standardized test (April 2010).

As I mentioned previously, for Rana and for the other two case-study teachers, I was able to observe their formative assessment practices—their use of the assessment system to instigate changes in student conceptions of the mathematical ideas targeted by the learning progression.

Rana's Practice in Year 1

Categorizing Student Answers as Right vs. Wrong. Rana's assessment practice in Year 1 illustrated categorizing students' responses into broad bins, such as right and wrong, without paying close attention to the mathematical ideas that guided students to arrive at their answers. In dealing with a wrong answer, Rana tended not to go into detail about how students solved a problem in order to understand how they reasoned about mathematical concepts. Rather, she tended to simply point out why an answer was wrong. In this excerpt, from the first year of her participation in the study (March 2009), students were asked to find the mean of seven measurements (42, 46, 45, 47, 43, 46, 46). A student, Justice, said he got 38.4, while other students said their answer was 45.

Excerpt 7

1	T:	Okay, I've got 45, 38.5 [38.4], anything else? Okay, how did you, okay, now what about 38.5 [38.4]? Could 38.5 [38.4] possibly be the mean?
2	Ss:	No. ((Several students answered))
3	Justice:	Maybe.
4	T:	Is there, is there anything in the 30's in that group of numbers?
5	Ss:	No. ((Several students answered))
6	T:	So, does the mean have to be in that group of numbers or close?
7	S1:	Yeah.
8	S2:	Yes.
9	S3:	Close to it.
10	T:	Okay.

Rana tried to point out how Justice could have checked on his answer by asking a series of questions in relation to an important attribute of the mean (i.e., central tendency) instead of finding out how he thought about the problem. In line 1, Rana problematized Justice's answer by asking, "Could 38.5 [38.4] possibly be the mean?" Justice's response, "Maybe (Line 3)," indicated that he did not appear to know what Rana was asking him to consider. Next, Rana directed students to inspect the measurements given to see if there were any numbers in the 30s by asking, "Is there, is there anything in the 30's in that group of numbers? (line 4)" Rana then proceeded to ask another question, "So, does the mean have to be in that group of numbers or close?" pointing out where the mean should be located in relation to the distribution of the given numbers.

This series of questions seemed to support sense making of the mean in relation to the data. Rana reflected on her instructional move in the post observation interview. She stated:

> For the mean they got 38.5 [38.4] and they're just wrong. I tried to kind of ask why could that not be because it's not even close to the answer … So I guess just focusing on it being central tendency and not something outside ... But I didn't talk about it too much. [Post Interview, March 2009]

Rana pointed out how Justice could have checked his answer ("I tried to kind of ask why could that not be ...") by considering central tendency. The series of questions Rana asked were related to an important attribute of the mean and a sense-making of mean as a measure, as she indicated, "focusing on it being central tendency."

However, Rana tended to generate a series of questions that students could answer simply by saying yes or no (lines 2, 3, 5, 7, 8, and 9), thus hiding the main concepts and lowering the cognitive demand for students (M. K. Stein, Grover, & Henningsen, 1996). Furthermore, the questions were not based on students' current state of understanding and did not reveal students' reasoning. In this interaction, Rana immediately directed students toward checking the correctness of Justice's answer, before she tried to understand how he thought about the problem.

Rana characterized her initial assessment practice as "categorizing student answers as right vs. wrong." Rana primarily talked about "right" answers rather than "wrong" answers as being a useful focus for instruction as she reflected on her first year of teaching. She stated:

> I didn't think that you really needed to talk about wrong answers because we don't want wrong answers, so we don't wanna talk about `em…When I see that their answer's wrong, I just go like, 'Okay, that kid's not gonna talk today.' [Post Interview, April 2010]

This excerpt indicates that Rana was oriented toward categorizing students' responses as "right and wrong." Also, Rana said she did not ask students who responded incorrectly to explain their thinking in detail because she believed that wrong answers did not contribute to the learning of the whole class.

Presenting Strategies to Get a Right Answer by Asking Content-General-Questions. In the lesson she taught in March 2009, Rana repeated the pattern of asking students to present different strategies without any follow-ups in assessment item talk. For example, the item, *Height of a Plant*, is designed to elicit a range of responses that allow teachers to make important distinctions in students' reasoning about measures of center in relation to qualities of distribution. The data set includes an outlier (66), and it makes the mean not be located in the center clump.

Height of a Plant

Fifteen students measured the **height** of a plant in **centimeters.**

Their measurements were:

27, 25, 21, 23, 66, 23, 30, 28, 26, 30, 31, 24, 26, 26, 23

1. What is most likely to be the actual **height** of the plant?

_________________cm.

2. Why did you choose that value? Show your work:

Students who calculate measure of center without considering qualities of the distribution are scored as CoS 2A (Calculate statistics indicating central tendency). For example, students may calculate the mean without considering the outlier. Students who first remove the outlier from the data, then calculate the mean of the remaining 14 data points knowing the effect of the outlier are scored as CoS 3D (Predict how a statistic is affected by changes in its components). Students who list several different ways to find the actual height of the plant and choose a statistic that is less affected by the outlier (for instance, the median) are scored as CoS 3F (Choose/Evaluate statistic by considering qualities of one or more samples). This highest level of performance is expected to be rare. The diversity in students' thinking can be orchestrated to step students up to CoS 3F by teacher's intentional comparison and contrast of different students' strategies, which makes this item as a good formative assessment item.

Figure 8. Description of Height of a Plant & related levels of performances.

Rana tended to share different strategies to get a right answer by asking content-general-questions, suggesting that she focused on getting it right rather than pushing students toward higher levels of mathematical ideas in different strategies, as illustrated in Excerpt 8. Students had been solving the assessment item, *Height of a Plant* (Figure 8) working in small groups. While students were solving the problem, Rana had talked with some groups of students and asked each team what they had decided as an answer and how they solved the problem. When Rana called the class together to discuss the assessment item, she solicited students' responses by asking "what did you decide was the actual height?" Many students simultaneously volunteered their answers. The excerpt starts when Rana was writing down the answers that students provided.

Excerpt 8

1	T:	We have, okay, I heard 23. I heard 24. 29. 26. 25. 24.5. Okay. So, 23. Tell us why you chose 23.
2	S:	We chose 23 because it had the most out of all of them.
3	T:	Okay. So, you chose 23 because it had the most. ((intercom announcement)) What about 24? Alright. So 24. Who chose 24?
4	S1:	I did.
5	T:	Okay. How did you guys get 24?
6	S1:	I got, I did the mean, but I crossed out 66 cause it's way off.
7	T:	Okay. So you did the mean, but you took off 66. Why'd you take off 66?
8	S1:	[Because it's way off.
9	S2:	[Way off.
10	T:	Those way off? So you don't think the 66 is a valid measurement?
11	S2:	No.
12	T:	So, what would've happened if you left 66 in?
13	S2:	It would've been [higher.
14	S1:	[About 28.6.
15	T:	Okay. So, then your mean would've been 28.6, and you think the only reason it would be that high is because that 66 is in there? Okay. So, do you guys understand what Marco and Troy did? They found the mean, but they >kicked out< 66 cause it was way off. So, did you kick out any other numbers?
16	Ss:	No.
17	T:	That was the only one you kicked out? Okay.

In this interaction, Rana asked several content-general questions: "Tell us why you chose 23 (line 1)" "How did you guys get 24? (line 5)" and "Why'd you take off 66? (line 7)" These questions seemed to be productive to estimate where students' levels of understanding were by eliciting the reasoning behind their answers. For example, the content-general question, “How did you guys get 24? (line 5),” revealed that a group of students calculated the mean without including 66 (line 6). The next instructional moves after the content general questions illustrate that Rana did not build on students’ responses toward higher levels of thinking, suggesting that she might not know the prospective learning progression.

First, the question “Why’d you take off 66? (line 7)” drew out important mathematical ideas from the students (“Way off”: CoS 1A). Students replied because it was “way off (line 8 and 9),” indicating that they noticed an important quality of the distribution (CoS 1A). This mathematical substance could be extended to explore more qualities about the distribution of the data set (e.g., most of the measurements are in the 20’s), which would set up the class to discuss locations of measures of centers in relation to the distribution. Rana’s follow-up question focused on the outlier as a specific point (“So you don’t think the 66 is a valid measurement?”), but did not link the outlier to the distribution of the data. This would obscure the important mathematical idea that 66 was away from the clump, where most of the measurements were.

Second, Rana asked a level-specific question (CoS 2A, “So, what would’ve happened if you left 66 in?”). This might have created an opportunity for students to compare the two means (with and without the outlier) and to think about the better choice for the best guess of the height of the plant, which might have led to CoS 3F (Choose/Evaluate statistic by considering qualities of one or more samples). However, Rana did not discuss the change in the mean in relation to the qualities of the distribution. In the remaining class, Rana elicited two additional responses from different groups. One group calculated the median, and the other calculated the median of the two modes (23 and 26). The diversity could be orchestrated to step students up to CoS 3F (Choose/Evaluate statistic by considering qualities of one or more samples). For example, the class can compare the mean to the median or the mean without 66 to the median. This comparison will make visible that the mean without 66 would be a better choice to find out an actual height of the plant because the mean without 66 would be located in the clump and very similar to the median.

Rana's post interview suggested that she did not intend to extend classroom discussion toward higher levels of performances:

> Throw out any outliers and then find the mean. But some of them chose the median; some of them chose the mode. I didn't really make a conclusion like one's better than the other. Cause they're all good in different situations but if they could justify themselves and say how they did it then it's legitimate. [Post Interview, March 2009]

Rana noticed the diversity in student thinking, but she did not focus on specific mathematical substance, instead, she focused on general verbal performance ("if they could justify themselves and say how they did it then it's legitimate"). It is true that all measures of center are useful in different situations. However, it is a valuable learning opportunity to discuss which method is the best choice in the particular distribution given, which can move students toward understanding at a CoS3F level. Rana might not know differences in students' strategies in terms of sophistication in conceptual understanding.

Asking a Series of Questions that Illustrate Procedural Steps to Get a Right Answer. In line with her interest in obtaining the correct answer, Rana focused on eliciting procedures to get the right answer when a class reviewed assessment items. Key mathematical ideas tended to be hidden in the step-by-step procedures.

An example that illustrates Rana's practice of eliciting a description of procedural steps can be seen in *Kayla's Project* assessment talk. Students were asked to estimate one data point when given with the mean and other data values.

Kayla's project

Kayla completes **four projects** for her social studies class. Each is worth **20 points.**

Kayla's Projects – Points Earned	
Project 1	16 points
Project 2	18 points
Project 3	15 points
Project 4	???

The mean score Kayla received for **all four** projects was **17.**

Use this information to find the **number of points** Kayla received on **Project 4.** Show your work.

Students' responses on these items can be mapped to CoS 2A (Calculate statistics indicating central tendency) as the lowest level and CoS 3D (Demonstrate knowledge of relations among its components) as the highest level. Students can use a guess and check strategy, randomly plugging in numbers and calculating the mean repeatedly (CoS 2A). However, they can also use an understanding of a mean as a fair share, which involves knowing that the mean multiplied by the number of measurements equals the sum of all measurements (CoS 3D). Or, students can use a deviation score approach, comparing the distance of each value to the mean of 17 (CoS 3D).

Figure 9. Description of Kayla's Project & related levels of performances.

When Rana called the class together to discuss the assessment item, Fresco told her, "I know this one." Thus, Rana decided to begin the discussion with him. The interaction in Excerpt 9 consists of Rana's content-general questions and Fresco's responses to reconstruct the step-by-step procedures he performed to arrive at his answer.

Rana mostly asked Fresco what he did, but did not ask him to justify his procedures. She asked “what” questions five times (lines 4, 6, 8, 12, and 16) and asked “why” question just one time (line 2).

Excerpt 9

1	Fresco:	I did like 17 times four, and then I got 68, then I, then I added all of these up ((pointing at the given measurements)), and I counted up to 68 ( )
2	T:	All right. Hold on, hold on. You’re going a little fast. Let’s slow down. So, tell me again. So, you did 17 times four. ((writing 17 x 4 on the transparency sheet)) Why 17 times four?
3	Fresco:	Because it equaled the mean like ( ) ((showing four fingers)).
4	T:	Okay. So, this is our mean ((writing mean under 17)). And this is what? ((pointing at 4))
5	Fresco:	How many numbers.
6	T:	The number of projects ((writing # of proj under 4)). Okay. So, then, after you did that, what did you get?
7	Fresco:	I got 68.
8	T:	Okay, and what’d you do with that 68?
9	Fresco:	Well, I just wrote it down.
10	T:	Okay. So, you just wrote it down. Got ya.
11	Fresco:	Then, and then I added like 16, and 18, and 15.
12	T:	So, you added 16, plus 18, plus 15, and what’d you get there?
13	Fresco:	I got, I don't know.
14	T:	((adding 16, 18, and 15 and writing 49))
15	Fresco:	Ya. Forty nine.
16	T:	Some number? 49? Okay. So, what’d you do with those numbers? How, how’d you get 19 as your answer?
17	Fresco:	I just counted up until I got 68.
18	T:	Okay. So, you figured out the difference between 49 and 68, and you figured out you would have to add 19 to get to 68? So, that would be your answer?

To make Fresco’s procedures visible to the other students, Rana notated Fresco’s procedures on the overhead transparency (Figure 10) as she elicited steps from Fresco. Rana broke up the procedures into three big steps, following Fresco’s idea (“I did like 17 times four, and then I got 68, then I, then I added all of these up and I counted up to 68," line 1). First, Rana related the numbers in Fresco's response (17 x 4) to the numbers in the problem. In doing so, Rana asked Fresco, "Why 17 times four? (line 2)" and "this is what? ((pointing at 4)) (line 4)" When Fresco clarified the numbers (lines 3 and 5), Rana asked Fresco what he got as an answer for the equation, "after you did that, what did you get?" (line 6). Next, Rana illustrated what Fresco did with the given measurements. Fresco described his procedure in line 11, "Then, and then I added like 16, and 18, and 15." When Rana asked what he got by adding the three numbers, Fresco did not know the sum of the given measurements ("I got, I don't know," line 13). Hence, it was unclear how Fresco moved onto the next procedure without knowing the sum of the given measurements. Instead of interrogating Fresco about how he got his answer, Rana calculated the answer for him (line 14). Finally, Rana elicited the last procedural step by asking, "What'd you do with those numbers? How, how'd you get 19 as your answer?" (line 16). Fresco seemed to use count up strategy to find the answer 19.

The class concentrated on the procedure, but they did not discuss relationships among components (CoS 3D). The relationship among components (CoS 3D) can be expressed as 16 + 18 + 15 + ❑ = 17 x 4, emphasizing the relationship between the sum of the measures and mean multiplied by the numbers of the measures. The relationship seemed invisible during the conversation and in the representation that Rana constructed.

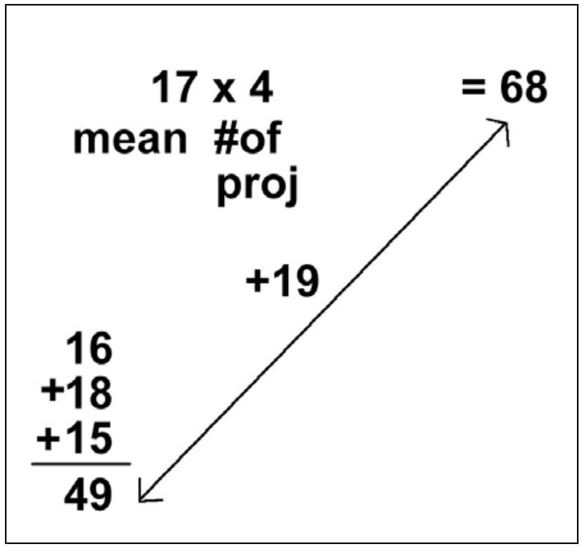


Figure 10. Facsimile of Rana's inscription on the overhead transparency.

Instead, Rana's notation on the transparency seemed to represent the exact procedures that Fresco took to get the right answer.

Summary of Rana's Practice in Year 1. The analysis shows that Rana demonstrated a hybrid of focusing "getting it right" and helping students experience mathematics as a form of sense-making. According to her reflection, Rana did not elicit how students who had a wrong answer thought about a mathematical concept because she believed it would not support learning for the whole class, suggesting that she viewed students' responses from a dichotomous perspective, right and wrong, at the outset of the study. Rana drew out strategies or procedural steps to get a right answer mainly by using content-general questions such as "what did you do?" "how did you do it?" or "why did you do that?" However, the analysis of classroom discussion also indicates that Rana asked some questions that directed students to make sense of the distribution of data in relation to the mean.

Rana's Practice in Year 2

Categorizing Student Thinking in light of Construct Maps or Scoring Exemplars. Rana showed some changes in her dichotomous perspective on students' responses in the second year. Instead of right or wrong, she identified mathematical ideas and levels of students' responses and elicited all levels of reasoning in discussion, including wrong answers and right answers arrived at via unconventional reasoning.

Rana was discussing *Two Spinners*[5] (Figure 2) assessing students' understanding of the probability of a compound event with the class. Excerpt 10 illustrates Rana's noticing of particular levels of students' responses in terms of the scoring exemplar and unpacking students' logic behind their answers. The first student's, Elena's, response was scored as "No Link," which means that her response was unrelated to mathematical ideas about probability. Elena used two numbers that were from the question and did not base her reasoning on the structure of either one spinner or two spinners. The second student, Leon's response was scored as Cha 3C, treating the two spinners as a simple event.

[5] See p.37 for more information about the item including possible outcome performances.

Excerpt 10

1	T:	Elena said the probability is one half? And she said explain why you chose this answer because there are two spinners and you get one prize. So what do you mean by that?
2	Elena:	It says two spinners for one prize.
3	T:	Okay. So we got the two ((pointing at Elena's writing)) from two spinners ((pointing at "two" in the problem text)) okay and the one from? It's just one from one prize. You saw one and two?
4	Elena:	((nodding her head))
5	T:	Does that make sense- Is that what you are saying? I am just making sure (that's what you're saying). Okay. Okay? Does anybody have any questions about Elena's answer?
6	Students:	No.
7	T:	Okay. Let's look at this one. Okay. Leon's. Wha Why Can you explain your answer to us?
8	Leon:	There are only two sides. Only one has gray so.
9	T:	What do you mean by there's only two sides?
10	Leon:	Like gray and white.
11	T:	Okay. So there we have gray and white so gray is one out of the two colors?
12	Leon:	Yes.

Here, Rana focused on asking students to explain their thinking (line 1: "explain why you chose this answer because there are two spinners and you get one prize. So what do you mean by that? line 7: "Can you explain your answer to us? and line 9: What do you mean by there's only two sides?"). These instructional moves (lines, 1, 7, and 9) are significantly different from how Rana started talking about Justice's wrong answer in the previous section. Here, she began the discussion by eliciting students' logic behind their responses, rather than by telling them why their responses were flawed.

Rana also made an instructional move directed at the other students - making visible students' reasoning by coordinating talk and gesture. She provided an elaborated explanation of Elena's response, relating where numbers in Elena's response came from the problem (line 3) and pointing at Elena's writing. Also, Rana asked Leon to elaborate what he meant by “two sides” (line 9), which was an important distinction to make before sharing more advanced ways of thinking about the item. Rana reflected on her choice to present Elena’s response in her post-instruction interview:

> I try to if they are completely off base like one girl that said there's two spinners and you get one prize I put her up there cause I wanted for the people to see that even though there's numbers in the question doesn't necessarily mean that you use them. [Post Interview, April 2010]

Rana had a clear understanding of what Elena did, suggesting that Rana started focusing on her student thinking. Also, it was clear that Rana was thinking of using Elena’s response for other students (“I wanted for the people to see”). This is a significant change from Year 1 where Rana indicated that she preferred not to talk about “wrong answers.”

The catalyst for Rana to become interested in the mathematical substance of students’ thinking beyond right or wrong was her attendance to the workshops and her use of the assessment system. She reflected on what helped her make the change.

> After I'd seen him [Rich Lehrer] explain to us his materials and then when I actually came back and used his materials it kinda, it just kinda all made sense because I saw all these things that the kids were doing that he [Rich Lehrer] said they would be doing and showing what level they're on and you know him saying well ask this question to get them to move, to understand this better. [Post Interview, October 2009]

The change seemed to be mediated first by talking about different ways that students might think about mathematical ideas at the workshop. Furthermore, when Rana used lessons and assessment items in her classrooms, she noticed that her students responded similarly to the ways presented and discussed at the workshop.

> So I don't know that I would've really noticed the, what he is lacking as much, I would have just rather missed it all, so I think that my knowledge [*that I learned*] … is helping me kinda more diagnose their specific shortcomings instead of just kinda saying, oh you don't get it at all. [Post Interview, October 2009, Italic added]

Rana explained how her diagnosis of student thinking became focused on mathematical understanding ("more diagnose their specific shortcoming") and not just on identifying and discarding wrong answers ("missed it all" and "you don't get it at all"). She also indicated that the construct map and scoring exemplar helped her identify different levels of students' thinking.

> When I'm giving the items, I look at the exemplars before and then have `em [scoring exemplars] out on a table, I mean, the whole time I was going back and looking at the exemplar to kind of see what to expect… It helps me put `em in order when we're gonna share, too. [Post Interview, April 2010]

In particular, the scoring exemplar supported Rana in anticipating what her students' responses would look like on assessment items and in ordering her students' responses during assessment talk.

Approximating Highlighting and Juxtaposing Practices. The practice of highlighting and juxtaposing consists of putting side-by-side student responses at different levels of sophistication according to the classification system and making visible distinctions in students' reasoning. In Year 1, Rana rarely compared students' mathematical ideas. In contrast, in Year 2, she enacted approximations of the practices of highlighting and juxtaposing mathematical ideas. Rana demonstrated several instances of juxtaposing different students' responses to make differences apparent. This transformation suggests that her noticing of levels of thinking about mathematical ideas influenced how she structured classroom interactions.

Excerpt 11 illustrates how Rana highlighted one student's way of thinking in order to build on it toward higher levels of thinking following the scoring exemplar. The class was talking about the assessment item, *Two Spinners* (Figure 2). In Excerpt 10 which illustrates classroom interaction right before Excerpt 11, the class talked about

Elena’s (getting random numbers from the question text) and Leon’s ways of thinking (only looking at one spinner and estimating the probability as ½). Ken considered both spinners but treated the compound event as a simple event, counting the four sections of the two spinners as possible outcomes and the two gray sections as target outcomes.

Excerpt 11

1	T:	Okay. What about this one? Ken, this one is yours. He says because out of the four possibilities, what are the four possibilities? What do you mean by four possibilities?
2	Monique:	Mine is like his. ((pointing at the board))
3	T:	WHAT what do you mean by four possibilities?
4	Ken:	Um.
5	Leon	Oh. I know what he [means.
6	Ken:	[Like. Just the
7	T:	So I need to move it down so that you can see the spinners. So you said because out of the four possibilities what do you mean by four possibilities?
8	Ken:	Um. Cause of like A, B, C, D I guess?
9	T:	Okay. Cause like we've got gray white gray white ((pointing at each section of the two spinners)). So one two three four ((pointing at each section of the spinners)).
10	Ken:	Hmm hmm.
11	T:	Okay. And then you say that two of them are gray. So two ((pointing at gray sections on the spinners)) out of one two three four are gray. Okay. Is that making sense?
12	Ss:	Hmm hmm ((Several students are nodding)).
13	T:	I mean do you see that? Leon was talking about he said there were two possibilities but Ken is counting this is one this is two this is three and this is four ((pointing at each section of the two spinners)). Does that make sense? Okay. So we have two fourths one half. Okay.

