Title
Improving Instruction: An Examination of a Network Improvement Science Effort to Support Instructional Change

Author
Hill, Mark Benjamin

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Improving Instruction: An Examination of a Network Improvement Science Effort to Support Instructional Change

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Education

by

Mark Hill

2019

ABSTRACT OF THE DISSERTATION

Improving Instruction: An Examination of a Network

Improvement Science Effort to Support Instructional Change

by

Mark Hill

Doctor of Education

University of California, Los Angeles, 2019

Professor Mark P. Hansen, Co-Chair

Professor Karen Quartz, Co-Chair

Improvement science is a promising framework for school efforts at improving classroom instruction. However, there is scant documentation on actual attempts to apply improvement science principles to better K-12 teaching practices. This research attempts to fill that gap by reporting the results of a two-year improvement science professional development effort undertaken by a five-school network, explicitly focusing on secondary math teaching practices. Through interviews with nine of 29 participating teachers, as well as the facilitators, there was clear consensus about which improvement science principles did and did not support the network's learning efforts. Observations were conducted over the two-year period. Document analysis included the Plan-Do-Study-Act (PDSA) logs of the participating teachers, as well the network's measures for improvement. Teachers reported positively on several improvement science principles. These include keeping their work problem-specific and user-centered and engaging disciplined inquiry. Additionally, teachers reported value in the distinct improvement

science principle of framing their efforts in terms of systems thinking and the concept of using measures for improvement to collect data on the ability of their efforts to impact important drivers of effective classroom instruction. However, both facilitators and teachers expressed concern about the ability to create accurate, responsive, and common measures for improvement in order to inform their decision-making. Additionally, the teachers reported intense time/bandwidth concerns about gathering and using measures for improvement to inform changes in their work processes. Finally, facilitators allowed teachers to change their driver of focus and to use qualitative, rather than quantitative, data to inform their PDSA cycles. Teachers appreciated this autonomy, but this decision hindered the network's ability to test a common hypothesis informed by data on a shared measure. These findings have important implications for any educational organization attempting to use improvement science principles, particularly measures for improvement, in an effort to reform classroom instructional practices.

This dissertation of Mark Hill is approved.

Louis M. Gomez

James W. Stigler

Mark P. Hansen, Committee Co-Chair

Karen Quartz, Committee Co-Chair

University of California, Los Angeles

2019

Dedication

To Monica, your endless support and belief lifted me up when I felt unequal to the task;

my mother, my role model, your integrity and love for education inspired me;

and my father, the first person to call me Dr. Mark Benjamin Hill.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I could not have finished this dissertation without the belief and support of my co-chairs, Mark Hansen and Karen Hunter Quartz. Their thoughtful feedback motivated me to produce a product of which I am extremely proud. The work felt Sisyphean at times, but Mark and Karen continually assured me that I was making progress. By addressing their insightful critique, my dissertation became lean and focused, shedding any fluff that was unclear or unsupported. I am truly humbled by the amount of time and effort Mark and Karen put into this effort. I cannot begin to describe how much I have learned thanks to their subject matter expertise.

I also want to thank the teachers and facilitators of the network I observed for this study. Educators are being analyzed more than ever and it takes courage to allow a researcher into your professional development effort, watching and taking notes. I am forever grateful to them for opening up and sharing their truth with me. In particular, Kristen Rohanna went above and beyond the call of duty to answer my questions and provide access to the results of her hard work on behalf of the network. It was impressive to witness the growth of the network under the guidance of you and your team.

Lastly, I want to thank my friends and family, without their encouragement I never would have had the belief to pursue this path. I hope to honor them by continuing to advocate for an educational system supports our most at risk students by working to close the achievement gap.

VITA

EDUCATION

2005 M.Ed., Education, University of California, Los Angeles, CA.

2001 B.S., Mechanical Engineering, Minor in Economics, Massachusetts Institute of Technology, Cambridge, MA.

RELEVANT PROFESSIONAL EXPERIENCE

2004 – 2019 High School Math Teacher
Lawndale High School
Lawndale, CA

2008 – 2019 Advancement Via Individual Determination Coordinator
Lawndale High School
Lawndale, CA

2015 Smarter Balanced Performance Task Rubric Workshop
UCLA Curtis Center
Las Vegas, NV

2014 Adaptive Schools Seminar
Lawndale High School
Lawndale, CA

2009 – 2010 Guiding Teacher for UCLA Teacher Education Program Novice Teacher
Lawndale High School
Lawndale, CA

2008 Union Site Representative

2007 – 2009 School Site Council
Lawndale High School
Lawndale, CA

Chapter 1. Introduction

Teaching the wide variety of students who attend American public schools is a complex undertaking. Significant time and resources have been devoted to improving this system. The past decade has seen an increased call for educators to use data to improve their practice. As such, K-12 public school teachers have become inundated with data. However, research shows schools have had great difficulty using this data to implement actual changes in the classroom that result in consistently higher student outcomes (Rohanna, 2017). Some of the documented difficulties include asking appropriate questions related to the data and connecting data use to day-to-day practice (Jimerson & Wayman, 2015). Educators have begun to look at other complex organizations for methodologies that could address these challenges (Lewis, 2015).

Improvement science originated in manufacturing several decades ago as a methodology for organizational improvement. At its core is a formal approach to *learning by doing*, and a strong emphasis is placed on the role of quantitative data in an organization's improvement process. Equally important is the idea of a network improvement community (NIC), which can be defined as a group of stakeholders all working on a common problem by sharing best practices and data in an effort to leverage numbers to speed up progress (Bryk, Gomez, Grunow, & LeMahieu, 2015).