In this excerpt, Rana attempted to highlight Ken’s way of thinking, which would help a transition toward an important mathematical idea, generating outcome spaces of a compound event. Ken wrote on his test, “Because out of the four possibilities to spin, two of them are gray and 2 are white equaling 2/4 = ½.” Rana highlighted Ken’s notion of "four possibilities" (Line 1, 3 & 7) by asking him to elaborate on it. She made further attempts to make Ken's way of thinking visible from lines 8 to 11. Ken explained four possibilities by assigning letters to them (A, B, C, and D), presumably referring to the four sections of the two spinners (line 8) but was not explicit. Rana helped students see that Ken was referring to the four portions of the spinners by rephrasing ABCD as "one two three four" as she pointed at the spinner sections in the picture (line 9). Next, Rana pointed to the gray sections to help students see where Ken got two (line 11). The elaborated illustration of Ken's method would be helpful when the class engaged in comparing his method of getting four (by looking at the two spinners) with a different way of getting 4 (by enacting the event). The comparison seemed to make a method for generating sample spaces visible to students.

Here, Rana also juxtaposed students' ideas in line 13. Rana imported Leon's response (only looking at one spinner and estimating the probability as ½), which the class had discussed several minutes before this excerpt. She put Leon and Ken's responses side by side, as she highlighted what they wrote for denominators for all possibilities, "Leon was talking about he said there were two possibilities but Ken is counting this is one this is two this is three and this is four." Rana reflected on her instructional move illustrated in Excerpt 11:

> I don't know move along slowly point out what they did see what they got kind of validate that okay you're on the right track like Ken when he said there's four chances you know I said he is on the right track he is almost there cause there are four possibilities but it's not the four you are talking about. He recognized that there were two spinners which is good. But you know the answer is not you got gray it's you got this and this. [Post Interview, April 2010]

The interview excerpt indicated that Rana recognized not only what Ken understood about the compound event ("He recognized that there were two spinners which is good") but also what he missed ("But you know the answer is not you got gray it's you got this and this."). Validating what Ken knew is a way to build from where he was and push toward a higher level of thinking.

Directly following the discussion about Ken's thinking (in Excerpt 11), Rana displayed Monique's response under a document camera to talk about her way of thinking, demonstrating further approximations of highlighting and juxtaposing. Monique chose ¼, which is the correct choice, and explained:

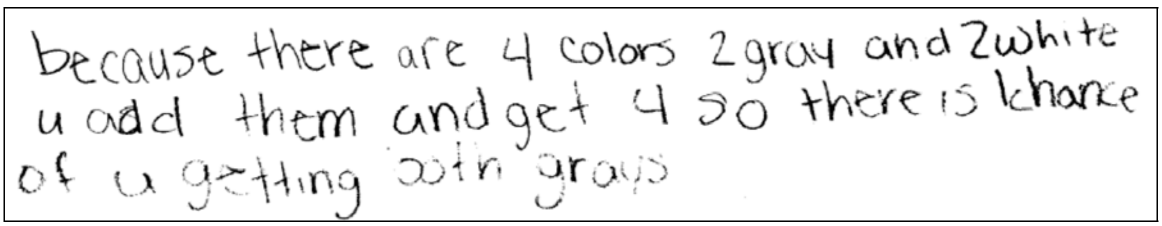

because there are 4 colors 2 gray and 2white
u add them and get 4 so there is 1chance
of u getting both grays

Figure 11. Monique's Response.

Although Monique chose the right answer out of the choices given, her explanation indicated that she did not understand how to generate outcome spaces for a compound event. She did not list any possible outcomes on her test sheet. The first part of her response ("there are 4 colors 2 gray and 2 white u add them and get 4") suggests that she was considering the four sections of the two spinners as possible outcomes, as Ken had. Monique confirmed this interpretation, telling the class "Mine is like his" during the discussion of Ken's method, (Excerpt 11, line 2). When discussing Monique's thinking, Rana only highlighted the second part ("so there is one chance of u getting both grays"), which was an important idea to consider in solving the item successfully.

Excerpt 12

1	T:	Now we have one more. Let's look at this and let's talk about how this one is different from everybody else's. So she chose one fourth, so she is the only person who chooses one fourth. But let's see why she chooses it. She says because there are four colors two gray and two white you add them you get four so there is a one (1) one chance of getting both grays. Okay. What do you mean by getting both grays? Cause everybody else was talking about like ((putting test sheets down to do hand gesture)) either you get this ((moving her left hand from the center of her body to the outside)) or get this ((moving her right hand from the center of her body to the outsider)). Tell me something she could get from this. What if I spin it, what can I get? ((gesturing spinning motion with both hands))
2	Monique:	Gray (.) and white.
3	T:	Okay. So she is saying I can get gray and white. Do you guys agree with that?
4	Ss:	Yes.
5	T:	Okay. If gray and white win?
6	Monique:	Uh-huh.
7	T:	Okay. What else could I get Monique?
8	Monique:	Gray and Gray.
9	T:	Okay. So let's change it. ((manipulating smart board)) Okay. So we can get gray white then we can get
10	Monique:	Gray and gray.
11	T:	Gray and Gray. Alright. What else could we get?
12	Monique:	White and white.
13	T:	Okay. So what else? What else can I get after white and white? Cause you said 1 out of 4. Cause you said there are four different ways to get it. ((trying to make the board work to write the sample space))

Rana drew students' attention to comparing Monique's response with the previously shared responses, directing them to think about "how this one [Monique's response] is different from everybody else's." She then shared her noticing of differences in students' responses ("So she chose one fourth, so she is the only person who chooses one fourth"), suggesting Rana wanted to show students that Monique chose ¼ in comparison to 1/2 as Leon and Ken chose.

Rana also highlighted an important mathematical idea. She asked Monique to elaborate on the second part of her response, "What do you mean by getting both grays?" juxtaposing and highlighting exactly how Monique's response indicated a different way of thinking from those provided by other students. Rana amplified the point one more time by contrasting with other students' ideas ("Cause everybody else was talking about like either you get this or get this").

What is illustrated above is a significant change in Rana's practice. She invested the class time to develop a shared understanding about an incorrect response instead of using the time for practicing procedures using a greater number of items.

As seen in Excerpts 11 and 12, Rana set the class up to participate in productive assessment talk by highlighting and juxtaposing students' responses. However, my analysis of the excerpts indicates that the juxtapositions she made did not seem to make critical conceptual distinctions visible.

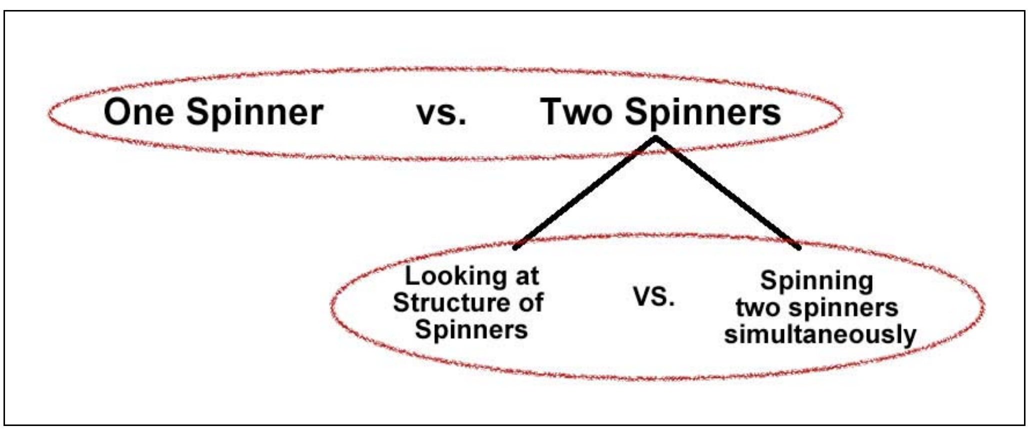


Figure 12. Important conceptual differences in thinking about Two Spinners

In excerpt 11, Rana attempted to highlight differences in Leon's and Ken' ways of thinking by juxtaposing them. In Figure 12, the upper ellipse (one spinner vs. two spinners) represents this contrast. However, she did not make more critical conceptual distinctions visible in understanding the mathematical concept, which is represented in the second ellipse in Figure 12. Rana did not enact her noticing about Ken's way of thinking ("He recognized that there were two spinners which is good. But you know the answer is not you got gray it's you got this and this") in juxtaposing students' responses. Rana's summary of other students' ways of thinking ("Cause everybody else was talking about like either you get this or get this") with her gesture seemed to signify that she was referring to target outcomes of spinning one spinner. She indicated the two color choices of one spinner by moving her hands from the center of her body to the outsider one by one. Ken's way of looking at both the spinners, yet looking at structure of spinners was not juxtaposed with spinning two spinners simultaneously, which would be more productive than jumping to juxtaposing spinning one spinner and spinning two spinners simultaneously.

Another reason I identify Rana's practice as *approximations* to juxtaposing and highlighting is Rana did not make use of highlighting and juxtaposing to provide other students opportunities to reason about the different ideas she juxtaposed. In Excerpt 11, Rana juxtaposed Leon's and Ken's ways of thinking side by side ("Leon was talking about he said there were two possibilities but Ken is counting this is one this is two this is three and this is four."), but did not invite students to participate in the conversation. Also, after Rana asked Monique to clarify what she meant by "getting both grays?" rather than waiting for Monique's clarification/elaboration so that students understood what

Monique did, Rana jumped to ask another question that was providing the procedure to solve the problem, “Tell me something she could get from this. What if I spin it, what can I get? (Excerpt 12, line 1)” Rana seemed to hybridize juxtaposing with asking a series of questions to illustrate procedural steps to get the right answer. Before Monique clarified her written response, Rana provided her interpretation of “by getting both grays” by gesturing a spinning motion with both hands. In the remaining interaction around Monique’s response, Rana employed an I-R-E structure to elicit procedural steps to get the right answer, consisting of one-on-one interaction with Monique.

Rana’s enactment of approximation of highlighting and juxtaposing seemed to be supported by Rich’s presentations on the assessment system at the workshop and most critically by Rich’s demonstration in the video annotated construct maps. Rana stated:

> … in the video exemplar, we see little clips of his classroom … get the kids to talk to each other about their thinking and then it's going to be more concrete and a lot of the questions that he asks are throughout the units are, well, he asks the kids to explain, well, how would you do this differently? How was yours different than theirs, or hey, this other kid, how was his different than his and they seem to get so much out of it. [Post Interview, October 2009]

Rana pointed out important instructional moves that she noticed from the video annotated construct maps. She identified teacher’s questioning, students’ explanation of their thinking, and teacher’s orchestration of juxtaposition. She continued talking about her hesitance to enact these kinds of instructional moves in Year 1:

> before I started doing it that way it seemed it would just be a waste like the kids wouldn't really come up with anything and I guess I was a little pessimistic but actually in class they're coming up with really good, and they're able to say what I want to say … I think that that's really valuable and even though it may, it may seem like it takes a lot more time than if I just say, hey here's how it is. If I just say, “Hey, here's how it is.” Then they're not going to really remember anything about it. They may not internalize at all, but if they're actually doing the comparison themselves then they'll, they'll get so much more out of it. [Post Interview, October 2009]

The interview excerpt indicates changes in Rana's practice and belief about her students' ability and learning. Pressure from standardized tests and inaccurate perceptions of her students' ability made her hesitate to have classroom discussion. However, once she attempted to model Rich's instructional moves, she saw that her students were able to come up with ways of thinking illustrated in the classification system and saw that it provided productive learning opportunities to students.

Summary of Rana's Practice in Year 2. The analysis shows a significant change in Rana's perspective on students' responses: from a dichotomous perspective to a focus on different levels of understanding. This change in perspective significantly influenced how Rana dealt with wrong answers during classroom discussion. In Year 1, Rana did not elicit how students who had a wrong answer thought about a mathematical concept because she believed it would not support learning for the whole class. However, in Year 2, Rana became interested in learning how students arrived at wrong answers and in making their thinking public for other students' learning. Rana started to support students in developing their own conceptual distinctions among different mathematical ideas, rather than simply presenting strategies for students to get right answers. Rana employed an approximation of highlighting and juxtaposing practices to orchestrate assessment item discussions.

Summary of Rana's Naturalization Process of the Assessment System

Rana exhibited practices and beliefs that were a hybrid of typical of traditional assessment practice (i.e., "getting it right") and reform-oriented teaching practice (i.e., "making sense of mathematical ideas") in Year 1. She classified students' responses into two broad categories, “right and wrong.” During whole class discussion about assessment items, she elicited descriptions of the procedural steps for obtaining a right answer or used a turn-taking discourse structure to share different students’ strategies that arrived at a right answer. In addition, she asked questions that helped students experience mathematics as a form of sense-making (e.g., helping Justice make sense of his incorrect calculation of the mean in relation to the distribution of data).

In Year 2, Rana’s assessment practices changed more toward reform-oriented assessment practice. First, she started to see the value of talking about wrong answers and considered them as building blocks toward higher levels of thinking. Second, she identified different levels of students’ understanding by using the scoring exemplars, and via highlighting and juxtaposing student responses, made the mathematical grounds of their responses visible to the class.

This change seemed to be supported by her participation in the workshop, and most obviously by her adaptation of the scoring exemplars and video annotated construct maps. Particularly, scoring exemplars and the video annotated construct maps supported Rana not only to differentiate forms of students’ reasoning, but also to provide a way of structuring comparisons among different students’ reasoning so that her students had opportunities to make the same conceptual differentiations she made. In doing so, she selected students’ responses that illustrated different levels of performances in the scoring exemplar and structured sharing of the selected responses following the sophistication, enacting approximations of highlighting and juxtaposing, an advanced form of a turn-taking structure.

Catherine: Asking Content-Specific Questions to Support Conceptual Changes

Catherine was in her second year of teaching 5th grade during her participation in the study. She majored in elementary education with a minor in elementary mathematics. Catherine had potential opportunities to develop attention to student thinking since it was seen as an instructional resource at her school. Catherine's principal was supportive of collaborating with researchers to improve teachers' quality of instruction. For example, the district math specialist and another master teacher held CGI (Cognitively Guided Instruction) workshops at the school. Catherine stated that encouragement and positive feedback on the workshop from these school leaders influenced her decision to participate in the workshop. Catherine also participated in another CGI workshop about fractions during the summer vacation immediately before she participated in this study, where she might have had further opportunities to talk about the importance of attending to the ideas that students were brining to instruction and to analyze students' ways of thinking about mathematical ideas.

In spite of the supportive school environment, there were constraints on Catherine's use of the assessment system that prevented her from having a full experience of how students develop statistical reasoning. For example, she felt pressure for her students to perform well on standardized tests, and it was important for her to align her teaching with the district pacing guide. The order in which the workshop introduced the mathematical concepts, so that ideas would build on each other, did not align well with her school's pacing guide. For example, the workshop started with measurement, but the paging guide mandated that this idea be addressed in March. In response, Catherine selected only the parts of the assessment system that were closely related to fifth grade standards to use in her teaching. She taught the lessons on data display, measures of center and chance, while using assessment items that most closely resembled questions on the state standardized test.

Catherine's Practice Early in the Year

Appropriating the Assessment System to Existing Practice. In early observations, Catherine adapted the assessment system in ways that minimally disrupted her existing practice. She selected particular learning activities and assessment items to meet her accountability requirements. For example, one of the requirements was discussion-based instruction. According to Catherine, the district had very specific goals for classroom discussion in that "they're just pushing student talk this year, and everything we do is that the students should be talking, you know, seventy-five percent of the time, and the teacher only twenty-five percent of the time." She pointed out specific elements in the lessons that were well fit with her district policy:

> These units I find fitting in very well with what we're working on as a district because so many of the activities, they are based around, okay, try this activity, do this, okay, discuss this. What are you thinking about? Why did you get that? Comparing answers with other members, so that fits in very well with what we're working on as a school and as a district. [Post Interview, January 2010]

Catherine noticed that Thought-Revealing-Questions in the lessons were effective in managing discussion-based instruction, but she did not address the big mathematical ideas salient in the lessons.

Catherine also selected assessment items that aligned with her school standards. For example, Catherine scored students' responses on an assessment item, *Buttoned*

Shirts (See Figure 13) and wrote the scores on a sheet of paper, as in Figure 14 below. The item assesses students' understanding of representing chance in probability.

A machine that puts buttons on dress shirts sometimes adds an extra button or leaves off a button. The company collected data about **20** shirts the machine made in an hour and **counted how often a button was missing or an extra button was added.**

Here is the data they collected:

	Missing a Button	Correct	Extra Button Added
Shirt Count	4	12	4

1. Given this data, what is the probability a shirt will be **missing** a button?

   ____

   Show how you calculated it.

Figure 13. Buttoned Shirts Item.

Figure 14 shows that Catherine grouped students' responses by the answers that students provided rather than by levels of understanding, suggesting that she did not use the scoring exemplar to analyze student thinking. The scoring exemplar provided levels and interpretations of these different responses in terms of mathematical understanding. For example, according to the scoring exemplar, the answer "4" was scored as Cha 1B, "Provides the frequencies rather than ratios." 20/4 was scored as Cha 2B- and interpreted as "The response indicates that the student understands that the probability is a relationship between the frequency and a total, but the student expresses the inverse relation, using the frequency as denominator. The student in this level may or may not have correctly identified the total." 4/8 was scored as Cha 2B and interpreted as "The response indicates that the student understands that the probability is a relationship between the frequency and a total, but the student is not able to correctly express the relationship by failing to correctly identify the numerator or the denominator." 4/20 and 20% were scored as Cha 3C. The response was interpreted as "Correctly quantifies probability as the ratio of the number of target outcomes to all possible outcomes by providing the correct percentages or ratios." 20% and 4/20 were the highest levels of understanding in different forms of representing the probability, but Catherine put them in different columns.

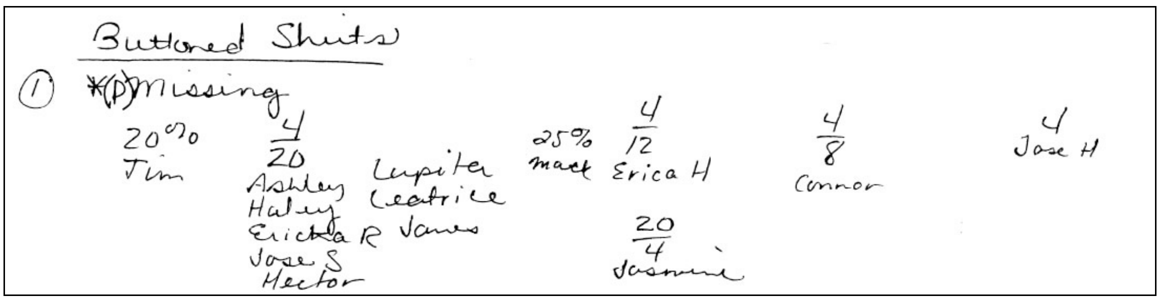


Figure 14. Catherine's scoring to mark a range of students' responses.

Her post observation interview suggested that she was differentiating different kinds of conceptual understanding from what the item was intended to assess: she was assessing students' understanding of proportional reasoning, whereas the item was intended to evaluate students' understanding of quantifying chance. This explained why she separated fractions from percents , although they represented the same probability. She stated:

> I could pull out that they had trouble quantifying probability, they're still struggling with fractions, and ratios. … So even, you know, even stepping aside from, from the big idea, from the big ideas of chance, I can pull out, you know, the other math that they're need work with or are struggling with yet too. [Post Interview, January 2010]

It is evident from the excerpt that Catherine was trying to meet her school demands using the lessons and assessment items. In this case, the lessons and assessment items were flexible enough for Catherine to interpret the meanings, and as a result they could be employed easily for her daily teaching practice. However, Catherine did not organize classroom discussion around big ideas that the item was intended to elicit.

When Catherine scored students' responses on assessment items, she applied her existing perspective. Even when she used the associated scoring exemplars, she translated different levels into dichotomous categories (i.e., correct vs. incorrect). She stated, "I really feel like, like even though some of the scoring guides, they have different [levels]… I guess I feel like either the way the problems were set up at least for the ones I scored, either they get it or they don't." Catherine did not seem to capitalize on the different levels illustrated by the scoring exemplars, and did not look at students' responses through the perspective provided by the scoring exemplar. For example, students often do not order data from least to greatest when they find a median, which means that students simply identify a middle number instead of a middle number of an ordered set of values. The scoring exemplar differentiated this as a different way of thinking about a median and indicated its mathematical significance. However, Catherine scored "finding a middle number in the unordered set" as a wrong answer and did not select any student's response as an opportunity to explore this interpretation of the median.

Letting Students Share Different Ideas. In early observations, Catherine employed a turn-taking discourse pattern and provided her students opportunities to explain/justify their ideas and withheld the information about right or wrong until several students came to revise their initial incomplete explanation by listening to their peer's correct explanation. In other words, Catherine let students be evaluators of their peers' justifications instead of playing the evaluator's role. Her instructional goal was to support students to "build on the knowledge they already have without me." Catherine explained rationale behind this practice:

> Again, for the most part, they, they helped themselves. My role in helping there was to you know continue to, to push them to share their, their ideas, their thinking and to explain why they drew this picture or why they wrote it out this way or what this, you know what this drawing means … you know they did more of the convincing to each other. [Post Interview, October 2009]

Following her image of role of a teacher, Catherine asked follow-up questions to elicit students' explanation/ justification. In line with Catherine aligning the curriculum materials with state standard requirements, she viewed students' responses as evidence of specific skills that state standards stated and as right or wrong. Catherine did not notice students' responses that were wrong but could be built on toward higher levels of performances. This perspective was evident in analysis of a data display assessment item talk, *Jumping Rope* (Figure15). The conversation happened in the second visit (October 2009). The assessment item assessed knowledge about data display, particularly how different displays were better than others for showing particular data patterns.

Jumping Rope

Dora counted how many rope jumps she can do in one minute. Here is the number of jumps she did in 20 trials of one minute each.

25, 26, 27, 27, 26, 28, 30, 26, 27, 28, 26, 25, 27, 29, 28, 19, 26, 25, 28, 29

1. Given this sample, make a graph that helps you think about how you expect Dora to perform in general.

Later, Dora's father gave her a lightweight jumping rope. He suggested that this rope will help her make more jumps in one minute. Dora counted her jumps with the lightweight rope. Here are the results of her 20 trials.

27, 28, 29, 29, 28, 30, 29, 28, 29, 30, 28, 29, 29, 30, 29, 29, 27, 30, 27, 28

2. Make a display that helps you think about Dora's performance using the lightweight rope.

The scoring exemplar illustrates five construct levels: At the lowest level, students who attend to values or groups of values without relating the data to the question would be scored as DaD 1A. Students who attend only to specific data points (such as maximum or minimum) would be scored as DaD 2A. Students who simply order the data and list them should be placed at DaD 2B. At DaD 3A, students create a display attending to repeated values (e.g., frequency) or clumps. Finally, at the highest level (DaD 4A), students would make visible both ordinal properties and continuity (e.g., scale) by using a number line display.

Figure 15. Description of Jumping Rope & related levels of performances.

Catherine had a very different goal from the intention of the assessment item. Catherine told the class, "I'm looking for us to notice is it okay that you all did not make the same type of graph? Yes right? As long as you can explain what your graph was showing us." When Catherine discussed the assessment item with students, Catherine stated that their rationales were all good as long as they were able to explain their rationales.

Employing a turn-taking discourse pattern, Catherine called on students to present their displays of the number of jumps in 20 trials of one minute using a heavier rope. Rene volunteered to present her display first. Rene noticed values that were the same and wanted to show the mode of the data (See Figure 16). So her display fulfilled her goal in that regard.

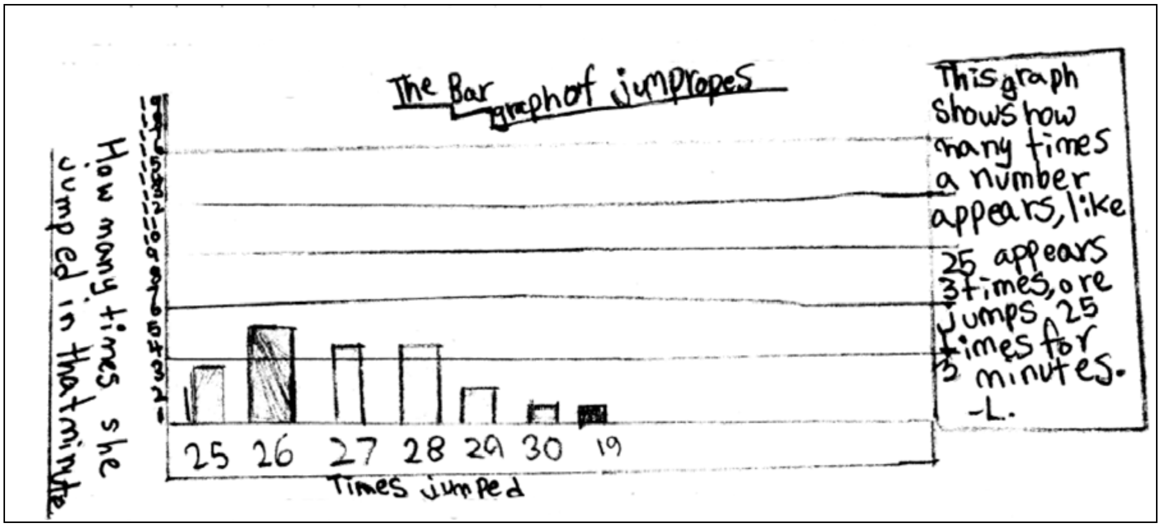


Figure 16. Rene’s Graph.

Interestingly, she omitted one data point (19) by mistake, but Catherine did not notice that Rene did not include 19. Catherine requested Rene to explain her display (“Tell us what you did and why you chose it”). Rene provided a long explanation:

> Well I did a bar graph because it's pretty much easy to read and these are the numbers of how many times she jumped a minute. This is 1 minute, this is another minute and each one of these are a minute, so it's easy to read that and how many times it appears is up on the big side, so this appears 3 times, 5 times, 4 times, 4 times and so on. So that's why I chose it cause it's easier to read.

After Rene’s explanation, Catherine replied, “Okay and I saw we had quite a few bar graphs. Do you have any questions for Rene about what her graph is showing?”

Catherine's response suggested that she focused on talking about traditional school mathematics standards (e.g., knowing graph names) rather than helping students understand what Rene's graph showed about the data.

The following interaction provided further evidence that Catherine did not notice mathematical significance in students' responses. For example, later a student pointed out that Rene omitted 19. This might have been a good opportunity to discuss how to change Rene's graph to show order and holes in the data. That would have helped a number of students at DaD 2A move up to the next level (DaD 2B) or at DaD 3A to DaD 4A. However, Catherine did not take up this instructional opportunity that she made by asking Rene what she would do with 19. Rene added 19 after 30 as you can see in Figure 16. However, Catherine did not problematize this. Instead, Catherine wrapped up the conversation by saying, "That [omitting a value] can happen, but that's why we need to go back and be careful. Okay. We check with those, but very nice job, Mark, being very observant on that piece of data. Let's take a look at another type of graph somebody did." She treated the missing 19 as a mistake, and did not seem to notice the mathematical significance in how Rene added 19 or what this suggested about her understanding of data display concepts. She emphasized non-mathematics related skill, "we need to go back and be careful."

While the scoring exemplar provided interpretations and suggested levels of student thinking about how different data displays showed different patterns in data in better ways, Catherine did not take this perspective when she discussed the item in her classroom. As a result, she appeared not to notice different levels of thinking and did not push student thinking toward higher levels. Instead she pushed students to notice different types of displays by taking turns to share and justify students' displays.

A further illustration of *not noticing* can be seen in the following exchange (Excerpt 13). The class was looking at two frequency displays of the two data sets, one with a heavier rope and the other with a lighter rope (Figure 17). Catherine asked students if they could know which display was representing which data set.

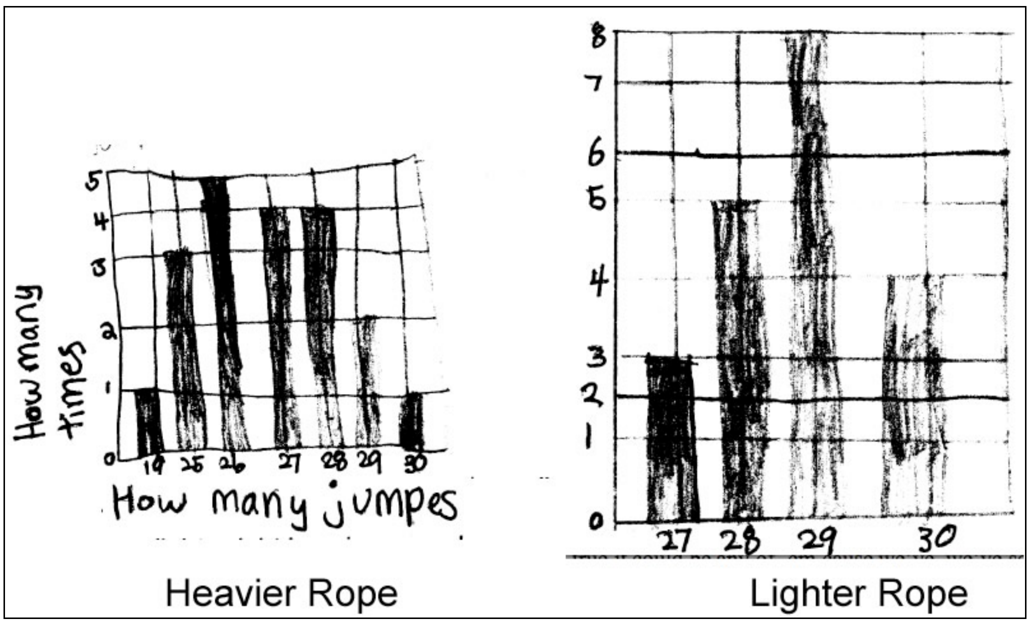


Figure 17. A student's displays of results from heavier rope and lighter rope.

Excerpt 13

1	Teodor:	Like the bottom one ((referring the graph on the right in Figure 17))? ( ) show of a <spike> in how much more she did? And um the top one ((referring the graph on the left side in Figure 17)) it really just stays under 29? It doesn't really go up to the 30's.
2	Rene:	It goes up to thirties.
3	T:	What Rene?
4	Rene:	((undecipherable))
5	T:	Yeah it goes up to (30 once). I loved the language Teodor used that in the second graph here that he saw he said a spike at 29. What does that tell us?
6	Teodor:	They can spike or increase at um that point.
7	T:	>Increases at that point<. (0.7) I'm not sure what you mean it increases at that point?
8	Teodor:	Cause like um top one with the spikes it goes up higher and she does more jumps in a minute than in the first one.
9	T:	Okay. She does more jumps in a minute. Do ya'll have any questions? I feel like we're a little, we're a little drawn out on this, so I wanna, I wanna keep us moving along.

Catherine's instructional moves suggest that she did not notice the significant mathematical ideas in students' responses. Two mathematically significant ideas were elicited. Teodor noticed the central tendency of the two displays (DaD 3A). He specifically pointed out that he would not consider performances far from the middle, saying "the top one it really just stays under 29? It doesn't really go up to the 30's." Disagreeing with Teodor, Rene interpreted that Teodor did not see that there was one 30 (line 2) and pointed out that there was one 30. Rene did not understand what Teodor meant. In response to the disagreement, Catherine simply confirmed that there was one 30. She did not ask Teodor to elaborate on his thinking, which might have helped Rene read data from aggregate perspective (e.g., shape) not just from case based perspective (e.g., specific points).

In the next turn, there was an opportunity that Catherine was able to talk about shape and what the shape told them about central tendency by “spikes”, but again she did not make use of the instructional opportunity. Catherine drew students’ attention to Teodor’s idea, “spikes,” which she asked him to elaborate (Line 5). However, Catherine had difficulties understanding Teodor’s thinking or connecting his thinking to the mathematical idea, shape of data – the primary intention of the unit. Teodor said, “Like a spike where increase at that point.” Catherine repeated, “Increases at that point” with very low and slow voice signifying she was trying to understand what he just said. Teodor provided further explanation in line 8 at Catherine’s request. Then, Catherine only repeated the correct conclusion part, “she does more jumps in a minute,” without unpacking representational evidence (“Cause like um top one with the spikes it goes up higher”). She ended the discussion by asking students to write what they learned from today’s class. Catherine did not talk about what they wrote, instead she moved onto the next assessment item.

Catherine in Excerpt 13 created instructional moments by asking students to elaborate their thinking. She expressed her interests and efforts to understand students’ thinking (line 7 in Excerpt 13, “I’m not sure what you mean it increases at that point?”). Catherine seemed to notice some key words from students’ responses, but did not use the sharing of different thinking to drive students’ understanding toward particular levels of understanding.

Summary of Catherine’s Practice Early in the Year. At the outset of the study, Catherine appropriated the elements of the assessment system to her existing practice (e.g., district policy on classroom discussion, school standards). The scoring exemplars did not affect Catherine's perspective on categorizing students' responses (i.e., correct vs. incorrect). Rather, Catherine lumped the levels except the highest level together and treated them as "wrong." This suggests the assessment item, as a boundary object, was used by Catherine to elicit students' responses but was subject to flexible interpretation (Bowker & Star, 1999). She structured the classroom discussion mainly by using a turn-taking discourse structure to let students share their explanations/justifications of mathematical ideas. She elicited important mathematical ideas but did not capitalize on them to develop student understanding. For example, Rene's adding 19 after 30 provides productive instructional moments in that the class could discuss the ordinal and the continuous scale. However, Catherine seemed not to notice the instructional moments.

Catherine's Practice Later in the Year

Categorizing and Interpreting Student Thinking in Light of Scoring Exemplars. In contrast to Catherine's characterization of student thinking solely as right and wrong in early observations, her final scoring of students' responses on *Two Spinners* item (See Figure 2 for the description of the item) showed some changes in how she categorized students' responses (April 2010).

Catherine scored all students' responses by using the scoring exemplar and annotated the levels of students' understanding on individuals' tests focusing on students' reasoning. Then she made a planning sheet for classroom discussion (Figure 18).

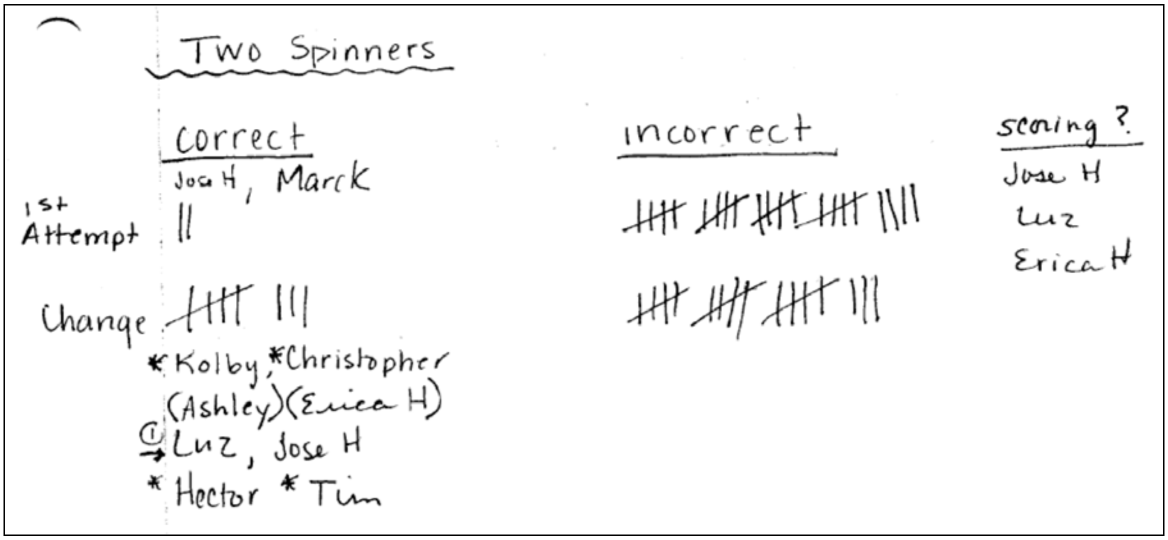


Figure 18. Catherine's scoring sheet of Two Spinners

Her planning sheet suggests that she attended to changes in student thinking. Catherine identified students who answered correctly in the first attempt and who changed their ways of thinking and answered correctly in the second attempt, as annotated in her scoring sheet (Figure 18). To keep track of the changes in student thinking, she asked her students to use different colors of pens to describe changes in answers and explanations. Catherine wrote * in front of the names of students who provided the highest level of reasoning. On the right side of the paper, she listed three students whose reasoning she could not interpret so she could ask further questions to understand their reasoning during classroom discussion. Catherine stated:

> This was the, kind of my planning sheet for today. … And this is where I'd put an arrow next to Mio and the number one. So, I was thinking, "Oh, I want him to go first." I mean first for, to say one-fourth. I had meant to start with some incorrect responses first. [Post Interview, April 2010]

Catherine wanted to start with Mio because of the inconsistency between his answer and written explanations: although he chose the correct answer (1/4) out of the four multiple choices, he based his reasoning on the structure of the spinners, not the outcome spaces. His rationale was "Because if I combined I will get 4 parts so I choose ¼." Catherine said:

> He said, "Well, these are my four parts. One, two, three, four." And he said, "And so, only one is for gray." Well, if I, when we look at it, two parts are gray obviously. So, I think that's why, when we got to one-fourth, I kinda wanted to start with him because the way he had it written on his paper there's four parts, but he didn't have out-, he didn't say outcomes and have them listed. He just said there was four parts. [Post Interview, April 2010]

Although Mio provided his logic behind choosing 4 ("4 parts"), he did not explain how he thought about 1. During classroom interactions, Catherine further probed Mio to understand how he thought about 1. In doing so, Catherine specifically asked Mio to explain his thinking about 1 ("You showed us your four parts that you thought. Where did where are you getting one from for one fourth?"), instead of asking him to explain his answer. This suggests that Catherine made instructional moves that were aligned with the significant landmarks of conceptual development illustrated in construct maps.

Noticing Significant Mathematical Ideas from Students' Thinking and Acting on them. In an assessment talk later in the year (April 2010), Catherine made particular instructional moves with an eye toward supporting students' learning based on diagnosis of student thinking facilitated by the scoring exemplars. The support came in the form of questioning and transforming the assessment system: (1) coming up with level-specific questions and (2) transforming an assessment item to make mathematical ideas visible.

Coming Up with Level- Specific Questions. In contrast to earlier in the year, when Catherine did not respond to significant mathematical thinking as revealed by students' responses in the early observations (e.g., Rene adding 19 after 30), she identified and explicitly addressed different levels of student thinking at the end of the year (April 2010). In this example, the class discussed “*Two Spinners*” item (See Figure 2). Jamie chose ½ because there were two spinners and only one landed on a gray section. She said, “Because like there’s two spinners right? And then a there’s two at the denominator and then there’s only like it says to land on gray section? And then there’s only one that landed on gray section.” The denominator (2) comes neither from the number of parts of spinner nor outcome spaces (i.e. combinations). Catherine noticed that Jamie was considering the particular outcome as was presented in the diagram in the item (Cha 1B), which was characterized as “outcome approach” (Konold, 1989). Based on her noticing, she attempted to support Jamie by reminding her of a critical feature of chance, a repeated process.

Excerpt 14

1	T:	What if we spun them again?
2	Jamie:	Okay. ((putting her head down))
3	T:	Like? What if we ((changing the picture by moving one arrow to white section)) what if the picture looks like that right now Jamie? Would that change your mind in any way? Leave the same? We still want to know or we still know that the way we win the game is that BOTH spinners land in the gray section. I just changed the picture? What happens every time we spin? Is that always gonna land on the same place?
4	Ss:	No.
5	S1:	There's one fourth.
6	S2:	Unless you're lucky.
7	T:	Yeah. That's why we are talking about just the chances we've got.

Catherine asked a level-specific question in response to Jamie’s way of thinking about chance. Catherine asked (line 1), “Okay. What if we spun them again?” This question was intended to help Jamie think about other possible outcomes (Cha 5B) and the repetition of trials. Jamie seemed not to understand Catherine's scaffolding question, and was ready to end the interaction (line 2). However, Catherine continued to help Jamie. This time, her support consisted of animating the item to illustrate what was possible if they spun the spinner again (line 3). Catherine changed the spinner on the left side; now both spinners landed on white. Catherine showed a different possible outcome when she spun again. Then she reinforced again the idea of repetition, "What happens every time we spin? Is that always gonna land on the same place?" Catherine told students that chance involves repeated trials to make a good prediction because of the factor, "luck," in line 7.

Catherine understood very precisely how Jamie was thinking about the problem. Instead of explaining it, she asked a level specific question and then asked follow-up questions to see how Jamie would think after Catherine provided some help. Catherine reflected on this instructional move in her post interview:

> I was really surprised that the kids, a couple of them looked just at where the arrows were and said, "One arrow's on grey. So, it's one out of one, two, three, four sections." That's why I went and turned the other arrow to white and said, "Well, what do you think now?" I re-, like I think that never, I mean, they, they weren't thinking about the repetition of it. They were thinking this is, you know, a singular event. We spin once and we're done, not what happens if we keep going. [April 2010]

The excerpt from the interview illustrates that Catherine interpreted how students thought about the question, rather than evaluating with right and wrong perspective. It was a critical insight that Catherine diagnosed that students were not thinking of repetition of the event by listening to what Jamie just said, and she acted to help Jamie transform her thinking by animating the repeated process only implied by the static spinner display.

Transforming an assessment item to make mathematical ideas visible. Catherine expanded the function of the assessment item as a learning context. Transformation of part of an assessment item seemed to make mathematical ideas visible to students. The idea of treating the two spinners as a simple event rather than a compound event was identified by the assessment item (Cha 3C). Most of her students thought in this way: treating the two spinners as a simple event. For example, instead of spinning two spinners simultaneously, Lorie decided to consider only one spinner because the two spinners just looked the same. She strengthened her argument by elaborating further that the two spinners became exactly the same spinner if she combined them. Catherine reminded Lorie that she needed to spin both spinners. However, that seemed not to help Lorie.

To address this way of thinking, Catherine modified the assessment item in two ways. Catherine called the first spinner “A” and the second spinner “B” to make visible that they were spinning two different spinners at the same time. In addition to naming the spinners, she colored the white part of Spinner B with blue. This might help some students only looked at one spinner because the two spinners looked exactly the same. Catherine reflected on her instructional move.

> I’d have each piece different colors so that they can’t combine them in some way. I don’t know if that will help or not, but in my mind, that’s the first thing I think of, and that’s why I thought, I thought, “Oh, she thinks we can combine them. So, what if I change the color on half of the spinner?” So, that might be something I’d start with and see how that makes them think about it. [Post Interview, April 2010]

The excerpt indicates that Catherine made the instructional move based on her understanding of Lorie’s thinking. Instead of explaining the right answer or calling on students who knew the answer, Catherine tried to first to make visible that they had to spin the spinners at the same time.

Catherine credited the scoring exemplar and video annotated construct maps in supporting her to come up with instructional moves. The scoring exemplars helped Catherine differentiate mathematical ideas in students' responses. Catherine said:

> I think the exemplars almost help me the most because it really shows me kind of what the range of student responses can be, and when you've gotta think about scoring those, you really start to see some of the smaller differences in students' thinking and how that affects their, you know, their answering or their ability to, you know, communicate about a certain problem type. [Post Interview, April 2010]

In addition to seeing different levels and "smaller differences in students' thinking," Catherine testified that the scoring exemplar helped her think deeply about the mathematical ideas that students' responses were based on and to come up with possible instructional supports. She said, "when I through the, the exemplar and read, that was great because I had to think, why are they coming up with this answer, what could make them think in this way, and what can I do to help change it?"

If the scoring exemplars helped Catherine develop supports for specific assessment items, the video annotated construct map seemed to encourage Catherine to think about a particular type of interactional structure, level-specific questions, that she employed the majority time of her instruction.

> … when I looked at those video clips trying to think, "Well, is there something on here, did I do, or what should I do next?" I heard something that he asked a student who was demonstrating, you know and I thought, "Oh, I should've said that, I bet that would have brought up," you know, cause sometimes when you're going on the fly, the questions you want just aren't there, and I'll lay in bed at night, and I'll be thinking about it and go, "Oh, why didn't I say this? Why didn't I ask this?" [From post observation interview on April 2010]

The excerpt suggests that video annotated construct map let Catherine think about content-specific-questions that was different from questions to elicit students' explanation.

Summary of Catherine's Practice Later in the Year. Later in her participation, Catherine not only characterized students' responses in terms of levels of performances illustrated by scoring exemplars but also planned classroom discussions to address her findings about student thinking. When Catherine reified students' responses to assessment items in alignment with the scoring exemplars, she was more focused on mathematical ideas in students' responses (e.g., "He just said there was four parts") and provided instructional supports for conceptual change. The forms of instructional support were: (1) to ask level-specific questions based on her diagnosis of levels of understanding and (2) to transform assessment items to make mathematical concepts visible to students so that they can reason about them.

Summary of Catherine's Naturalization Process of the Assessment System

Catherine demonstrated some important changes in her formative assessment practice during the school year (2009-2010). In early observations of her participation in the study, Catherine created a hybrid of traditional assessment practice and reform mathematics instruction principles: she tended to characterize students' understanding of mathematical concepts from a dichotomous perspective (i.e., right and wrong), but students were not informed about the correctness of their ideas. Catherine asked several students to share their thinking and to explain and justify their solutions. She believed that sharing different students' reasoning would provide other students opportunities to construct their own explanation/ justification. However, she did not coordinate the discussion to help students see the mathematical significance of different levels of student thinking. As a consequence, students' ideas were not contested or guided toward higher levels of thinking. At the end of the year, Catherine created learning opportunities that extended beyond sharing different ways of thinking. In particular, she identified student thinking at different levels by using the scoring exemplars, and this practice seemed to help her recognize significant mathematical ideas in students' responses. In discussion, she asked content-specific questions that were tailored for particular levels of understanding. In addition, she altered the assessment items to make mathematical ideas more visible to students, which supported students to reason about previously invisible mathematical ideas.

The analysis suggests that Catherine's existing interest in student thinking was augmented by the classification system in a way that she was able to provide effective support for learning. The classification system (e.g., scoring exemplars, paper and video annotated construct maps) funneled the scope of interpretation, and this focus played an important role in making transformations in Catherine's teaching practice. The scoring exemplars helped Catherine interpret students' responses in terms of mathematical ideas. Also, video-annotated construct maps supported Catherine to learn that teachers' content-specific-questioning functioned as a critical lever for students to understand mathematical ideas.

However, her instructional structure was coordinating one-on-one interaction, mainly between the teacher and a student. For example, Mio's idea of only considering 4 parts of the two spinners and Lorie's argument of considering 2 parts of one spinner seemed to be a potentially productive juxtaposition for learning. Catherine discussed these different ways of thinking separately and did not bring them together for fruitful contrast.

Nancy: Attuning Instructional Trajectories to Learning Progression

Nancy had a bachelor's degree in elementary education with an emphasis in mathematics. She had been teaching 5th grade mathematics for 15 years. In informal talk, Nancy was referred to as a lead mathematics teacher in her building, indicating that teachers acknowledged Nancy's expertise in teaching mathematics.

Nancy was part of a supportive teaching community at her school. She described her principal as open-minded and supportive, particularly in encouraging her to participate in the data modeling workshop and to implement the assessment system in her classroom. She shared the assessment system with her principal and invited him to observe her classroom discussion as she taught with the assessment system. Nancy also worked with a district math/science coordinator, who first participated in the initial data modeling workshop by herself and then recruited all the 5th grade math teachers in the school, noting the potential benefits of the workshop for supporting student learning. Nancy and her colleagues[6] were very often engaged in talking about the assessment system and using it to plan instruction together.

Nancy's Practice in Year 1

Expecting Forms of Student Reasoning. In contrast to other case teachers, Nancy used curriculum materials (i.e., lessons and assessment items) in coordination with the classification system (i.e., construct maps and scoring exemplars) at the start of her participation in the study. As a result, rather than focusing on identifying whether students' responses were "right" or "wrong," Nancy thought deeply about students'

[6] Two 5th grade math teachers also attended the workshop with Nancy and the math coordinator.

understanding of mathematical concepts and recognized important mathematical ideas in students’ responses and ordered them in terms of sophistication. In the first year, Nancy used the classification system to anticipate the kinds of responses students would provide to particular problems. For example, when asked about how she used the classification system to plan an assessment conversation in the first year, she responded in an interview:

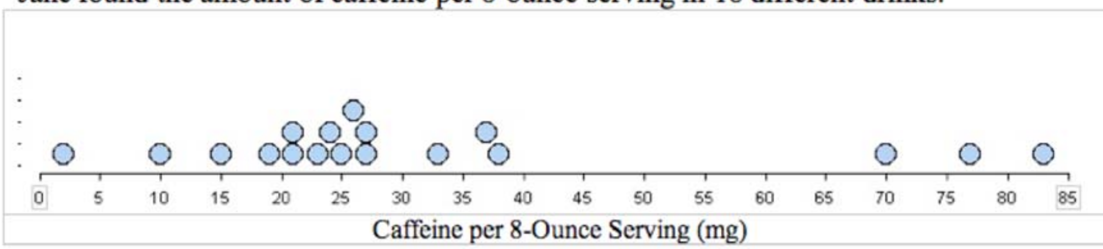
> Well I I just just read over it [a scoring exemplar] yesterday and I thought … Okay, using a continuous scale [DaD 4A], I thought some of them are not going to do that. I expected them to do a lot of this ((point at DaD 3A on the scoring exemplar)) and this … I thought some of them would do this ((unidentified)) and some of them did, and I expected all of them to at least be here ((unidentified)). I did not expect to see this [DaD 1A]. [Post interview, April 2009]

Nancy read scoring exemplars and envisioned students’ possible performances. As a result, she was ready to identify students who demonstrated particular levels of performance on the classification system. Nancy used her roving time, when students solved problems by themselves, to select students’ work for sharing. The range of selected and shared student work suggested that Nancy was able to categorize students’ responses in view of the classification system.

Pinpointing a Better Performance. It is notable that highlighting and juxtaposing was demonstrated in Nancy’s assessment talk in the very early stage of participating in the study. Anticipating and interpreting student thinking in light of the classification system seemed to allow Nancy to see the significant mathematical ideas in students’ responses. The recognition seemed to support Nancy to highlight and juxtapose significant mathematical ideas in students’ responses. However, Nancy employed the instructional move to point out a higher level of performance.

An episode of highlighting and juxtaposing with an eye toward pinpointing a better performance occurred during the first observation of Nancy's assessment talk in the first year (April 2009), when the class was discussing *Caffeine in Drinks* (Figure 19), an item adapted from the Connected Mathematics Program.

Jane found the amount of caffeine per 8-ounce serving in 18 different drinks.



1. Based on the graph above, which statement about the median and the mean is true?

    a. The median is larger than the mean.
    b. The median is smaller than the mean.
    c. They are the same.

2. Please explain why you chose your answer.

This item assesses students' understanding of the effect on statistics of changes in the components of a distribution. The two main distinctions in students' reasoning that the scoring exemplar makes are: (1) Students who rely on calculation are scored as CoS 2A (Calculate statistics indicating central tendency) and (2) Students who use components of the distribution, in this case the three outliers, to infer changes in the mean and the median are scored as CoS3D (Predict how a statistic is affected by changes in its components).

Figure 19. Description of Caffeine in Drinks & related levels of performances.

Nancy called on Tobi, who demonstrated the highest performance (CoS 3D: Predict how a statistic is affected by changes in its components), to share his response. The sharing was facilitated by Nancy's highlighting of mathematical ideas in Tobi's response: Tobi pointed out in his response: (1) a clump in the data (line 6, "most of the measurements are on the left side”) and (2) outliers (line 6, “there’s only three points towards the end of it”). Then he concluded the mean would be located outside of the clump (“when you divide it would be higher than most of this, ” line 6) by doing the calculation in his head.

Excerpt 15

1	T:	Thank you Tobi, what do you think?
2	Tobi:	I think that the median can’t be larger than the mean because of all of the=
3	T:	That’s good. ((nodding her head))
4	Tobi:	=All of the points are on the lower side of the plot.
5	T:	Can you go up there and show us up there what you’re talkin about? Okay, Tobi is gonna point something out and I want you to notice what he’s showing us. Go ahead Tobi. He’s observed something that I’d like you to notice.
		((requesting Tobi to speak up))
6	Tobi:	I think that it, the median can’t be lower [higher] than the mean because all of the, most of the measurements are on the left side, on the lower part of the bar and there’s only three points towards the end of it and that when you added them all together it would be a higher number and when you divide it would be higher than most of this and that will be ( ).
7	T:	Okay, I think, I’m gonna try to say what he’s saying and only louder. You correct me if I mess up okay? He said did you notice there’s three points that are really high up here but MOST OF the points are down here and so he said when you add it all up to get the total it’s gonna be pretty high because of these three numbers but then most of the points are down here. So what about that? If most of the points are lower and you’ve got these three really high ones here, what’s gonna happen? Thanks Tobi. Did you want to say some more about it? No? Kristine, what do you think? Thanks for getting us started.

In this excerpt, Nancy made several important moves to use Tobi’s sharing as an opportunity for other students to learn. First, she made sure that other students would be able to follow Tobi’s reasoning by asking Tobi to go to the front of the room and point at the distribution on the overhead projector (line 5). She alerted students to attend to Tobi’s idea and to try to notice something significant in it ("He's observed something that I'd like you to notice"). She revoiced Tobi as she restated his explanation, explicitly contrasting the three points that "are really high up," or the outliers, and "most of the points," or the clump. She also said the important idea ("MOST OF") very loudly to emphasize its importance. This highlighting, the coordinated use of talk (emphasis) and gesture (pointing) (Hall et al., 2002), made visible an important idea illustrated in the classification system: Visual noticing about the distribution (CoS 1A) is an initial but foundational performance that situates students' reasoning in distributions, shifting away from reliance on calculation. Here, Nancy was highlighting these foundational ideas for the class.

Nancy also invited other students to respond to these ideas by asking content-specific questions. When she asked "So what about that? If most of the points are lower and you've got these three really high ones here, what's gonna happen?" she generated learning opportunities for other students, supporting their learning with Tobi's noticing and her specific question about it. Asking content-specific questions after highlighting allowed Nancy to gather further information to support her next instructional moves. For students, it provided conceptual assistance in that students know what they should reason about the mathematical ideas highlighted. The sentence, "Thanks for getting us started," also suggests that Nancy called on Tobi to initiate discussion, not to announce the correct answer.

Although Nancy started the conversation successfully by positioning students to reason along with Tobi, the following classroom interaction illustrates that Nancy juxtaposed two different strategies at different levels with an emphasis on telling students that she preferred Tobi's method, rather than Kristine's calculational method. Instead of responding to Nancy's question about the outliers, Kristine shared how she calculated the median and mean, which was a lower performance (CoS 2A) than Tobi's strategy (CoS 3D) but still legitimate.

Excerpt 16

1	Kristine:	Because um the median is the number in the middle if you put them all in order, smallest to largest, um they're pretty much in order from smallest to largest starting, well at half of 18 would be 9 so I went to the number that was the 9 number and that was 25 so I'm guessing it's gonna be somewhere around 25 and I used my=
2	T:	=Okay.
3	Kristine:	calculator to find the mean and the mean was around 32 so.
4	T:	So you actually pretty much calculated the median and the mean? Okay you could have done that. You could have calculated it. I was more interested in how you could know without actually calculating it. You know how I like shortcuts and I like to know without having to actually do the work. She said she knew that there were 18 points here and so that 9, the 9 is next to the middle so she counted up 9 and she knew it would be somewhere around 25 and then she said she estimated that, now could you know exactly what these points were right here?
5	Kristine:	No.
6	T:	No, but did you just estimate? Okay, best guess on those and then she said she thought that the mean would be higher. Anything else? Now Tobi just did it by looking at it and he said I know the mean's gonna be higher cause you got these three points right here. What are those points gonna do?
7	David:	It's gonna make you put the mean far and fall higher.
8	T:	It's gonna raise the mean because?
9	David:	Because it's separating into the, it's higher numbers so when you do that it's just gonna make it, and when you write on, it's gonna.

In this interaction, Nancy positioned Tobi's method as a better way to get the answer than Kristine's method. Important mathematical ideas in Kristine's response were not highlighted for other students who were unable to solve the question by employing calculation. Kristine knew how to calculate the median and mean (CoS 2A) and noticed that measurements were already in order in the graph, demonstrating her ability to read a number line graph (DaD 4A). Also, Kristine located the mean and median on the distribution (“around 25” and “around 32”). These ideas are all levels of performance on the constructs that are identified as significant conceptual achievement and might have been useful to help other students understand the graph and attributes of the distribution as well as reviewing the measures of center.

Instead, Nancy moved on to making very explicit her response to Kristine’s method. Nancy said (line 6), “You could have calculated it. I was more interested in how you could know without actually calculating it. You know how I like short cuts and I like to know without having to actually do the work.” Also, Nancy provided her observation of the possible difficulty and extra work involved in Kristine’s strategy by highlighting that Kristine had to estimate some data points, “could you know exactly what these points were right here?” Then Nancy provided Tobi’s method as an example that met her criteria for solving the question in a better way, “Tobi just did it by looking at it.” Nancy highlighted one more time Tobi’s visual discovery (line 8), “he said I know the mean’s gonna be higher cause you got these three points right here.”

Nancy’s constraint on solving the question by only using visual discovery seemed to be intended to push students to infer changes in statistics by considering relations among components (CoS 3D). However, it may have been too big a conceptual jump for students to make without additional support. Nancy did not help students use the visual qualities to enact calculation in their head or conceive of the mean as a balance point. Students do not necessarily have to do algorithmic calculation to solve the question, but they have to simulate calculation mentally as Tobi did (“When you add them all together it would be a higher number and when you divide it would be higher than most of this and that will be…”: Excerpt 15, line 6).

As she continued to talk about the *Caffeine in Drinks* item on the next day of instruction, Nancy’s instructional moves indicated that she was perplexed with her students’ understanding of the assessment item, but was unable to ask content specific questions to pinpoint which conceptual blocks students were missing. Nancy started the math class by recollecting that the class had discussed the three outliers would change the mean but not the median.

Excerpt 17

1	T:	Okay. So you think the, the three outliers would change the mean but not change the median. Is that what we talked about yesterday? Who, I know there's a couple of you that still don't really get why that's true. Is there, does anybody not understand that? I thought there was somebody. Everybody gets that?
2	Ss:	Yes. ((in unison))
3	T:	You all understand that perfectly?
4	Ss:	Yes. ((in unison))
5	T:	Okay. We're ready to go on then. I thought we needed to do a little more, but I guess not. You understand that. What, can anyone explain why that would happen? Lexi, explain why that would happen in a loud voice.
6	Lexi:	Because if, if you moved 'em back they'd still be the highest numbers.
7	T:	So what does that mean?
8	Lexi:	That means it, it's not going to change any of the other numbers because it's basically just keeping 'em there but you're really just moving them a little.
9	T:	So what statistic will it not change?
10	Lexi:	The median.
11	T:	It won't change the median because it's still the three highest numbers. Okay, why will it change the mean? Can anyone tell me why it would change the mean? Will?
12	Will:	Because it's smaller so that means that the thing whatever you divide by is going to be smaller.
13	T:	The total amount that you divide by. Okay, I think you do have it. Good. Very good.

In this interaction, Nancy did not come up with content specific questions that would have helped her discover who understood the item conceptually. More specifically she did not ask level-specific questions to locate where students were in terms of the learning progression.

First, Nancy depended on students' self-reporting to test her conjecture that students did not understand the question. Her conjecture was: "I know there's a couple of you that still don't really get why that's true (line 1)." Her questions to follow up her conjecture were (line 1 and 3): "Does anybody not understand that? Everybody gets that?

You all understand that perfectly?” This self-reporting assessment might make it difficult to assess students’ understanding because it does not elicit exactly what students do and do not know. In line 2 and 4, several students said they understood the item without any hesitation. Based on students’ self-reporting, Nancy appeared to decide to move on, but she changed her mind and decided to call on students to test their understanding.

Second, Nancy did not ask follow-up questions after her initial question to collect further evidences of students’ understanding. When she asked the class about the effect of outliers on statistics (CoS 3D), Lexi remembered what Nancy did yesterday (line 6, “Because if, if you moved ‘em back they’d still be the highest numbers”). Lexi’s explanation did not address the effect of outliers on statistics (CoS 3D). Rather she was talking about the effect of repositioning the outliers on magnitude and order of them. Lexi’s response was not related to the Conceptions of Statistics construct. Instead of probing further on where Lexi’s understanding was in terms of the learning progression about Conceptions of Statistics, Nancy moved on asking another CoS 3D level of question (line 9, “So what statistic will it not change?”). Although Lexi answered correctly to Nancy’s new question, it was not clear whether Lexi understood the idea conceptually or she just provided a memorized fact.

Third, Nancy assessed whole class understanding using a transformed IRE structure in which only two students responded. The first IRE was from line 9 to 11, and the second one from 11 to 13. The two students that Nancy called on, Lexi and Will, were students that she characterized as “high” in her post interview. This could lead Nancy to over-generalize students’ understanding. In line 12, Will said, “whatever you divide by is going to be smaller” and Nancy took it as evidence of understanding. Nancy revoiced it as “the total amount that you divide by,” possibly for other students. She then evaluated Will’s response by saying, “I think you do have it. Good Very Good,” suggesting that she was convinced about students’ understanding.

The episodes in this section show that Nancy identified different levels of students thinking and attempted to make mathematical ideas visible by highlighting and juxtaposing. However, she seemed not yet to have the image of the continuum of the different levels of performances and did not enact instructional moves that connected conceptual building blocks. This lack of the image of development might have hindered her in coming up with productive instructional moves to support students’ conceptual change.

Summary of Nancy’s Practice in Year 1. Nancy focused on identifying students’ understanding of particular mathematical ideas by using the classification system and remained focused on disciplinary substance (Coffey et al., 2011) before, during, and after assessment talk. Nancy put the different ways of thinking about mathematical concepts side by side to provide opportunities for other students to see and reason about ideas that were previously not visible to them. However, the first year employment often focused on pointing out a higher level of performance.

Nancy’s Practice in Year 2

Identifying Variations in Students’ Levels of Understanding & Deepening Mathematical Disciplinary Perspective. In the second year, Nancy made use of construct maps and scoring exemplars to reify all students’ responses in terms of levels of the classification system, which provided more information about her students’ ways of thinking. Figure 20 illustrates an example of Nancy’s scoring.

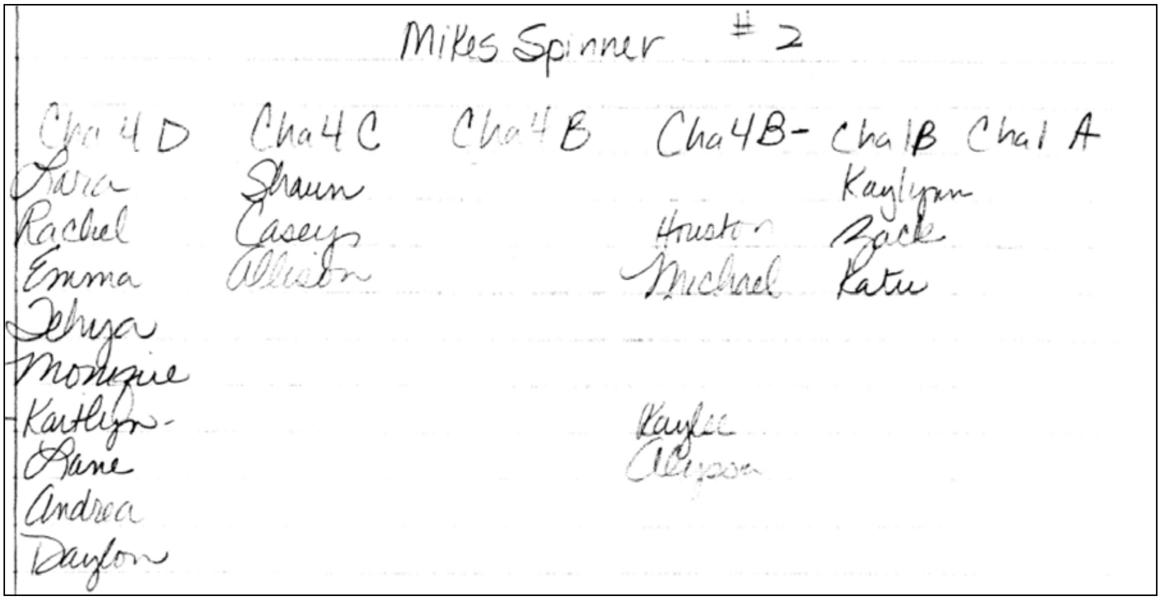


Figure 20. Nancy’s scoring sheet of an assessment item.

Scoring in this way allowed Nancy to understand the variation in her students’ thinking. She used this information to select which assessment items she would discuss, targeting those where numerous students showed low levels of performance. For instance, she indicated that she chose to review a data display item:

> because I really am concerned that they don’t notice the gaps and that they didn’t, only one student noticed the benchmark the same size. [Post Interview, October 2009]

Here, Nancy mentioned the specific performance (DaD 4A: Display data in ways that use its continuous scale to see holes and clumps in the data) that she wanted to focus on during assessment talk, suggesting that the scoring allowed her to target this understanding in her classroom instruction. She also pointed out that only one student indicated understanding of size of interval, which guided her to decide to have an assessment talk about the item.

In addition, Nancy deepened her understanding of the relationship between students' responses and mathematical significance in Year 2. It was beyond anticipating the range of students' responses in Year 1. Nancy explained in her post interview:

> Well probably I hadn't thought about the fact that if they mention specific data points that that's more, that that's a higher level. That they're actually using proof from the display … I probably would not have noticed the difference between that if it hadn't been for the exemplar saying they have to mention specific data points to be on this level and if they don't mention specific data points then they're only on this level. [Post Interview, January 2010]

The scoring exemplar supported Nancy in recognizing key mathematical aspects of students' responses. In particular, she made distinctions between student responses ("specific data points") and understood the mathematical significance of those distinctions ("using proof from the display"). As Nancy continued to talk, it became clear that this noticing informed her in deciding next instructional moves.

> Then I'm like oh, that is more perceptive that they would say that there's 12 here and 8 here where I might have just lumped it all together if I hadn't been looking at that [scoring exemplar]. I would have said yeah there's more. That's the same as saying there's 12. [Post Interview, January 2010]

The practice of identifying mathematical ideas and levels of understanding in student responses is very important in supporting students to move to higher levels of thinking. As Nancy noted, she would not have made instructional moves to help students' progress from visual summaries ("there's more") of data to quantification ("there's 12 here and 8 here") without noticing these differences in student responses. Over the course of the study, Nancy mentioned several examples that she indicated she could not have differentiated without the scoring exemplars.

Juxtaposing Different Ideas to Make them Under the Attention of Students and Position Students To Reason about them. Juxtaposition is a way of supporting students to construct mathematical conceptual distinctions by highlighting key contrasting aspects in different students' ideas. Here is a specific example of the practice, juxtaposing to put students in a position to discuss, that came from Year 2 (March 2010), when Nancy discussed "Two Spinners" in her class (See Figure 2 in Methodology section for detailed descriptions of important ideas in *Two Spinners*). The key mathematical idea of this item is to generate outcome spaces of a compound event by enacting the event. In the following excerpt, two important ways of thinking about outcome space in a compound event were juxtaposed. Don and Eric seemed to understand the difference between outcome spaces and physical spaces on the two spinners. However, other students, including Baylee, seemed to be confused although they did consider two spinners.

Excerpt 18

1	Don:	I did change my answer to one fourth cause after what Eric said I realize that there's only one fourth of chance cause there's four outcomes that you can get.
2	T:	Okay. You don't think there's fifty fifty chance of winning anymore half chance of winning. Okay. Um. Baylee what you are gonna say?
3	Baylee:	Um. I chose one fourth because there are there two spinners but there's four there's four parts there's two parts on one spinner and then two parts (I just realize that) four parts and that's why I chose the four and that's how I took out anything that didn't have four in it. And then I got the one because there's only one gray part on each one.
4	T:	Okay. I guess my question is why is it four? Is it because there's four spaces on there that we're looking at one two three four ((pointing at each section)) or is it because there's four different outcomes? I heard two different answers. I heard several people say well I think it's a fourth because there's four spaces and then I heard someone else say well no it's a fourth because there's four different outcomes.

Here, Nancy made two contrasts. First, she made explicit how Don changed his mind. Don pointed out that he was thinking of outcomes, which was the evidence of his understanding. He said, “I realized that there’s only one fourth of chance cause there’s four outcomes that you can get” (line 1). When Don said he changed his mind based on what his classmate, Eric, said, Nancy reminded students of his previous answer to contrast with his current thinking, “You don’t think there’s fifty fifty chance of winning anymore, half chance of winning (line 2).”

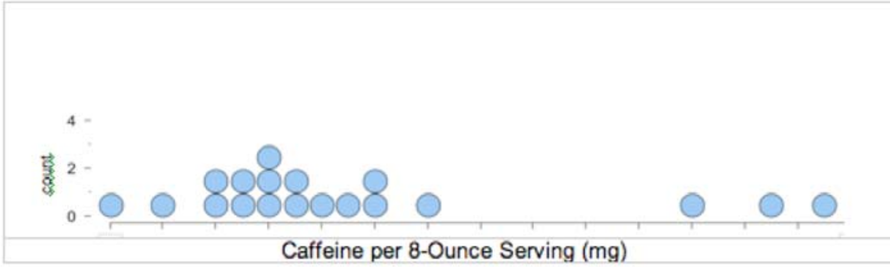
Mostly importantly, Nancy juxtaposed two important conceptual distinctions necessary for students to understand outcomes in a compound event. In contrast to Don, Baylee selected the right answer (1/4) based on parts rather than outcomes. She said, “I chose one fourth because there are there two spinners but there's four there's four parts there's two parts on one spinner and then two parts” (line 3). In response to the two students’ ideas, Nancy juxtaposed the ideas by asking a content specific question. Nancy said, “I guess my question is: why is it four?” drawing students’ attention to the commonality in Don and Baylee’s ideas, “Is it because there's four spaces on there that we're looking at one two three four or is it because there's four different outcomes?” This highlights critical differences in how they thought about the outcome space. In juxtaposing the two different ideas, she did not provide any signal that indicated the right answer. Rather, Nancy asked the question to position students to be the judge of mathematical ideas given more mathematical distinctions among students’ reasoning.

The instructional move that Nancy demonstrated here is a sophisticated form of juggling different levels of students’ responses to enable the whole class to reason about the different ideas. Nancy employed highlighting and juxtaposing as she coordinated talk (e.g., using a louder voice to emphasize), gesture (e.g., pointing at mathematical ideas in the graph), and inscription (e.g., presenting related displays through a projector) (Goodwin, 1994; Hall et al., 2002).

The analysis also identified that there was a significant change in purposing the same form of interactional structure: Nancy used several of the same practices (e.g., highlighting, juxtaposing, and content-specific questions) in both years, but she was able to use them more effectively in the second year. For example, in her work around the *Caffeine in Drinks* item, in the first year Nancy juxtaposed the computational method and the visual discovery method and challenged students to use the visual discovery method by highlighting one of the conceptual building blocks (visual qualities of the distribution). In contrast, in the second year Nancy drew important contrasting viewpoints from students' talk to make these ideas under the attention of other students, who might not notice and were positioned to reason about the ideas.

Attuning line of instructional moves to learning progression. The assessment talk about *Caffeine in Drinks* in the second year consisted of lines of instructional moves that were aligned with the learning progression expressed in the Conceptions of Statistics construct map. *Caffeine in Drinks* was revised in the second year in that data values on the x-axis were hidden (See Figure 21).

Jane found the amount of caffeine per 8-ounce serving in 19 different drinks.
This is the graph that she made.



1. Based on the graph above, which statement about the median and the mean is true?

   a. The median is **larger** than the mean.
   b. The median is **smaller** than the mean.
   c. The median and the mean are the **same.**
   d. It is impossible to tell without knowing the amounts of caffeine in each drink

2. Please explain **why** you chose your answer.

The intention of the revision is two-fold. One is to differentiate effectively between students who rely on calculation (CoS 2A) and those who are able to employ relational thinking (CoS 3D). If a student relied on calculation, s/he would choose D (i.e., it is impossible to tell) because s/he has no numbers to use or attempt to assign numbers to the data points to allow calculation. If a student were a relational thinker, s/he would integrate observation of qualities of the distribution and calculation of statistics, noting that the three outliers would increase the total of measurements and consequently increase the mean. The challenge of this item is to compare the median and mean, which requires the understanding of measures of statistics as measures of distribution (CoS 3C). Although a student might know that "outliers increase the mean" as memorized fact, comparing mean and median requires a more sophisticated conceptual understanding, in particular, that the median and mean measure the center of distribution. Second, students who tend toward calculation but have developed aspects of relational thinking might be encouraged by the revised item to think about whether it is in fact possible to answer the question without numbers, encouraging them to employ relational thinking.

There are three conceptual building blocks for the highest level of performance, predicting how a statistic is affected by changes in components of a distribution. First, students need to notice visual qualities of the distribution such as the clump and the three outliers (CoS 1A). Second, students should know that the median is the middle of the ordered data (CoS 2A). Although there are no numbers, the data points are ordered as they are represented in the line graph. So it is possible for students to find the median. Third, students need to understand the median and mean in relation to the distribution (CoS 3C). For example, students need to understand statistics as measures of distribution in that the mean would be located somewhere in the center clump without the three outliers because it is a measure of center, not just a number produced by formula.

By connecting the three conceptual building blocks, students can reason about the effect of the three outliers on the mean and median and compare them (CoS 3D). Knowing that the outliers will increase the sum of values and consequently the mean (dividend) is useful for estimating the location of the mean and comparing it to the location of the median.

Figure 21. Revised form of Caffeine in Drinks

Establishing mean and median as measures of center. As a first step to support conceptual change, Nancy began discussion by focusing on how to calculate the median when data values are unknown. Students argued strongly that it was impossible to calculate the median "because there're no numbers that tell you," suggesting that they did not make use of the information given by the graph (i.e., ordered data). In response, Nancy asked a content specific question, "Okay, but when we put numbers on here, will this ((pointing at a data point)), ha-, you're, you're saying there's no way to tell if this one's more or this one's more. Which one's more?," to help students see that data points were ordered, therefore providing important information. In this way, Nancy helped her students talk about the magnitudes of the data points without using numbers.

Nancy supplemented this move with further instructional support. In the following excerpt, she asked a student to mark the median on the distribution, then linked the median to attributes of the distribution. This is different from the first year, when Nancy pushed students to find the answer "without calculating it." Nancy was asking students how a student was able to find the median without knowing any values. In this interaction, Nancy helped students understand the meaning of median in relation to qualities of the distribution, which is an important conceptual understanding that supports performance at CoS 3D (Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components).

Excerpt 19

1	T:	So, she's saying here's the median ((pointing the median with the pinky)). Do you guys think the, why would you've been able to guess that the median would be right in here ((Circling around the measurements on the left side of the distribution))?
2	Ss:	Cause that's where most elves[7] are.
3	T:	That's where most of them are. Most of them. So what is the median telling us?
4	S:	The
5	S:	Right there.
6	S:	What's in the middle.
7	T:	The middle or the
8	S:	Center.
9	T:	Center. So.
10	S:	On the data.
11	T:	you're saying that the median is probably gonna be right there in the
12	S:	Center of the data.
13	T:	Center of most of the numbers, right?

The first instructional move was to help students connect the median to the clump. This was mediated by highlighting the clump and median, gesturing at them, and by asking level-specific questions. For example, Nancy asked (line 1), "Why would you've been able to guess that the median would be right in here?" as she circled around the measurements on the left side of the distribution. This is different from the first year, when Nancy highlighted the clump but did not connect it to median.

The second instructional move was that Nancy made sure that students understood the definition of median in relation to the clump (Konold & Pollatsek, 2002) by asking (line 3), "What is the median telling us?" The question revealed that students did not yet understand the median in relation to the distribution, as students just said the

[7] Nancy changed the context of the problem from Caffeine in Drinks to Elves because she conjectured that difficulty with interpreting the graph was due to the unfamiliarity of the problem context.

middle or the center. Nancy emphasized that not only the median was the center of the data, moreover it was most likely located in the center clump (line 13).

These two instructional moves played the role of linking visual qualities of the distribution (CoS 1A) and calculation of statistics (CoS 2A) to a higher level of thinking (CoS 3C: seeing the statistic as a measure of a characteristic of the distribution). Nancy asked students to calculate the median, but also connected the position of the median on the distribution to the clump, which helped students understand the meaning of the median not just as a point but as a measure of the distribution.

Inside and outside of cluster. In talking about the mean, Nancy also related the meaning and calculation of the statistic by asking content specific questions. She asked questions like: (1) What do you know about the mean just by looking at the graph?, intending to draw students' attention to visual qualities of the distribution, (2) Is it gonna be just like the median?, intending to build on the previous agreement that the median is in the middle of the cluster and to prompt students to infer the position of mean on the distribution and (3) What does mean do?, instigating a discussion of the definition of mean as a balance point. Students were largely silent when Nancy asked these questions, suggesting that they found it difficult to reason about these ideas. In response, Nancy's instructional moves were to (1) visualize changes in the mean in a simplified distribution and (2) support students to develop relational language to talk about the mean in relation to the distribution.

Visualizing changes, the first instructional move, was mediated by Nancy's use of an interactive computer program (Figure 22). The computer program calculated median and mean as Nancy dragged data points on the X-axis, helping students to visualize changes in the statistics. The way in which Nancy used the computer program seemed to be very productive for allowing students to explore the relations between median and mean and the distribution.

First, Nancy made visible significant changes in mean and median by manipulating the computer program. Nancy put three data points next to each other, making visible that mean and median were located in the same place as centers of measurements in this particular case. In addition to the visualization, the class agreed that the median and the mean were the same. Next, she put a fourth data point on 100, resulting in a significant contrast from the previous distribution of three points: Here, the median increased a little bit but was still in the cluster of the three data points. However, the mean increased so that it was no longer in the cluster of data points.

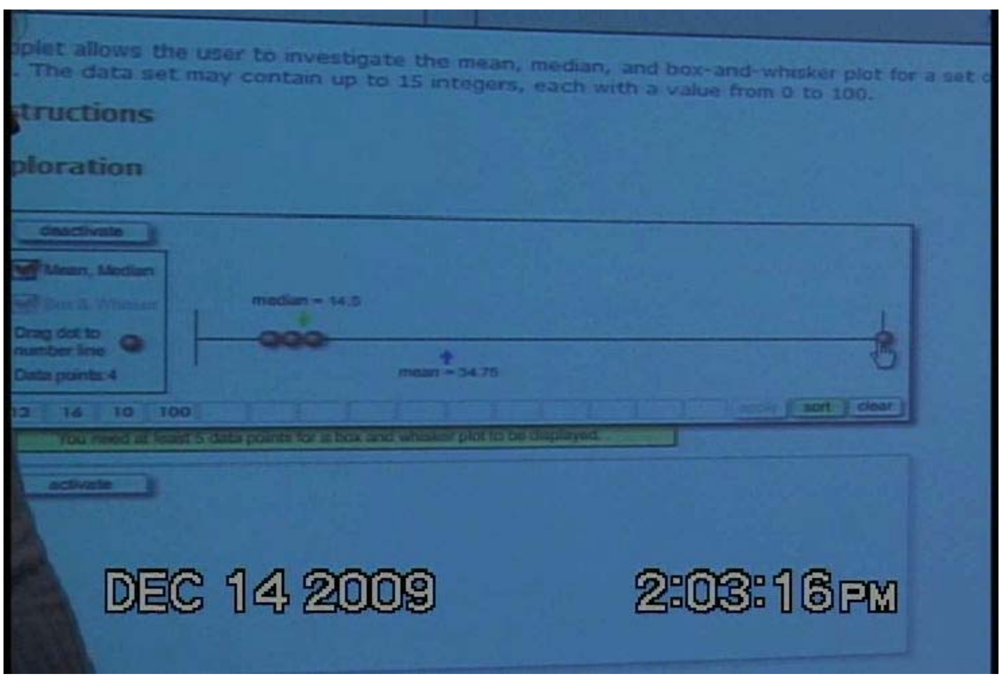


Figure 22. Screenshot of the interactive computer program Nancy used.

Excerpt 20 illustrates how Nancy employed discourse to make use of the representation.

Excerpt 20

1	T:	Let's put it on 100. Okay, where's the median?
2	S:	14.5.
3	T:	The median is still in this little group right here this cluster of data ((circling around the three measurements next to each other)) right? Where's the mean?
4	S:	34.75.
5	T:	It's way outside the cluster.
6		….
7	T:	Why is the mean not in the cluster here? ((showing the interactive computer program))
8	Bob:	Because you've got one that's WAY at 100 points, and then, those are all ((Students talking over each other))
9	T:	Okay, well, let me bring this down a little bit, if that's the problem ((moving 100 toward the three points)).
10	S:	It goes closer to it, but not quite.
11	S:	I know where mean will be. It'll be like; it'll be like on that one, near the cluster, but not in the cluster. It'll be like
12	S:	Near the cluster but close to it.

What is significant about this exchange is the way in which Nancy helped students to build conceptual language to move students from calculating statistics (CoS 2A) to focusing on the relation of the statistic to the distribution (CoS 3C). When she asked where the median was, students said it was on 14.5 (line 2). She then highlighted its relation to the clump, saying (line 3) "The median is still in this little group right here, this cluster of data right?" When discussing the mean, she again highlighted the relation to the clump; as students again read the number (line 4, "34.75"), she said (line 5) "It's way outside the cluster."

Another instructional move was that Nancy treated a student's response as a conjecture and engaged the class to test the conjecture. This instructional move is to leverage students toward CoS 3D (Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components). In line 8, Bob argued that the change was caused by 100. In contrast to Nancy's use of IRE discourse to confirm students' responses in year 1[8], here she used Bob's statement to provide further opportunities for exploration. In line 8, Nancy said, "Okay, well, let me bring this down a little bit, if that's the problem." Students' responses in line 10 to 12 provided evidence that Nancy's instructional move was effective; students started to talk about the mean and median in relation to the clump by using relational language (i.e., closer, near). The students started using the term "cluster" to explain the changes of statistics, indicating that Nancy's instructional support helped the students to talk about mean and median in relation to the important qualities of the distribution.

In addition to using the practices more effectively, Nancy brought in other learning support tools, situating key mathematical ideas in other mathematical contexts to make them more visible to students. All these practices in year 2 were better coordinated to link learning performances toward a learning progression.

Summary of Nancy's Practice in Year 2. Nancy's assessment discussions in Year 2 shows her development of understanding the learning progression of mathematical concepts and the implementation of this knowledge in her assessment talk. Nancy attended to different levels of learning performances possibly elicited by an assessment item (i.e., anticipated learning performances) in year 1. Building on the knowledge about anticipated learning performances, she attended to the image of the

---

[8] [8] Nancy initiated the conversation by asking, "How will it change the mean?" A student responded, "It'll be lower because the three outliers are now replacing further and lower numbers." Then Nancy evaluated, "Did you hear what she called those? She said it's gonna change the mean because the three OUTLIERS are now closer to the center."

progressive development of the anticipated learning performances in Year 2. She gave a deep thought about identifying intermediate learning performances and coming up with instructional moves to link anticipated learning performances and intermediate learning performances in Year 2. In doing so, she elaborated on the assessment item in ways that highlighted the difference between mean and median as measures of center, and helped students relate these measures to visible qualities of the displays.

Summary of Nancy's Naturalization Process of the Assessment System

Nancy demonstrated sophisticated assessment practices and uses of the assessment system in both years, but also made important changes during the two years. Even at the outset of the study, Nancy demonstrated the coordinated use of the assessment system to understand the level of sophistication of students' responses, to identify important mathematical ideas in responses (Coffey et al., 2011), and to inform her classroom practice. When teaching, she made important mathematical ideas visible and promoted conceptual change by asking content-specific questions and highlighting and juxtaposing mathematical ideas that she identified from students' responses.

The interactional structures (e.g., highlighting and juxtaposing) that Nancy employed to support conceptual change were refined in the second year. In Year 1, Nancy highlighted and juxtaposed different levels of students' responses to show students where she wanted them to move toward and pressed students to attain particular learning performances without mediating between current levels of students' understanding and higher levels of mathematical understanding. In Year 2, Nancy often positioned students to discuss different ways of thinking about mathematical ideas by juxtaposing ideas with additional instructional moves. Most importantly, she appropriated the assessment system to come up with line of instructional moves that aligned with a learning progression: connecting anticipated learning performances of an assessment item with intermediate learning performances.

The development of understanding of the learning progression and construction of instructional trajectories seemed to be supported by her use of video annotated construct maps as well as scoring exemplars and paper version construct maps. She stated:

> The one with the broken ruler, and just how they, the students answered the question when they didn't understand. And then the questions that he proposed and the way the other students talked just to, and then when they would understand, you know, and start, and you could see the progression in their thinking and so give you ideas about how to question and think about it. [Post interview, October 2009]

Although there was no affiliated video exemplar for each assessment item, Nancy generalized forms of instructional moves that were attuned to learning progression to her practice. Nancy pointed out that she was able to see how the researcher-teacher's instructional moves (e.g., questioning) supported conceptual change (i.e., moving students from "when they didn't understand" to "when they would understand"). She also highlighted that the exemplar video helped her think about instructional structures (e.g., give you ideas about how to question and think about it) to support students' learning.

Her ongoing development of the sophisticated assessment practices seemed to be related to her routinization (Akkerman & Bakker, 2011) of employing the assessment system in her practice: she developed a routine in which she taught the lessons, tested students' understanding by using the assessment items, then scored students' responses and generated scoring sheets before facilitating assessment talk. Her routinization seemed to support her to engage in constant negotiations with disciplinary perspectives and practices represented in the assessment system and with other participants and researchers at the workshop, providing learning opportunities.

CHAPTER V

CHANGES TO THE ASSESSMENT SYSTEM AS A CONSEQUENCE OF COLLABORATION

Over the course of the collaboration, teachers contributed to the revision of the assessment system. For example, teachers suggested changes in document formats (e.g., using familiar language and contexts for students) and caught errors (e.g., grammar and mismatch between text and representations). Teachers also contributed to the content of the assessment system by providing their own classroom objects and learning activities to be represented in the lessons and scoring exemplars. In addition, student responses from teachers' classrooms replaced the hypothetical responses that researchers used to exemplify levels of performance on the scoring exemplars. Although all elements of the assessment system were revised as teachers and researchers collaborated, the video annotated construct map went through the most significant changes in terms of its functions and forms. The video annotated construct map was originally designed to illustrate discrete level of performances elicited during classroom discussion, but later it showed teachers orchestrating dynamics of learning performances during assessment talk. The changes were motivated by teachers' feedback on the intelligibility of the video annotated construct maps which emerged as they implemented the assessment system to support conceptual development. Here I focus on one significant change that was made to the video annotated construct map in order to better support the transformation of teacher practices.

From Illustrating Levels of Performances to Illustrating a Learning Progression in Action

The video annotated construct map started as a video-annotated illustration of the paper version of construct map that provided more contextual information for teachers to see particular forms of students' reasoning about disciplinary content. The development of the video annotated construct map was motivated to enhance "boundary permeability" (Akkerman & Bakker, 2011, p. 144), considering the nature of teachers' practice. The forms of students' reasoning illustrated in the construct map were very different from skills and performances of traditional mathematics standards and unfamiliar to teachers. Thus, video exemplars were created so that teachers could become familiar with the distinctive students' ways of thinking as they saw them in action. Initially, the video examples were drawn from classrooms led by researchers.

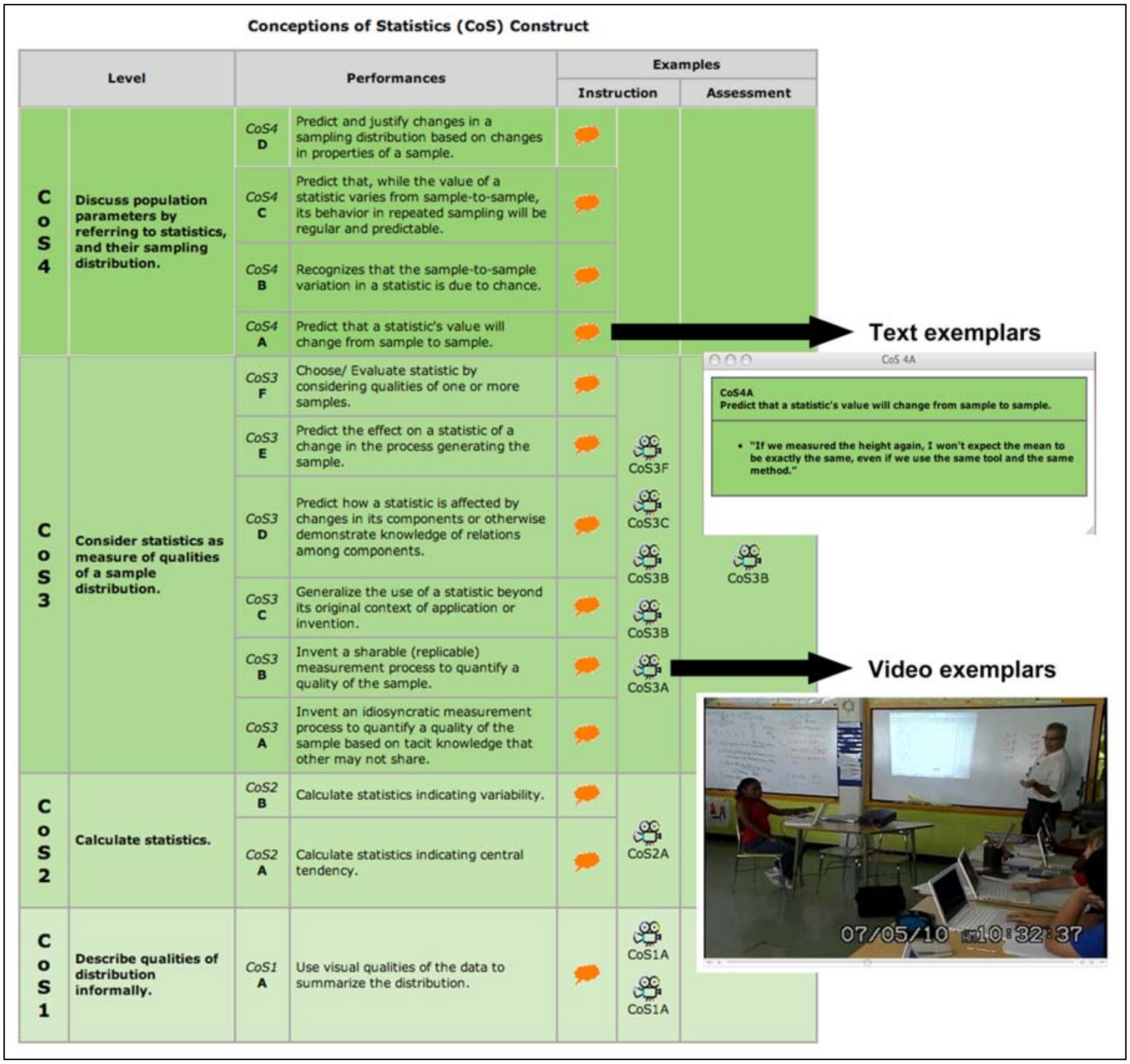


Figure 23. Screenshot of video-annotated construct map of Conceptions of Statistics.

Figure 23 shows one of the video annotated construct maps that were created in the beginning of the collaboration. It had the same structure as the paper version of the construct map but contained both text exemplars and video exemplars. Each performance was exemplified with edited video clips from previous design studies to make the construct maps more accessible to teachers. Some performance levels were illustrated with video clips from both lesson talk and assessment item talk.

When researchers shared the video annotated construct map with teachers, they were very interested in seeing how the activities in lessons were enacted by the researcher-teacher, or what questions the researcher-teachers asked during the course of a lesson. For the researchers, the intention of the video annotation was to illuminate student thinking, with the expectation that teachers would employ these as guides to noticing forms of student thinking as they emerged in teachers' classrooms. However, teachers were more oriented towards understanding the practice of orchestrating classroom talk in ways that leveraged the forms of student thinking illuminated by the video exemplars. For example, Rana said in her interview when she was asked about whether video-annotated construct maps made a difference in how she thought about a mathematical idea,

> …seeing Rich teaches with this body measurement that kind of did help because I knew what questions to ask because I saw him ask, because I saw him ask and responding to the kids so it influenced the way I thought about data displays and how kids think and how, how to move them forward with, with their, with their understanding of displays and how to pull, how to pull the information out of them because of it. [Rana, November 2008]

Rana's interview excerpt suggested that she paid attention to the researcher-teacher's instructional moves (e.g., questions), which was not the original intention of the video-annotated construct maps.

Teachers' feedback about the video-annotated construct maps motivated transformation of forms of the video exemplars. Longer episodes of classroom teaching were incorporated, initially drawn from the design research, into the construct maps. In particular, the video exemplars of assessment talk were expanded so as to promote formative assessment talk. The intention of this form of video exemplar was to make more visible: (1) the levels of sophistication of different forms of reasoning that could be elicited by assessment items beyond “right and wrong” and (2) the instructional effectiveness of assessment items that draw fruitful contrasts among students’ ways of thinking.

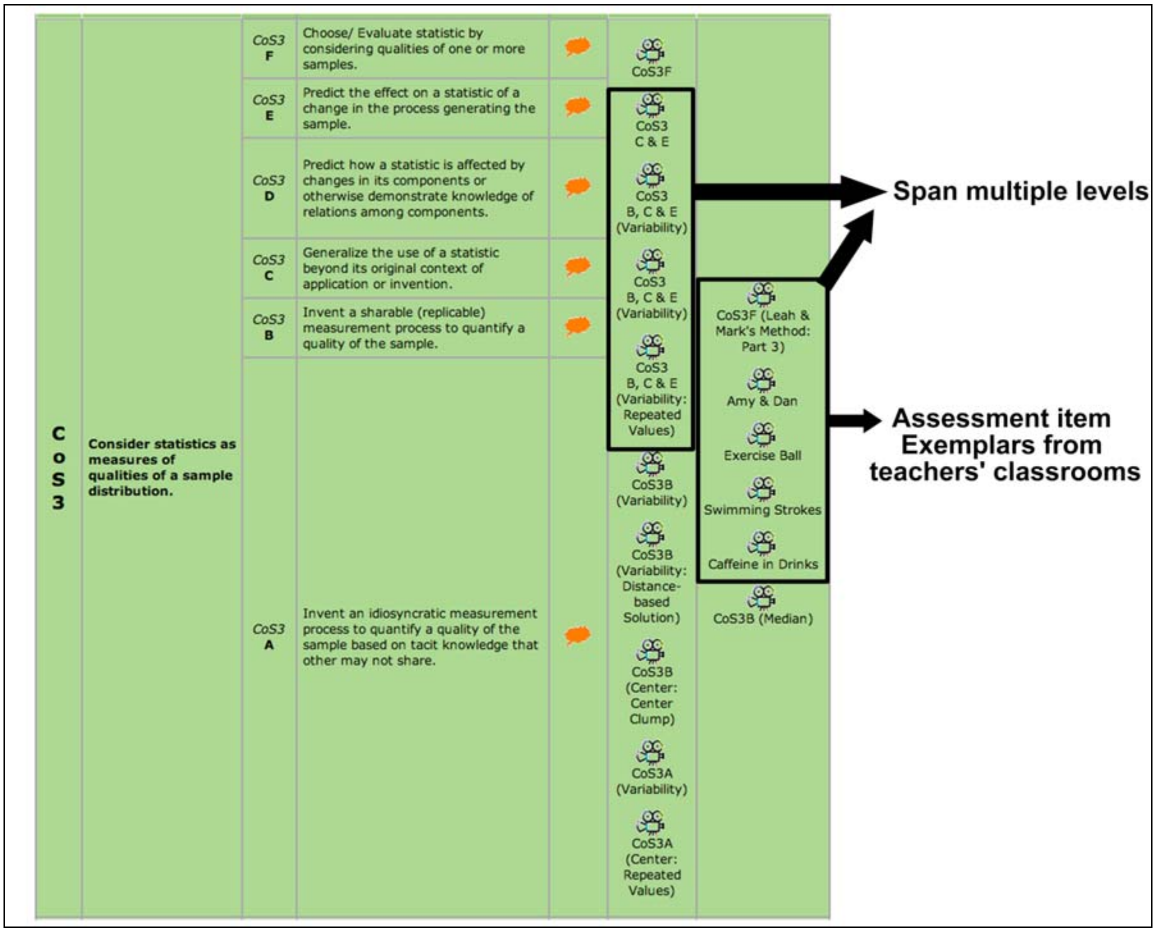


Figure 24. Revised video-annotated construct map.

In response to teachers’ comments in interviews and workshops, the video annotated construct map was revised (Figure 24). This revised version was a hybridized form of two different practices: assessing and teaching. The scope of the video exemplars was increased to illustrate not only the levels of performance elicited during the instructional conversation but also the dynamics of learning performances orchestrated by a teacher's instructional moves. Formative assessment talk video exemplars suggested how the different learning performances were discussed toward more sophisticated conceptions of data and statistics by particular instructional moves (e.g., highlighting and juxtaposing). These revisions transformed the video annotated construct map from illustrating an outcome space of performances to illustrating the integration of the outcome space and its orchestration by teachers.

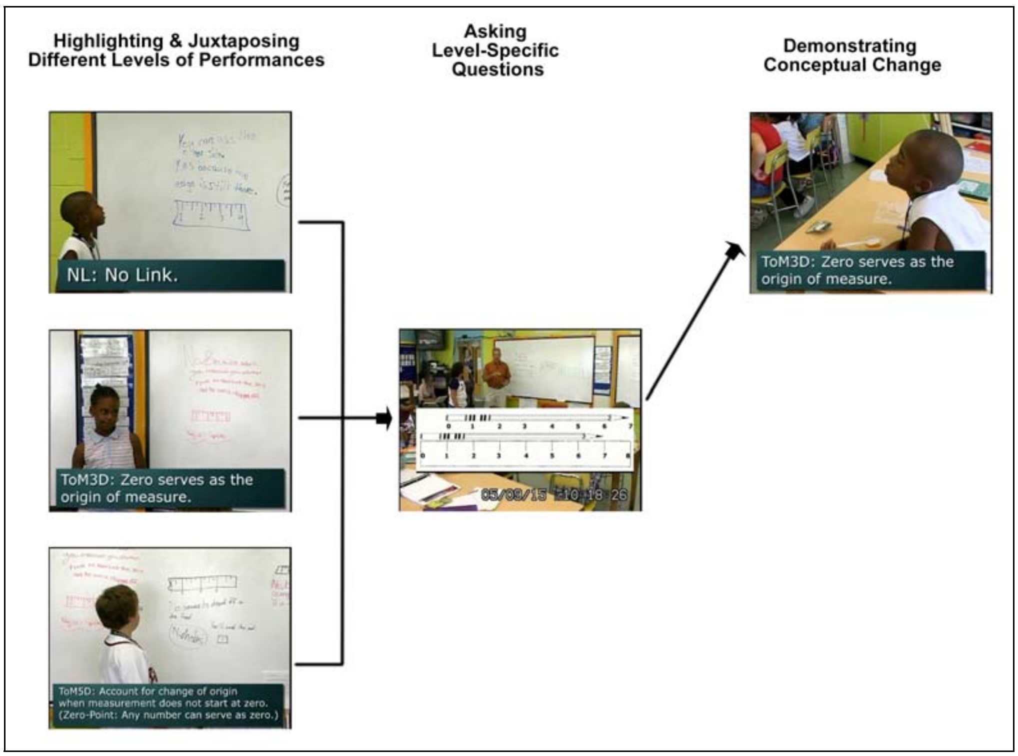


Figure 25. Structure of the video exemplar.

An example of the structure of a video exemplar is illustrated in Figure 25. The revised video examples include labels of students' current states of understanding (e.g., NL: No Link, ToM 3D: Zero serves as the origin of measure) to highlight the relation between students' responses and mathematical significances. They also include teachers' instructional moves, though these are not labeled. In the example in the figure, the teacher asked level-specific questions and used inscriptions to make visible students' thinking and mathematical ideas about the origin of measure. The revised video exemplar shows how a student at a low level of performance moved up to a higher level of performance, illustrating conceptual change. This revised structure of the video exemplar seemed to be visible to Nancy:

> Just how they, the students answered the question when they didn't understand. And then the questions that he proposed and the way the other students talked just to, and then when they would understand, you know, and start, and you could see the *progression* in their thinking and so give you ideas about how to question and think about it. [Post Interview, October 2009]

Nancy noticed the beginning states of students' reasoning, teacher's instructional moves, and then conceptual change evident in students' reasoning, as expressed by her use of the term *progression*.

Teachers' transformed practice was videotaped and made in the form of the video exemplars to be used by other teachers (Crystallization according to Akkerman & Bakker, 2011). Initially, video exemplars were drawn from the design studies, but over time they began to be drawn from the classrooms of participating teachers. As teachers implemented the assessment items and the suggested forms of formative assessment talk, their instruction had interactional structures similar to the one illustrated in Figure 25. These classroom interactions were videotaped and reified as video exemplars. A teacher, Maggie, who watched Nancy's video stated:

> So I watched all of those to kind of guide me and took notes on questions that Nancy asked and ways to extend it or variate it that she did in class that worked well to get the kids to see what happens to the mean and the median, which is why I moved the outliers down closer. She didn't move 'em over the left side but I thought when I got, when I started doing that and moved 'em down and they

> understood the mean, I thought okay let's talk about median, what changes median? [January 2010]

The excerpt suggests that Maggie made use of Nancy's video example to plan her instruction. Not only did she take the questions from the videos, but she also created her own instructional moves based on the video examples.

Summary

The process of how the video annotated construct map was transformed over time exemplifies the fruitfulness of the collaboration between the researchers and teachers and the importance of adapting initial boundary objects to support the ongoing transformation of practices within a community. The construct map in paper format went through several process of transformation: from illustrating learning performances in action to illustrating instructional trajectories in coordination with learning performances. This transformation was intended to accommodate the adaptor's unique practices and support their adaptation of the learning progression to their practices.

CHAPTER VI

INTEGRATION OF TRAJECTORIES OF CHANGE ACROSS THE FOUR CASES AND THE ASSESSMENT SYSTEM

Trajectories of Transformation in Teachers' Assessment Practices

The analysis of classroom observations identified different forms of construct-centered instructional moves, as illustrated in Table 5. Likewise, the analysis of teacher interviews suggested that these different forms of construct-centered instructional moves were mediated by different elements of the assessment system. In this section, I organized the different forms of construct-centered instructional moves in relation to Akkerman and Bakker's (2011) learning mechanisms of boundary objects and the elements of the assessment system to illuminate the roles of specific elements of the assessment system in mediating particular learning mechanisms and particular forms of construct-centered instructional moves.

Table 5. Learning mechanisms and forms of instructional moves in relation to the elements of the assessment system

Learning Mechanism	Coordination	Reflection	Transformation	
Employed Elements of the Assessment System	Lessons and Assessment Items	Lessons, Assessment Items, Scoring Exemplars, and Construct Maps		
Forms of Naturalization	Appropriating to existing practice	Reifying student thinking by scoring exemplars/ construct maps	Making mathematical ideas in student thinking visible	Attuning instructional trajectories to learning progression
Teaching Practices	**Evaluating Student Answers from Right and Wrong Perspective** Teacher focuses on whether student's answer is right or wrong.	**Categorizing Student Thinking in light of Construct Maps or Scoring Exemplars** Teacher looks for students' performances that match with those represented in construct map or scoring exemplar.	**Highlighting and Juxtaposing Significant Mathematical Ideas** This is different from simply saying A said and B said. The instructional move should be followed upon by inviting students to reason about contrast made.	**Asking Level-Specific Questions, Thought-Provoking Questions** Teacher asks construct-related questions to provide disciplinary perspective in response to students' thinking.
	Eliciting Student Thinking by Content-general Questions Teacher elicits student's thinking or justifications, but the driving questions are content-general, such as asking: What did you do?, How did you do it? and why did you do it?	**Showcasing Different Levels of Student Thinking by Students' Responses** Teacher notices different levels of student thinking, but he/she structures presentation of the different levels of student thinking in linear fashion and does not make meaningful connections among them.	**Augmenting the Assessment System** Teacher transforms elements of the assessment items to make key mathematical ideas more visible.	**Linking Different Levels of Performances to Support Conceptual Change** Teacher coordinates multiple levels of students' thinking by employing several instructional moves (e.g., highlighting, juxtaposing, and asking level-specific questions).

Table 5, continued

	Asking Series of Questions that Illustrate Procedural Steps to Get the Right Answer Teacher asks series of questions to illustrate procedural steps, which does not provide opportunities for students to think about mathematical ideas by themselves.			

Teachers' naturalization process of the assessment system to their teaching practices in this table is characterized as follows: (1) appropriating to existing practice, (2) reifying student thinking by scoring exemplars/construct maps, (3) making mathematical ideas in student thinking visible, and (4) attuning instructional trajectories to a learning progression.

The first column consists of how teachers appropriate the elements of the assessment system to existing practice. Sub-categories include: (1) evaluating student answers from right and wrong perspective (e.g., Rana in Year 1 and Catherine in the beginning of her participation in the study), (2) eliciting student thinking by content-general questions (e.g., Rana in Year 1 and Catherine in the beginning of her participation in the study), and (3) asking a series of questions that illustrate procedural steps to get the right answer (Rana in Year 1).

Evaluating student answers from a right or wrong perspective suggests that teachers do not pay much attention to students' reasoning and that they have a traditional perspective on viewing students' responses. Eliciting students' different ways of thinking about a mathematical concept and their justifications is characterized as an important index of high leverage practice (NCTM, 2000). However, teachers mainly employ content-general questions such as "What did you do? How did you do? and Why did you do that?" and do not go further beyond getting students' responses public (M. L. Franke et al., 2009). In addition, these content general questions do not usually uncover the details of student thinking about mathematical ideas.

The discourse patterns seen in this type of practices are IRE (Mehan, 1979) or turn-taking (Ball, 1993; M. Franke, Kazemi, & Battey, 2007), which mostly illustrate procedural steps to get the right answer.

The first column falls into *Coordination* in terms of Akkerman and Bakker's (2011) learning mechanism. Some teachers (e.g., Rana in Year 1 and Catherine in the beginning of her participation in the study) mainly used lessons and assessment items with loose alignment with the classification system of the learning progression. This resulted in different interpretations of intended mathematical ideas of lessons and assessment items. It still allowed teachers to meet their job requirements but results in less disruption in existing practices.

The second column consists of teacher's practices that reify student thinking by using the suggested classification system (i.e., construct maps and scoring exemplars), which led to change in teachers' perspective (*Reflection* in Akkerman & Bakker's term). Sub-categories include: (1) categorizing student thinking in light of construct maps or scoring exemplars (e.g., Nancy in Year 1, Rana in Year 2, and Catherine later in her participation in the study) and (2) showcasing different levels of student thinking by students' responses (e.g., Rana in Year 1 and Catherine early in her participation in the study).

The analysis showed that when the teachers (e.g., Rana, Catherine and Nancy) routinized their coordinated use of the classification system with the curriculum materials, they developed a mathematical disciplinary perspective. The practice of reification seems to be critical in transforming how one orchestrates assessment talk. Rana and Catherine demonstrated significant changes in their teaching practices when they started categorizing student thinking based on the classification system. When the teachers had assessment talk in their classrooms, their classifying work seemed to be used for different instructional ends. For example, Rana started to believe that "wrong answers" or lower levels of student performances could be used to support students' learning. However, she seemed not to be sure yet what instructional moves would be effective in doing that. As the teachers developed deeper understanding of relations between students' expressions and mathematical ideas, they demonstrated more sophisticated forms of instructional moves, as described in the third and fourth columns.

The third and fourth columns consist of assessment practices that indicate how teachers' changes in classifying students' responses influence teachers' orchestration of assessment conversations that support conceptual changes in student thinking. The two columns fall into *Transformation* in Akkerman and Bakker's framework (2011). Transformation has been noted as the most difficult learning mechanism to enact, one requiring rigorous intervention (Akkerman & Bakker, 2011).

The third column consists of practices that some teachers used to make significant mathematical ideas in students' responses visible. These teachers seemed to be clearer about the seeds of disciplinary ideas evident in what students said and did. The subcategories include: (1) highlighting and juxtaposing significant mathematical ideas (e.g., Rana in Year 2 and Nancy in Year 1 and 2) and (2) augmenting the assessment system (e.g., Catherine later in the study and Nancy in Year 1 and 2). The teachers looked for mathematically significant performances of students that matched with those represented in construct maps or scoring exemplars. The teachers made key mathematical concepts visible by highlighting (Goodwin, 1994) or revoicing student thinking (O'Connor & Michaels, 1993) that was related to performances on the classification system. The teacher purposefully made connections among different levels of students' thinking to make mathematical ideas more visible. These practices were possible because the teachers could identify important mathematical ideas from students' responses in light of the classification system. The teachers also augmented the assessment system. For example, they transformed elements of the assessment items to make key mathematical ideas more noticeable.

Finally, the fourth column illustrates assessment practices that attune a line of instructional moves to learning progression. This category refers to the practice of attuning instructional moves not only to particular learning performances but also to the larger picture that those learning performances depict. Teachers who use this practice consider learning performances as landmarks that students exhibit on the way to understanding the "big idea" of each construct and are able to situate learning performances in a continuum moving toward this big idea. Sub practices within this category include: (1) asking level-specific questions (e.g., Catherine later in the study and Nancy in Year 1 and 2) and (2) linking different levels of students' performances with an eye toward conceptual changes (Nancy in Year 2). Asking level-specific questions is a way for teachers to guide students from their current states of student thinking towards a more complex disciplinary perspective. Connecting students' performances with an eye toward conceptual changes is to make several instructional moves that orchestrate different levels of students' performances toward higher levels of thinking. The teacher coordinates multiple levels of students' thinking by productively employing several instructional moves (e.g., highlighting, juxtaposing, and asking level-specific questions).

Comparing Four Cases in terms of Naturalization of the Assessment System

In this section, I summarize significant patterns of how teachers adapted the assessment system and changes in their assessment practices. And I discuss both the common and unique aspects of the cases. Then I relate the patterns to the teachers' feedback on the intelligibility of the elements of the assessment system in order to find evidence for how the elements of the assessment system mediated the changes. Table 6 illustrates each teacher's changes in perspective and practice in terms of the forms of construct-centered instructional moves. A circle (●) indicates an instructional move that a teacher demonstrates. A double circle (⦿) indicates an approximation of an instructional move.

Table 6. Changes in Perspectives and Practices of Each Case Teacher

Learning Mechanisms		Coordination			Reflection		Transformation			
Employed Elements of the Assessment System		Lessons & Assessment Items			Lessons, Assessment Items, Scoring Exemplars, and Construct Maps					
Forms of Naturalization		Appropriating to existing practice			Reifying student thinking by scoring exemplars/ construct maps		Making mathematical ideas in student thinking visible		Attuning to learning progression	
Teaching Practices		Evaluating ST as “right and wrong”	Eliciting ST by content-general questions	Illustrating procedural steps	Categorizing ST by the classification system	Showcasing ST	Highlighting and Juxtaposing mathematical ideas	Augmenting the assessment items	Asking level-specific questions	Building toward higher levels of thinking
Theresa	Pre		●		⦿	●				
	Post		●		⦿					
Rana	Pre	●	●	●		●				
	Post			●	●		⦿			
Catherine	Pre	●	●			●				
	Post				●			●	●	
Nancy	Pre				●		⦿	●	●	
	Post				●		●	●	●	●

Mathematical Disciplinary Perspective

The most significant change in teachers’ perspectives was that teachers came to align their perspectives with mathematical disciplinary ideas. When teachers started participating in the study, they tended to coordinate their perspectives with traditional school mathematics or seek ways to adapt the assessment system to meet school accountability requirements. As illustrated in *Table 6*, Rana and Catherine fell into more toward “evaluating student thinking as right and wrong” in the beginning of their participation in the study. For example, Rana, in Year 1, classified students’ responses to assessment items into binary categories. She used her classification to select students who would present correct problem solving strategies in front of the class. This suggests that she used assessment items in coordination with her existing dichotomous perspective. She did not want to discuss wrong answers because she believed that it would confuse students. In the case of Catherine, she used a scoring exemplar to assess students’ responses, but she converted different levels of performances into a dichotomy (right and wrong) when she used the classification system. This suggests that Catherine did not see the kinds of mathematical ideas that the scoring exemplar intended to highlight in “wrong” answers. When asked about the effectiveness of lessons in engaging students in learning, Catherine explained that they helped her meet her district guideline of classroom discussion: 75% of students’ talk and 25% of teachers’ talk. She appropriated the curriculum materials to meet school accountability requirements rather than seeking mathematical ideas. In the case of Theresa, she applied a rough categorization of mathematical ideas to students’ responses rather than specific categories of the constructs, which fell into “an approximation to categorizing student thinking by the classification system.”

As teachers negotiated their preexisting perspectives with the one presented in the classification system, they demonstrated changes in how they viewed students’ responses to assessment items. Table 6 illustrates that Rana and Catherine categorized student thinking by the classification system" later in their participation in the study. Although Nancy categorized student thinking by the classification system throughout her participation in the study, there seems to be qualitative differences in categorizing students' thinking. My analysis identified further distinctive forms of teachers' alignment with the mathematical discipline as teachers categorized students' responses by the classification system: (1) anticipating particular forms of students' reasoning and (2) developing understanding about relations between students' expressions and mathematical ideas.

Anticipating Particular Forms of Students' Responses. The classification system helped teachers identify particular forms of students' responses they could anticipate. The anticipation helped teachers notice learning performances represented in the classification system and the mathematical significances of students' responses. As an example of illustrating a rudimentary form of *anticipating*, Theresa described in post interviews that she shifted her instructional goal from traditional school mathematics standards (e.g., elements of conventional graphs such as key and title) to big ideas of mathematics (e.g., data structure expressed in graphs). She did not talk explicitly about the categories of the constructs, rather she had rough categories in sense-making of the data. Accordingly, she sought and elicited the big ideas of mathematics during classroom discussion. Rana became more aware of what students were lacking in light of her anticipation of forms of mathematical reasoning. Rana selected significant responses of students based on her anticipation of particular forms of responses during assessment talks. Nancy read scoring exemplars before class and came with images of students' responses that she could expect. And she was able to locate her students' responses on the classification system. In contrast to Theresa's rough description about the categories of the constructs, Nancy mentioned specific levels of performances of the constructs.

Teachers were able to elicit particular forms of students' reasoning. Also they centered classroom discussion around mathematical substance. This was accomplished through identifying particular forms of students' reasoning by using the classification system.

Developing Understanding about Relations between Students' Expressions and Mathematical Ideas. This form of practice refers to teachers' interpretation of students' responses beyond noticing what students did. In other words, teachers were concerned what student's response implied about his or her mathematical understanding. My analysis suggests that the classification system helped teachers make close connections between mathematical disciplinary ideas and students' ways of expressing them. For example, at the end of the study, Catherine reified students' responses in terms of the classification system and questioned her students' logic behind their answers: "Why are they coming up with this answer, what could make them think in this way?" As an early adaptor of the classification system, Nancy did not ever demonstrate a dichotomous perspective. However, she reported in Year 2 that the scoring exemplar helped her differentiate distinctions in mathematical understanding (e.g., "using proof from the display") from students' ways of expressing mathematical ideas (e.g., "more" vs. "12 here and 8 here"). She refined her categorization of students' responses from "anticipating particular forms of students' reasoning" to "developing understanding about relations between students' expressions and mathematical ideas." This form of interpreting students' responses seemed to inform construct-centered instructional moves, as discussed in the next section.

Transforming Interactional Structure in Relation to a Learning Progression

The analysis suggests that the assessment system, especially the classification system, supported teachers to conduct construct-centered pedagogical practice. Since teachers were able to identify more mathematical substance beyond a right answer, they can tailor their instructional moves specifically to the mathematical substance.

A transformed I-R-E or turn-taking pattern or a hybrid of the two discourse patterns were commonly used by Theresa, Rana, and Catherine in early participation in the study. Table 6 shows that Theresa, Rana and Catherine demonstrated combinations of "eliciting by content-general questions," "illustrating procedural steps," and "showcasing student thinking" in the beginning of the study. These discourse patterns were aligned with the teachers' perspectives on students' responses (e.g., right or wrong and different strategies to get a right answer).

Teachers tended to enact a turn-taking pattern when they recognized multiple strategies to get a right answer. For example, when Rana talked about the item *Height of a Plant*, she let students share different strategies to get a right answer. The different strategies represented different sophistications in conceptual understanding, but she did not push students toward a higher level of understanding. Her evaluation criteria of the different strategies were distant somewhat from specific disciplinary understanding (e.g., whether students could justify their answers or not). Another situation of enacting a turn-taking discourse pattern was when a teacher tried to implement reform oriented mathematics instruction. Catherine often employed a turn-taking pattern to provide opportunities for the whole class to hear other students' thinking and construct explanations based on peer's thinking. Catherine tried to refrain from telling students a right answer because of her belief about learning, but often this saving did not provide students enough instructional supports.

Several teachers employed a hybrid of a transformed I-R-E and turn-taking discourse pattern as a way to accommodate both traditional school mathematics and mathematical discipline perspective. This pattern was observed when teachers started to anticipate particular forms of students' responses. For example, Theresa in Year 1 knew that certain forms of students' responses indicated better understanding about data display. She elicited higher levels of performances by employing Thought-Revealing Questions in lessons. When students provided types of answers that Theresa anticipated, she communicated very strongly that students were right, a reminiscent of I-R-E. Students took turns to share their reasoning about data structure, but Theresa did not make any instructional moves to connect the students' thinking elicited by a turn-taking discourse pattern.

The analysis indicates that teachers developed construct-centered instructional moves to support students' conceptual change during their use of the classification system: (1) tailoring instructional moves to current states of students' understanding, (2) coordinating students' responses across multiple levels for productive learning, and (3) attuning the instructional trajectory to learning progression.

Tailoring Instructional Moves to Current States of Students' Understanding. The analysis indicates that the classification system supported teachers to figure out how to act on current states of students' understanding toward the next achievable levels of mathematical understanding. For example, Catherine, in her early participation in the study, did not notice mathematical ideas in students' responses (e.g., putting 19 after 30 and describing shape of distribution as "spike") and did not respond with content specific instructional moves. However, at the end of the study, she took construct-centered instructional moves in response to students' thinking. In doing so during the assessment talk about *Two Spinners*, Catherine first made connections between students' responses and the big idea of chance, repeated process. When Catherine heard Jamie's response (Excerpt 14), she realized that Jamie was not thinking about the repeated process of chance, identified as foundational by the construct map. She then asked level-specific questions to help Jamie consider repeated process of chance. As another example, Nancy asked a series of content-specific-questions (Excerpt 19 and Excerpt 20) that guided students to engage with mathematical ideas that they were not able to identify or consider by themselves.

Another instructional move tailored to current states of students' understanding was to transform assessment item to make mathematical ideas visible. For example, Catherine altered the representation of the assessment item, *Two Spinners*, to help students reason about a compound event. In contrast to asking content general questions, these forms of content-specific instructional moves seemed to remain classroom discussion focused on mathematical substance during assessment talk.

Coordinating Students' Responses across Multiple Levels for Productive Learning. The analysis also indicates that the classification system helped teachers coordinate multiple levels of understanding during assessment talk. Teachers enacted the practices of highlighting and juxtaposing students' responses to let students compare and contrast different ideas. This practice should be differentiated from teachers asking students simply to compare their answers. Instead, here teachers deliberate about which parts of students' responses should be highlighted and which responses should be juxtaposed to make the most effective comparisons.

This form of transformation seems to have variations in its relationship with forms of mathematical perspective that teachers develop over time. When teachers had some image of *anticipated particular forms of students' responses*, they enacted the practice of highlighting and juxtaposing. However, they focused on presenting higher level of thinking through highlighting and juxtaposing. For example, Nancy, in Year 1, highlighted Tobi's noticing about the distribution and juxtaposed Tobi's and Kristine's strategies as a way to pinpoint a better strategy to solve the problem. Rana also evidenced in Year 2 that she focused on communicating differences in levels of students' strategies. She knew how students would respond to certain assessment items (e.g., Two Spinners) and levels of sophistication in students' responses. Rana described different levels of performance as "staircase":

> If you're familiar with the different kinds of responses you might get, you could put `em in order, talk about `em right then, and then, move the kids up a little staircase of understanding, and you can just do it right then. It's, it's nothing. I mean, it's really simple to do. [Rana, April 2010]

Rana seemed to believe that making visible the learning progression by representing all levels of performances during a whole class discussion would help students learn. So her role as a teacher was to elicit different levels of students' performance. As a way of doing so, Rana employed approximations of highlighting and juxtaposing, unpacking different levels of students' responses, including wrong answers.

In contrast, *understanding relations between students' expressions and mathematical ideas* seemed to facilitate coordinating students' responses on multiple levels in more productive way. As contrasting examples, Rana (Excerpt 12) and Nancy in Year 2 (Excerpt 17) both demonstrated highlighting and juxtaposition when they discussed the assessment item, *Two Spinners (Figure 2)*. However, the teachers selected different combinations of students' responses to be compared. Rana juxtaposed possible outcomes by spinning one spinner versus spinning both spinners simultaneously. In contrast, Nancy juxtaposed students who considered looking at the structure of the two spinners versus who thought about enacting two spinners simultaneously in creating total outcome spaces. Nancy made a more strategic contrast that would make significant conceptual differences in students' ways of thinking more visible to other students.

Attuning Line of Instructional Trajectory to a Learning Progression. This form of practice refers to the practice of aligning instructional trajectory with learning progression. Not only does a teacher need to understand individual levels of mathematical ideas, but also has to understand how the individual levels of mathematical idea fit into a progressive pathway toward understanding a big idea of mathematics. This entails: First, a teacher has to identify the distribution of her students' current states of understanding in terms of learning progression. Secondly, a teacher has to identify intermediate conceptual building blocks that were not expressed by her students, but yet are part of the learning progression. Then, a teacher makes instructional moves that mediate current states of students' understanding and targeted understanding. Teachers do not necessarily need to have an image of a progressive pathway of developing a big idea to enact identifying, highlighting, juxtaposing, and asking content-specific questions, because the minimum requirement is to recognize particular learning performances from students' responses. In contrast, attuning instructional moves to address the larger learning progression requires teachers to be able to assemble learning performances into a learning progression.

The analysis indicated that teachers' instructional trajectories often were not coordinated with the learning progression. For example, in Theresa's assessment talk about *Range* in Year 2, her students identified outliers and clump (CoS 1A and DaD 3A) but they did not understand measures of spread in relation to distribution (CoS 3C). As Theresa tried to push students toward understanding the effect of components of distribution to statistics (CoS 3D: predict how a statistic is affected by changes in its components), she explained the mathematical concepts for students. She did not constitute her instructional trajectory with instructional moves to help students move from CoS 1A to CoS 3D. As another example, Nancy noticed a mathematical idea (CoS 1A) and highlighted the important mathematical idea (Excerpt 3). In attempting to push students' understanding from CoS 1A to CoS 3D, she did not support intermediate conceptual building blocks of CoS 2A (Calculating statistics) and CoS 3C (Understanding statistics as measures of center). Rather Nancy ended up calling on students who already understood the mathematical concepts (CoS 3D) and asked them to explain for the class. In contrast, Nancy demonstrated the practice of attuning an instructional trajectory to the learning progression in Year 2. She supported students to move toward CoS 3D by juxtaposing, asking specific questions, and making links between different levels of performances.

This form of structuring classroom interaction in relation to learning progression is the most sophisticated form. This study suggests that understanding the learning progression is a key to enacting this form of classroom interaction. It leads a teacher to ask fruitful content specific questions, to evaluate and pinpoint states of students' understanding, and to make instructional moves in light of prior elicitation and diagnosis.

Relating Variations in Assessment Practice to Teacher's Ratings of the Intelligibility of the Elements of the Assessment System

The analysis of the cases suggested the classification system played a critical role in transforming teachers' formative assessment practices. Here I present an analysis of teachers' ratings of the intelligibility of each element of the assessment system in their teaching practice, providing further evidence of the correlation between ways of using the assessment system and changes in teachers' assessment practice.

Teachers provided different ratings on the curriculum materials (see Figure 26) and the classification system (see Figure 27). Their ratings suggest that teachers perceived each element of the assessment system to have different implications for their teaching practice. Figure 26 shows the case teachers' ratings of the intelligibility of lessons and assessment items in relation to their teaching practice, as they responded in the last interviews conducted with them (see Appendix II for the survey questions).

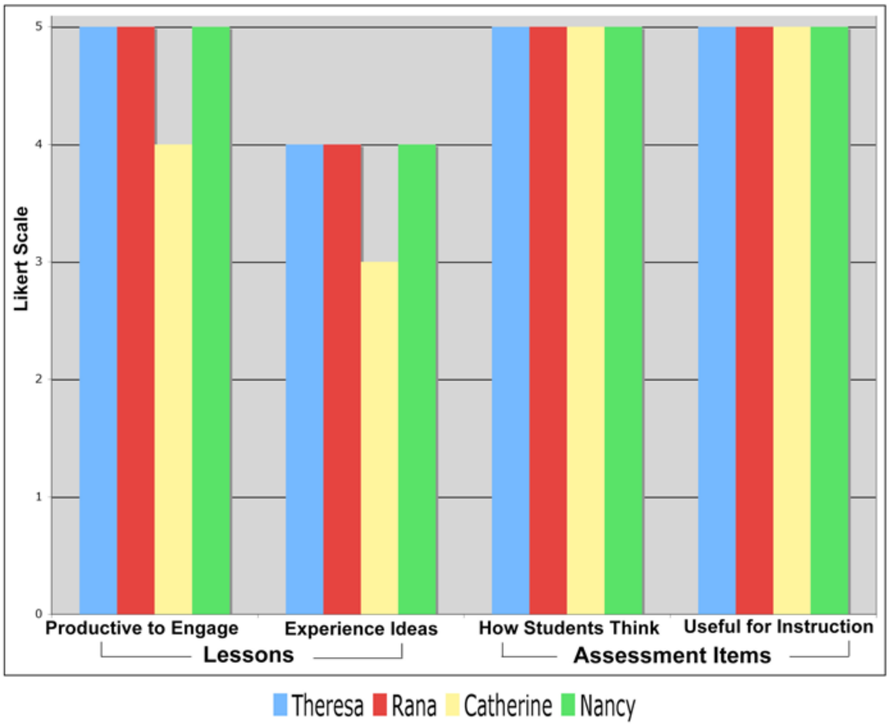


Figure 26. Intelligibility of the assessment system to teachers' teaching practice.

Close inspection of the figure reveals that the four case teachers agreed about the intelligibility of the curriculum materials to their practice: They rated "agree or strongly agree" on questions concerning the intelligibility of lessons and of assessment items for teaching practice (e.g., lessons show productive ways to engage students, lessons show how students experience ideas, assessment items show how students think and are useful for instruction). Particularly, the four case teachers all strongly agreed that the assessment items were useful for instruction and for revealing how students think. These ratings coincide with findings from classroom observations and interviews. All four case teachers readily used the curriculum materials in their classrooms.

In contrast, the case teachers responded more diversely regarding the classification system, as illustrated in Figure 27.

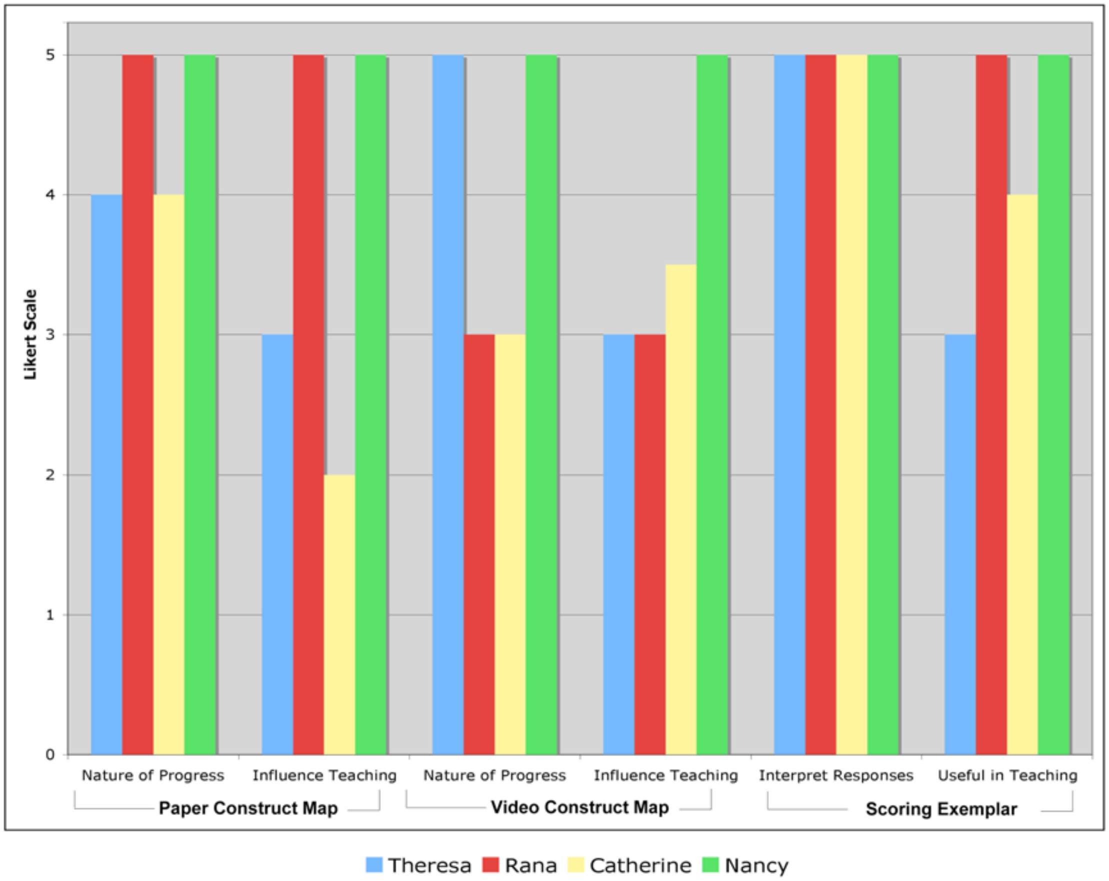


Figure 27. Teacher's ratings of the intelligibility of classification system.

The case teachers all strongly agreed that scoring exemplars were useful to interpret students' responses, as the scoring exemplars were intended as an assessment tool. However, it is noteworthy how the case teachers responded to questions concerning the usefulness of the classification systems for teaching. Nancy, who attuned her instructional trajectory to the learning progression, strongly agreed that the classification system was useful and influenced her teaching. She rated all categories as "strongly agree." These ratings suggest that her sophisticated form of instructional moves was supported by the extended use of all different types of the classification system. Rana, who illustrated learning progression in action by presenting students' responses, strongly agreed that the paper construct maps and scoring exemplars were useful in teaching. It is conjectured that these two components of the classification system supported her in identifying students' responses in association with each level of performance. Catherine, who acted on particular levels of understanding but did not yet juxtapose them, rated scoring exemplars high, presumably because they were useful for interpreting students' responses. Finally, Theresa, who reverted back to her existing practice, said she could not decide on usefulness of the classification system in her teaching.

It is also conjectured that the classification system contributed to variations in forms of teaching practice. The case teachers expressed most disagreement in relation to the paper and video annotated construct maps' usefulness for teaching. The average rating of three teachers (Theresa, Rana, and Catherine) on the influence of paper version construct maps to teaching was 3 and that of video annotated construct maps was 3. In contrast, Nancy strongly agreed that both paper and video construct maps influenced her teaching. This suggests that, unlike the other teachers, Nancy found implications for the different types of the classification system in her teaching practice.

Moreover, teachers' ratings of "influence on teaching of video annotated construct map" showed the most variation in teachers' responses. Their ratings on the survey question seemed to correlate with how they integrated the learning progression in their teaching. Also, the teachers' recall of particular episodes in the video annotated construct map suggests that they inferred different implications about teaching from the map.

Theresa seemed to look for logistics of implementing lessons and coordinating assessment talks in video exemplars. Theresa responded that she was undecided on the video annotated construct map's impact on teaching. She elaborated on her rating:

> Because there again it's just this [video annotated construct map] is not first nature or second nature this is you know I still have to deliberately think about this. [Post Interview, April 2010]

Theresa reported that she generally did not think about looking at video annotated construct maps when she planned classroom discussions. However, when she did look at video annotated construct maps in preparation for her class, she seemed to look for specific video exemplars for specific learning activities or assessment items. She said:

> I watched some of the others that were about the same things, but … I couldn't relate that back to those lessons, you know I couldn't figure out how to use that information to you know, an alternate for these. [Theresa's Post Interview, April 2010]

Theresa did not seem to be able to relate students' ways of thinking about measures of spread in a learning activity to a different activity about measures of spread. She described herself as "stealer of Thought-Revealing-Questions" in Year 1, and she was not able to use the video annotated construct map to focus on student thinking. Instead, she looked for video exemplars that were directly related to an activity or questions that she could directly import into her classroom.

Rana rated the survey item, video annotated construct maps influence my teaching, as "undecided" and described the implication of the video annotated construct maps to her teaching practice.

> I like to see that for the teacher, how the teacher questions … it's helped me see how the teachers kind of restraining themselves and don't, do a lot less instruction and just a lot more questioning. So, it's made me question the kids more than just actually talking, letting the kids teach each other. [Rana's Post Interview, April 2010]

What she learned from the video annotated construct maps was a more general sense of what teachers should do (e.g., "do a lot less instruction and just a lot more questioning and letting the kids teach each other") rather than connecting general instructional moves to content-specific instructional moves. Her interview excerpt supports the classroom observation that Rana was approximating juxtaposing and highlighting. She facilitated the process of seeing differences in students' responses and provided students opportunities to share their ways of thinking, as she described as "do a lot of less instruction and just a lot more questioning." Rana seemed to view teacher questioning as a tool to encourage students to talk more. In contrast, Catherine's view on questioning seemed to be more focused on specific content. Catherine stated:

> When I looked at those video clips trying to think, "Well, is there something on here, did I do, or what should I do next?" I heard something that he asked a student who was demonstrating, you know and I thought, "Oh, I should've said that, I bet that would have brought up," you know, cause sometimes when you're going on the fly, the questions you want just aren't there. [Catherine's Post Interview, April 2010]

Catherine was looking closely at the teacher's content-specific question to a student's answer, as illustrated in the video annotated construct maps.

Nancy, who demonstrated the most effective orchestration of classroom discussion, strongly agreed that the video annotated construct maps and paper construct maps influenced her teaching. Nancy stated:

> Just how they, the students answered the question when they didn't understand. And then the questions that he proposed and the way the other students talked just to, and then when they would understand, you know, and start, and you could see the *progression* in their thinking and so give you ideas about how to question and think about it. [Nancy's Post Interview, October 2009]

Nancy pointed out that she was able to see how the researcher-teacher's instructional moves (e.g., questioning) supported conceptual change (i.e., moving students from "when they didn’t understand” to “when they would understand”). She also highlighted that the exemplar video helped her think about instructional moves (e.g., give you ideas about how to question and think about it) to support students’ learning.

In sum, the analysis suggests that the variations in forms of mathematical disciplinary perspectives and instructional moves are related to coordinated and extended use of the elements of the assessment system. Particularly, the analysis indicates that the most sophisticated forms in perspective and practice demonstrated by Nancy were supported by the coordinated and extended use of the different forms of the classification system.

CHAPTER VII

DISCUSSION

In this dissertation, I described how the researcher-created assessment system as a set of boundary objects mediated the collaborative efforts between teachers and researchers in reorienting assessment toward improving the quality of instruction and supporting student learning. The analysis suggests that the assessment system coordinated the collaboration by providing focal points around which the two professional groups negotiated their interpretations of the conceptual development of statistical reasoning. More importantly, the analysis provides evidence that the assessment system enacted the learning mechanism for reflection, supporting teachers in developing new perspectives: understandings of the big ideas of data, chance and statistics and of the learning progressions of statistical reasoning. In addition, the assessment system supported the teachers in transforming assessment practices in their classrooms[9]. The teachers demonstrated construct-centered orchestration of assessment talk: structuring classroom interaction centered on important mathematical ideas represented in the classification system and/or aligning the instructional trajectory with the learning progressions to support student learning.

This study also illustrated the process of naturalizing learning progressions to the teachers' daily practices. Within the case teachers' changes in perspectives, there were variations in terms of noticing and interpreting mathematical disciplinary ideas expressed in students' verbal- or written-responses (i.e., anticipating particular forms of students'

[9] For transformation of researchers' practice, see Lehrer et al. (2011).

responses and developing understanding about relations between students' expressions and mathematical ideas). In relation to changes in practices, this study identified variations in forms of orchestrating levels of students' performances (i.e., tailoring instructional moves to current states of students' understanding, coordinating students' responses across multiple levels for productive learning, and attuning a line of instructional trajectories to a learning progression). The variations in forms of adapting learning progressions seemed to be mediated by different elements of the assessment system. The analysis suggests that the coordinated use of the curriculum materials with the classification system was critical in adapting the assessment system for improving instruction.

Finally, the elements of the assessment system (e.g., video-annotated construct maps) were transformed to coordinate the collaborative effort, to support transformation of professionals' practices more effectively, and to reify the transformed teachers' assessment practices.

In this chapter I discuss implications of the analysis and findings I have presented. First, I will discuss implications of coordinating collaboration via researcher-created objects, linking the findings to the theoretical framework of boundary objects. Secondly, I will discuss implications of the findings in supporting teachers to enact formative assessment discussions as a specialized form of dialogue to make conceptual progress. I will then discuss continuing challenges and future work.

The Role of the Assessment System as a Set of Boundary Objects

This study's conceptualization of the researcher-created classification system as a set of boundary objects may provide practical implications in emergent collaborative efforts between researchers and teachers around learning progression. By employing the theoretical framework of boundary objects, this study traced the process of naturalization of the classification system in teachers' daily teaching practices. The findings can inform the design of boundary objects that can mediate ongoing collaborations between teachers and researchers.

Coordinating Collaboration

The study suggests that a learning progression can constitute a medium where researchers and teachers are able to coordinate their collaboration around a shared goal within and across their boundaries of practice. More importantly, this study suggests that enabling a learning progression to function as a productive boundary object requires significant attention to the nature of the adaptor's job requirements. For example, in our work, we made efforts to translate results from research practice into teacher-friendly objects (e.g., lessons and video-annotated construct maps) by using formats that were consistent with teachers' daily job requirements. It was intended to increase boundary permeability (Akkerman & Bakker, 2011) of the classification system. The analysis indicates that the teachers' experiences with the curriculum materials encouraged the teachers to reflect on their existing perspectives and practices. For example, Theresa in Year 1 described how she came to understand core disciplinary ideas of data display as she used Thought-Revealing-Questions in the lesson. Although it was not part of this dissertation, teachers' preexisting traditional school mathematics perspectives were disrupted when the researchers engaged them in negotiating their experiences with the curriculum materials at the workshop. Teachers' experiences with the curriculum materials in their classrooms supplied substance for negotiation as teachers saw similar forms of student reasoning in their classrooms when they enacted learning activities. The negotiations were centered on supporting teachers to reinterpret their classroom experiences in light of the classification system.

Supporting to Develop a Disciplinary Perspective of Mathematics

This study suggests that learning progressions as a classification system can be an effective tool to disrupt the historically developed classificatory system for assessment in modern schooling (i.e., right or wrong) and eventually overwrite it with a disciplinary perspective on mathematics. The analysis of the data suggests that the learning progression centered classification system disrupted teachers' preexisting traditional school mathematics perspective (e.g., dichotomous perspective in viewing students' responses shown by Rana and Catherine in the beginning of their participation in the study) over time. The disruption by the classification system supported the teachers to develop their understanding of mathematical ideas, toward one better aligned with a mathematical disciplinary perspective. For example, at the end of the collaboration, Rana developed some images of anticipated students' responses and was able to notice them as she conducted assessment talk. As a more sophisticated form of aligning with the discipline of mathematics, Catherine and Nancy demonstrated their development of understanding about relations between students' expressions and mathematical ideas.

In addition, this study identifies a trajectory of constructing discipline-oriented perspectives that teachers may go through as a possible process of naturalizing a learning progression. In this study, the case teachers developed different depths of understanding of mathematical disciplinary ideas and of making connections between mathematical ideas and forms of students' expressions of the ideas (e.g., anticipating particular forms of students' responses, developing understanding about relations between students' expressions and mathematical ideas). The different degrees of coordinating mathematical ideas and students' mathematical logic can be understood as the process of naturalization in relation to their experience with the classification system. Ideally, one might expect a process of naturalization that Nancy demonstrated in this study. Nancy provides evidence of the different degrees of understanding in terms of her trajectory of developing a perspective on the classification system: she started by anticipating particular forms of students' responses in Year 1, then developed the most sophisticated understanding about the relationship between mathematical disciplinary knowledge and students' expressions of knowledge in Year 2.

Transforming Practices

The analysis of this study provides empirical evidence that a learning progression can support teachers to orchestrate construct-centered assessment talk. The forms of discursive practices involved coordination of their students' current levels of understanding with the learning progression, with an eye toward guiding students' attention to significant mathematical substance and positioning students to evaluate and investigate disciplinary mathematical ideas. They highlighted and juxtaposed different levels of students' responses in order to make them into a discursive substance for class discussion, asked level-specific questions to provide students alternative disciplinary perspectives to consider, transformed the initial forms of assessment items to make significant mathematical ideas more observable by students, and objectified prospective conceptual pathways in action built upon current states of students' understanding. These instructional moves occurred when the teachers used the classification system to monitor students' progress in conceptual understanding, suggesting the critical role of the classification system in organizing construct-centered instruction. This was in contrast to when they positioned themselves as evaluators of students' work using an I-R-E discourse pattern as seen in the beginning of the collaboration. In sum, the classification system began to function as a mathematical horizon that supported teachers' efforts to orchestrate productive mathematical conversation about assessment items.

Co-constitution of Transformation of Boundary Objects and Practices

This study suggests how to transform learning progressions to make them accessible and usable by teachers and what should be considered in the transformation to support adaptors to open up the black box (e.g., unpacking of meanings of the classification). In this study, transformation of the elements of the assessment system was guided by teachers' feedback after they used them in their daily practices. Initially, the researchers created the video annotated construct maps to illustrate discrete levels of performances elicited during classroom discussions. This version of the video annotated construct maps was intended to fulfill the same function as the paper version construct maps, helping teachers distinguish different levels of students' performances in action.

The translation kept the original nature of the classification system intact, maintaining the identity of the learning progression across the communities of different professionals, but facilitating localization of the learning progression for the teachers. The interactions at the workshops suggested that the transformation of the video annotated construct map was necessary to provide teachers information that supported their practices (illustrating a practice of formative assessment). As it was transformed from “weakly structured in common use” (Bowker & Star, 1999, p. 297) to “strongly structured in individual-site use,” the video annotated construct maps helped the teachers think about how to orchestrate students’ answers (e.g., questions to ask) as well as how to anticipate and interpret possible student answers. Once transformed, the video exemplars illustrated not only levels of performance elicited during the instructional conversation but also the dynamics of learning performances orchestrated by teachers’ instructional moves. Also, assessment talk video exemplars were added to illustrate how the employment of interactional structure was exploited for conceptual change when the teachers tailored instructional moves in response to substantial mathematical ideas expressed by students.

This study demonstrates that adaptors of boundary objects can participate in the creation of the boundary objects as they contribute materials from their practices. The collaboration resulted in crystallizing teachers’ practices into video exemplars that were embedded in the video annotated construct map. This suggests that the teachers’ position were changed from users of the boundary objects to creators of the boundary objects.

Implications for Mathematics Education

Teaching practices, mathematical content knowledge, and assessment systems have tended to be researched separately, but recently national reform documents highlight and problematize this separation and encourage the use of learning progressions as a tool for coordinating them (NRC, 2005). The study provides empirical evidence that learning progressions can be effective tools to coordinate assessment and instruction centered on important mathematical ideas in moments of classroom interaction.

First, this study shows how teachers can be supported in connecting formative assessment practice with disciplinary perspectives and in moving beyond instructional tactics and strategies. This study shows that learning progressions can support teachers to construct “discipline-relevant criteria” (Coffey et al., 2011, p. 1131) and remain focused on mathematical substance. The analysis illustrated that the classification system was a critical resource for the case teachers to develop a discipline-specific perspective on evaluating students’ responses: to notice mathematical substances that students expressed and to interpret the students’ current states of understanding in relation to the mathematical horizon.

This study challenges educators to move beyond some of the classification systems suggested by some reform efforts in mathematics education. For example, Franke et al. (2009) used “correct and complete, ambiguous or incomplete, and incorrect” to characterize qualities of students’ explanation/justification. Although these categorical systems are more descriptive about students’ reasoning than “right or wrong,” they are still at the level of content-general criteria. As another example, Jacobs et al. (2010) used “robust evidence, limited evidence, or lack of evidence” to differentiate their teachers’ ways of noticing, interpreting, and deciding-how-to-respond to students' responses, but these criteria are very subjective and content-general. The field needs to develop more content-specific classification systems to inspect qualities of students' reasoning and teachers' interpretations of students' reasoning. The analysis here evidenced that the learning progression centered classification system supported the teachers to develop a discipline-specific classification on evaluating students' explanation/justification. For example, in this study, Catherine demonstrated her belief that learning mathematics was to be able to construct explanations at the beginning of her participation. However, she was not specific about qualities of explanation/justification. Catherine came to characterize students' justification/explanation in terms of levels of understanding on the learning progression, as evident at the end of her participation.

Secondly, pushing beyond improving formative assessment practice in terms of aligning with disciplinary ideas, the study shows that formative assessment should be part of instruction for effective learning, and vice versa. This study evidenced that the classification system became a tool for the case teachers to coordinate assessment and instruction, resonating with Ball's pedagogical practices; "with my ears to the ground, listening to my students, my eyes are focused on the mathematical horizon (Ball, 1993, p. 376)." The case teachers demonstrated that not only did they focus on mathematical ideas in students' responses, but they also paid attention to connections among students' different levels of thinking and to mathematical disciplinary ideas. The instructional moves demonstrated by the case teachers were constituted with particular levels of performances of statistical reasoning constructs (e.g., asking level specific questions, juxtaposing different levels of performances). Noticing different forms and levels of students' responses and understanding progressive development of these forms and levels seemed to support the teachers to make productive instructional decisions during moments of interaction. One particular instance illustrating this point is Nancy's first year practice, when she depended primarily on students' self-reporting assessment, in comparison to her second year practice. In the first year, she was not sure if her students understood the item, but was not able to come up with level-specific questions to figure out where her students' understanding fell on the spectrum of learning progressions. In Year 2, her instructional moves illustrated that she constantly assessed the current state of students' understanding and kept the formative assessment discussion on the appropriate mathematical horizon (Ball, 1993) to leverage students' current states of understanding toward higher ones. This constant evaluation of current states of students' understanding seemed to help Nancy to identify the mathematical horizon and to make instructional moves accordingly, illustrating the coordination among mathematical ideas, assessment, and instruction as mediated by the classification system. This study meets a call for reforming formative assessment practice in classrooms.

Challenges and Future Work

This study contributes to the emergent research about learning progressions. In particular, as an early study, it provides empirical evidence as to how teachers adapted the learning progression based assessment system about statistical reasoning to inform their teaching practices and how the process of adaptation was supported. The analysis illustrated that the learning progression centered instructional moves provided more fruitful learning opportunities for students. However, this study did not test students' achievement in relation to these learning opportunities. Additional research is needed to provide evidence of the connections between learning progression-based instruction and students' achievement.

In addition, future work needs to consider ways to accelerate the process of naturalizing the classification system to transform teachers' perspectives and practices. A transformed video-annotated construct map is expected to speed up teachers' adaptation of the learning-progression centered instruction: integration of instructional trajectories and an learning progressions. Researchers should explore more efficient ways to integrate the transformed video-annotated construct maps to professional development programs.

APPENDIX I

WORKSHOP AGENDAS

Arkansas Workshop
October 15, 2009

Agenda

Material	Name cards for participants Rulers (15 cm. ruler &1 meter stick) Big post-it paper / markers/ Sticky notes DaD & MRC Construct Maps: Visual, Text & Multimedia Construct Maps Quiz 1 Quiz 1 Item exemplars Lesson 1: Body Measure Thumbnail sketches of students' displays Computers/ Speakers

Measuring Task & Scoring ToM Items (8AM – 9:20AM)	Introduce the measuring task • What would students do when they measure? • Try out of repeated measurement of one person's arm-span with a 15cm. ruler and with a meter stick & put measurements on sticky notes Score 5 Theory of Measurement items using exemplars while waiting for a turn to measure the length of Rich's arm-span
Group Discussion about Scoring & Items (9:20AM – 10AM)	Group discussion 1. Do the scoring exemplars make sense? • How is each scoring exemplar intelligible? • What is the relationship between scoring exemplar and construct maps? • Which items do seem to work to advance instruction?
Break (10-10:15AM)	
Construct Maps (10:15 AM – 11:30 PM)	Analyze students' displays (PPT) 2. What would students do with the data? 3. What are students noticing about the structure of the data?

	4. What does it show and what does it hide? Introduce Data Display & MRC Construct Maps • What would progress look like when representing data? • What would progress look like when comparing displays? Locate students' displays on Display Construct map
Lunch Break (11:30AM – 12:20PM)	Min-Joung gets feedback about Multimedia ToM.

Classroom Video (12:20PM – 1:10PM)	Formative Assessment: Homemade Bowling item in action
	• Watch the video clip **without** subtitles:
	  • What do you notice about the item? What is it trying to test?
	  • What do you notice about students' thinking?
	• Watch the video clip **with** subtitles:
	  • Was this form of the video helpful? If so, how?

Lesson & Quiz (1:10PM – 1:40PM)	Read Lesson 1
	
	Read Quiz 1
	• For each item, decide what each item might assess and predict a range of student performances.
	  • Using items for instructional purpose
	  • What logistics does it require to teach with items?

Break (1:40PM – 1:50PM)	
FADS (1:50PM – 2:30PM)	Formative Assessment Delivery System

APPENDIX II

INTERVIEW QUESTIONS

Reflection

Pre instruction:

1. Which assessment items are you planning to use for instructional purpose? Why did you choose these items?
2. What did you learn as you scored students' responses based on the scoring exemplars?

Post instruction:

1. What do you think about the math class? Did anything surprise you?
2. What would a student have to know about measurement to correctly answer a question like this one [Ask for each item]? What about students' thinking did you notice about this item? What difficulties did you notice that students have when they solve the question? How did you help the student?
3. Was there anything about students' thinking that you wanted to explore more?
4. Have you seen any changes in students' thinking today?

Perception of Boundary Practice

1. If another teacher were to ask you what this collaboration between teachers in Northwest Arkansas and Vanderbilt is all about, what would you say?
2. What do you like the most about the partnership?
3. What do you like the least about the partnership?
4. Is there any part of your experience with the partnership that you would like to see continue? If so, what and why? Probe: What kinds of factors would facilitate that (what you just described) continuing, and what do you think the barriers would be?
5. I am going to ask you about the workshops and instructional materials we provided.

    a. What do you think of the workshops on [mmddyy]? Was there anything at the workshop that you found particularly helpful and if so, how? Anything for you was a waste of time or should be changed?
6. How do you get to know about the data modeling workshop?
7. What did you have in mind when you decided to participate in the data modeling workshop?

8. What did you expect?
9. Why did you decide to come?
10. Have your original goals been achieved? Do you have new goals as a result of participating in the workshop?
11. What do you think are your roles or responsibilities in the collaboration between Vanderbilt and Northwest Arkansas?
12. If you were inviting other teachers, who would you invite? Why?
13. How would you tell them what is about the workshop?
14. If you were to invite someone near, what would you tell them strengths about the workshop if you wanted them to learn about?
15. What would you tell them weaknesses about the workshop if you wanted them to learn about?
16. Does the workshop align with your teaching practice or requirements from your school district?
17. How does the workshop conflict with your teaching practice or requirements from your school district? How do you handle the conflicts?
18. If we continue this workshop series next year, do you plan to attend the workshop?
    a. [If no] Why would you not participate?
    b. [If yes] What do you want to happen in the workshop?
    c. Which topics might be most helpful to focus on?

Perception of Boundary Objects

When we work with you, we provide several different kinds of materials. Please rate your response on each item from one to five. One is strongly disagree and five is strongly agree.

Lessons:

Lessons suggest productive ways of engaging students in learning:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Lessons help me think about how students might experience or reason about mathematical ideas:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

I use some or all of the lessons in my classroom:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Please tell of a time when you recall that the lessons made a difference in your classroom.

Assessment items:

Assessment items help me see how students are thinking:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Assessment items are useful for instruction:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Please tell of a time when you recall that the assessment items made a difference in how you thought about a mathematical idea and/or how students might think of that idea or how you taught (if any specific memories come to mind). If nothing specific comes to mind, that's OK. We'll just move on the next one.

Paper version construct maps:

Paper version construct maps help me see the nature of progress:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

The progression outlined influences my teaching:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Please tell of a time when you recall that the paper version construct maps made a difference in how you thought about the nature of progress of a mathematical idea and/or how students might think of that idea (if any specific memories come to mind). If nothing specific comes to mind, that's OK. We'll just move on the next one.

Video-annotated construct maps:

Video-annotated construct maps help me see the nature of progress:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

The progression outlined influences my teaching:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Please tell of a time when you recall that the video-annotated construct maps made a difference in how you thought about a mathematical idea and/or how students might think of that idea (if any specific memories come to mind). If nothing specific comes to mind, that's OK. We'll just move on the next one.

Exemplars:

Exemplars' helpfulness on interpretation of students' responses:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Exemplars are useful in teaching:

Strongly disagree	Disagree	Undecided	Agree	Strongly agree
1	2	3	4	5

Can you elaborate on your response to the question? Why are you "Strongly agree"? What make you decide that?

Please tell of a time when you recall that the exemplar made a difference in how you thought about a mathematical idea and/or how students might think of that idea (if any specific memories come to mind). If nothing specific comes to mind, that's OK. We'll just move on the next one.

Please tell me what you most and least like about each material we provided.

a. Lessons
b. Text Construct maps
c. Multimedia Construct maps
d. Assessment items
e. Scoring exemplars

We provided many items. Which items do you like best? Why?

Perceptions of Mathematics

1. Has your participation in the partnership between Vanderbilt and teachers in Northwest Arkansas changed your knowledge of, or the way you think about math and/or science?

- *[If no]* Is there any more that you want to say about this topic?

- *[If yes]* Tell me a little about what kinds of changes you have experienced and tell me how those changes have occurred--what has supported them or caused them?

2. Has your work in this project helped you think about how students' "reason" about data, statistics, chance, and measurement?

- *[If no]* Is there any more that you want to say about this topic?

- *[If yes]* Would you tell me what this experience has been like for you?

3. How, practically, have you used information about student thinking/ knowledge in instruction?

4. What do you think the big idea of [measurement, data display, conceptions of statistics, chance]? How do you think students develop the big idea of [measurement, data display, conceptions of statistics, chance]? How would you teach the big idea of chance [measurement, data display, conceptions of statistics, chance]?

5. Have you experienced changes in what you know about how students think about ideas in data and statistics as you participated in the workshops?

   - [If yes] Would you tell me what this experience has been like for you?
   - [If no] Is there any more that you want to say about this topic?

6. We have shared six construct maps that we thought they are all related to data and statistics. We talked about them separately, but as you taught the lessons, have you seen any relationships among the construct maps?

7. Which construct maps, if any, were most helpful to you? Why?

8. When you think about teaching your students, which concepts or ideas about data, statistics and chance seem most important to help them learn? Which are least important?

9. According to the construct map, what are some important changes in how students reason about [measurement, Conceptions of statistics, chance, data display]? Is the construct map's view of how reasoning changes and develops consistent with your experience? Are there parts of it that you doubt?

APPENDIX III

CONCEPTIONS OF STATISTICS CONSTRUCT MAP

Level			Performances	Examples
CoS4	Investigate and anticipate qualities of a sampling distribution.	*CoS4* **D**	Predict and justify changes in a sampling distribution based on changes in properties of a sample.	▪ Students predict that the variability of a sampling distribution of the median will change if the sample size is decreased from 30 to 3 and explain why (grade 5 FP study).
		CoS4 **C**	Predict that, while the value of a statistic varies from sample-to-sample, its behavior in repeated sampling will be regular and predictable.	▪ "If we measure the teacher's arm span again and again, we will get different means and spreads. However, they will not be very different next time." ▪ "If we tested another 75 Type A batteries, I would expect the median to be similar to the median we got this time and around this area (between 165 and 175)."
		CoS4 **B**	Recognize that the sample-to-sample variation in a statistic is due to chance.	▪ "From sample to sample, the medians change, just by chance."
		CoS4 **A**	Predict that a statistic's value will change from sample to sample.	▪ "If we measured the height again, I won't expect the mean to be exactly the same, even if we use the same tool and the same method."
CoS3	Consider statistics as measures of qualities of a sample distribution.	*CoS3* **F**	Choose/Evaluate statistic by considering qualities of one or more samples.	▪ "It is better to calculate the median because this data set has an extreme outlier. The outlier increases the mean a lot." ▪ "The estimate of the mean will be better if we can increase the number of cases, because the mean measures central tendency and more cases increases our confidence in this tendency."
		CoS3 **E**	Predict the effect on a statistic of a change in the process generating the sample.	▪ "If we use a more precise measurement tool, our spread number will get smaller." ▪ "The average deviation of rates of change of fast plants will increase as the plants grow, because of the growth spurt."

	CoS3 *(continued)*	*CoS3* **D**	Predict how a statistic is affected by changes in its components or otherwise demonstrate knowledge of relations among components.	▪ "If we increase the highest value, the mean will change, but the median will not." ▪ "If I know the mean and all but one of the data values, I can find the missing value."
	Consider statistics as measures of qualities of a sample distribution.	*CoS3* **C**	Generalize the use of a statistic beyond its original context of application or invention.	▪ "Nick's measure of spread works because when the data get more spread out, it increases." ▪ Students use average deviation from the median to explore the spread of the data across multiple samples.
		CoS3 **B**	Invent a sharable (replicable) measurement process to quantify a quality of the sample.	▪ "In order to find the best guess, I count from the lowest to the highest and from the highest to the lowest at the same time. If I have an odd total number of data, the point where the two counting methods meet will be my best guess. If I have an even total number, the average of the two last numbers of my two counting methods will be the best guess."
		CoS3 **A**	Invent an idiosyncratic measurement process to quantify a quality of the sample based on tacit knowledge that others may not share.	▪ "In order to find the best guess, I first looked at which number has more than others and I got 152 and 158 both repeated twice. I picked 158 because it looks more reasonable to me."
CoS2	**Calculate statistics.**	*CoS2* **B**	Calculate statistics indicating variability.	▪ "We found the range by subtracting the minimum value from the maximum value."
		CoS2 **A**	Calculate statistics indicating central tendency.	▪ Students calculate mean, median, and mode when they are given a set of data and put these as labels in their displays. However, they may not understand that each is a measure of central tendency.

CoS1	**Describe qualities of distribution informally.**	*CoS1* **A**	Use visual qualities of the data to summarize the distribution.	▪ "There is a big clump here." ▪ "The measurements are really spread out." ▪ "The majority is in the middle." ▪ "The real value might be where most of measurements are."

REFERENCES

Akkerman, S. F., & Bakker, A. (2011). Boundary Crossing and Boundary Objects. *Review of Educational Research, 81*(2), 132-169.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The elementary school journal, 93*(4), 373-397.

Ball, D. L., & Forzani, F. M. (2011). Teaching Skillful Teaching. *Educational Leadership, 68*(4), 40-45.

Bickel, W. E., & Hattrup, R. A. (1995). Teachers and researchers in collaboration: Reflections on the process. *American educational research journal, 32*(1), 35-62.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education, 5*(1), 7-74.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The elementary school journal, 97*(1), 3-20.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C.-P., & Loef, M. (1989). Using Knowledge of Children's Mathematics Thinking in Classroom Teaching: An Experimental Study. *American Educational Research Journal, 26*(4), 499-531.

Cobb, P., & McClain, K. (2006). The Collective Mediation of a High-Stakes Accountability Program: Communities and Networks of Practice. *Mind, Culture, and Activity, 13*(2), 80-100.

Cobb, P., McClain, K., Lamberg, T. d. S., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational researcher, 32*(6), 13-24.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom matehmatical practices. *The Journal of the Learning Sciences, 10*(1 & 2), 113-163.

Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching, 48*(10), 1109-1136.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*(6), 1047-1085.

Davis, E. A., & Krajcik, J. S. (2005). Designing Educative Curriculum Materials to Promote Teacher Learning. *Educational Researcher, 34*(3), 3-14.

Denzin, N. K., & Lincoln, Y. S. (2000). The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 1-28). Thousand Oaks, CA: SAGE Publications.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, Massachusetts: The MIT Press.

Fennema, E., Carpenter, T. P., & Franke, M. L. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403-434.

Franke, M., Kazemi, E., & Battey, D. (2007). Mathematics Teaching and Classroom Practice. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225-256). Reston, VA: National Council of Teachers of Mathematics.

Franke, M. L., Webb, N. M., Chan, A. G., Ing, M., Freund, D., & Battey, D. (2009). Teacher Questioning to Elicit Students‚Äô Mathematical Thinking in Elementary School Classrooms. *Journal of Teacher Education, 60*(4), 380-392.

Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J., Ayala, C., Brandon, P., Shavelson, R., et al. (2008). On the Fidelity of Implementing Embedded Formative Assessments and Its Relation to Student Learning. *Applied Measurement in Education, 21*(4), 360-389.

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. New York: Routledge.

Goodwin, C. (1994). Professional vision. *American Anthropologist, 96*(3), 606-633.

Hall, R., Stevens, R., & Torralba, T. (2002). Disrupting Representational Infrastructure in Conversations Across Disciplines. *Mind, Culture, and Activity, 9*(3), 179-210.

Hallinan, M. T. (1996). Bridging the gap between research and practice. *Sociology of education, 69*(Extra Issue), 131-134.

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., Rowan, B., & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal, 42*(2), 371-406.

Jacobs, V., Lamb, L. C., Philipp, R., & Schappelle, B. (2009a). *Professional noticing of children's mathematical thinking*. Paper presented at the American Educational Research Association Annual Meeting.

Jacobs, V., Lamb, L. C., Philipp, R., & Schappelle, B. (2009b). *Responding on the basis of children's mathematical thinking*. Paper presented at the American Educational Research Association Annual Meeting.

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional Noticing of Children's Mathematical Thinking. *Journal for Research in Mathematics Education*, *41*(2), 169-202.

Klingner, J. K., Ahwee, S., Pilonieta, P., & Menendez, R. (2003). Barriers and facilitators in scaling up research-based practices. *Exceptional children, 69*(4), 411-429.

Konold, C. (1989). Informal conceptions of probability. *Cognition & Instruction, 6*(1), 59-98.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal of research in mathematics education, 33*(4), 259-289.

Lagemann, E. C., & Shulman, L. S. (1999). *Issues in education research: problems and possibilities*. San Francisco: Jossey-Bass Inc.

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

Lehrer, R., & Kim, M.-J. (2009). Structuring variability by negotiating its measure. *Mathematics education research journal*, *21*(2), 116-133.

Lehrer, R., Kim, M.-J., Ayers, E., & Wilson, M. (in press). Toward establishing a learning progression to support the development of statistical reasoning. In J. Confrey & A. Maloney (Eds.), *Learning over time: Learning Trajectories in Mathematics Education*. Charlotte, NC: Information Age Publishers.

Lehrer, R., Kim, M.-J., & Schauble, L. (2007). Supporting the Development of Conceptions of Statistics by Engaging Students in Measuring and Modeling Variability. *International Journal of Computers for Mathematical Learning, 12*(3), 195-216.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition & Instruction, 14*(1), 69-108.

Lehrer, R., & Schauble, L. (2000a). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*(1&2), 51-74.

Lehrer, R., & Schauble, L. (2000b). Modeling in mathematics and science. In R.Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 101-159). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635-679.

Mehan, H. (1979). "What Time Is It, Denise?": Asking Known Information Questions in Classroom Discourse. *Theory into Practice, 18*(4), 285-294.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progressions for carbon cycling in socio-ecological systems. *Journal of research in science teaching, 46*(6), 675-698.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*(3), 722-730.

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: The national council of teachers of mathematics.

Nolen, S. B., Horn, I. S., Ward, C. J., & Childers, S. a. (2011). Novice teacher learning and motivation across contexts: Assessment tools as boundary objects. *Cognition and instruction, 29*(1), 88-122.

O'Connor, M. C., & Michaels, S. (1993). Aligning Academic Task and Participation Status through Revoicing: Analysis of a Classroom Discourse Strategy. *Anthropology & Education Quarterly, 24*(4), 318-335.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2&3), 131-156.

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

Shaughnessy, J. M. (1997). *Missed opportunities in research on the teaching and learning of data and chance.* Paper presented at the MERGA 20, Aotearoa.

Simon, M. A., & Tzur, R. (1999). Explicating the Teacher's Perspective from the Researchers' Perspectives: Generating Accounts of Mathematics Teachers' Practice. *Journal for Research in Mathematics Education, 30*(3), 252-264.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of research in science teaching, 46*(6), 610-631.

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of research in science teaching, 46*(6), 699-715.

Stein, M., & Coburn, C. (2008). Architectures for Learning: A Comparative Analysis of Two Urban School Districts. *American Journal of Education, 114*(4), 583-626.

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

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building Student Capacity for Mathematical Thinking and Reasoning: An Analysis of Mathematical Tasks Used in Reform Classrooms. *American Educational Research Journal, 33*(2), 455-488.

Sztajn, P., Confrey, J., Wilson, P. H., & Edgington, C. (2012). Learning Trajectory Based Instruction. *Educational Researcher, 41*(5), 147-156.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. New York, NY: Cambridge University Press.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Erlbaum.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of research in science teaching, 46*(6), 716-730.

Wilson, M., & Bertenthal, M. (2005). *Systems for State Science Assessment*. Washington DC: National Academy Press.