An improvement science implementation begins by all members of the network community coming together to determine a specific problem of practice they would like to solve, such as low math scores as measured by state testing (Gomez et al., 2015). The next step is to identify and target for change key work processes where major differences in student outcomes originate. In this example, the literature suggests that when students are asked to think critically about math, test scores improve. Individual teachers will have different ideas about what constitutes critical thinking, which increases the variability in student outcomes from class to class. At this point, research and best practices are used to create a framework detailing a list of drivers that support the specific improvement, such as how students are made to think critically about math in this school district. Instruments such as short student surveys can be used to collect data on how teachers provide opportunities for students to think critically about math (Bryk et al., 2015). Ideally, teachers would be involved in the creation of these measures (Gomez et al., 2015).

Measures for Improvement

Different types of measures are used for different purposes, and "we cannot improve at scale what we cannot measure" (Bryk et al., 2015, p. 14). The measurements generally used for accountability, such as standardized test scores, are useful for highlighting where improvements need to be made. However, they lack the ability to provide teachers with actionable information about how to change their actual practice to bring about that improvement. This is a subtle distinction that most schools don't address, leading to ineffective use of school data.

For example, when two teachers compare class scores on a common assessment, it is possible to determine which class demonstrated greater mastery of the material. However, due to the many possible explanations for differences in scores across classes, it is not possible from this data to conclude that a particular action by either teacher contributed to their results. In contrast, if we think about our previous example, where students were asked to think critically about math, then both teachers could use a short survey to question their students about the quality of critical discussion throughout the unit (Jackson, Henrick, Cobb, Kochmanski, & Nieman, 2016). This data could then be compared with the student achievement data. If the classroom with fewer quality opportunities (as reported by the students) for critical thinking also demonstrated lower performance compared to the classroom with more quality opportunities, that teacher would have data supporting a change to a particular practice in this case, improving the quality of critical discussions.

If the group that reported critical discussion of lower quality received higher scores, or if, after increasing the quality of discussion scores didn't improve, then the teachers would need to revisit the research to identify other practices that may raise math achievement. Measures for improvement for these new practices could be created to gather data. In this way, improvement science is an iterative process (Langley et al., 2009). It is learning by doing but with a structured and methodical approach that always links data to specific teacher actions, thus eliminating one of the most frequent complaints educators give for why they cannot effectively use data to change their practice to improve student achievement.

Measurements for improvement are specifically designed to capture data that teachers can use to change their practice to achieve better student outcomes in accordance with a pedagogical theory (Bryk et al., 2015). These measurements differ from accountability measurements in several significant ways. For example, a measure of student achievement (such as a standardized test) is neutral when it comes to the actual work process employed by the teacher to cover the content. With this data a teacher could learn that students scored poorly on a topic, but it would not be clear what specific work process could be changed to lead to better outcomes in the future. By contrast, a measure for improvement is designed to provide evidence concerning the implementation of specific work processes that are thought to influence outcomes (Langley et al., 2009). An educator can use this data to make reasoned changes to specific classroom practices.

Research shows that effective use of data can lead to school improvement (Farrell & Marsh, 2016a; Gelderblom, Schildkamp, Pieters, & Ehren, 2016; Glover, Reddy, Kettler, Kurz, & Lekwa, 2016). Furthermore, much research exists on what it means for teachers to be data literate and the skills and supports that are needed for teachers to use data to improve their practice (Coburn & Turner, 2011; Datnow & Hubbard, 2016; Ho, 2016; Jimerson & Wayman, 2015). Particularly important are the context and culture surrounding how a given school determines what data is credible. Ho's (2016) findings outlined much of what is generally accepted at a policy level about data collection and use and provided some important guidelines about the effective use of data to improve instruction.

Ho (2016) defined different types of data usage – namely, for accountability, organizational learning, and improving instruction. It is the latter two that are under-documented in educational research. A consensus has evolved among researchers that not enough is known about the actual processes by which teachers make specific changes to their practice using data (Little, 2012). The analysis of a school site implementing improvement science principles provides an opportunity to document how teachers create and use data expressly for the purpose of organizational learning and improving instruction. This description can add to the discussion as educational organizations search for a successful methodology to support teachers as they seek to use data to improve their instructional work processes (Little, 2012).

An important component to the success of improvement science is the involvement of teachers in adapting pedagogy to suit their local context. Although sometimes teachers are given the measure they will use to gather data rather than creating them, in most cases teachers will do the actual data gathering. Teachers will be the ones to determine whether to use that data to change their practice and the ways in which their practice will change. This analysis is particularly helpful in revealing how teachers determine the credibility of the data that is generated (Little, 2012), which is a vital precursor to data usage (Coburn & Turner, 2011).

Research Questions

Improvement science has been well documented in manufacturing and health care but is relatively new to education. While there is some research on improvement science in education, there is very little concerning the actual use of measures to inform an instructional improvement effort in a K-12 setting (Yeager, Bryk, Muhich, Hausman, & Morales, 2013). Therefore, what follows is an examination of an improvement science implementation, with a focus on the improvement science principles most relevant to the collection and use of data. The overarching question, in what ways did the network's professional development employ improvement science principles in their improvement effort, will be explored through the following research questions:

1. In what ways was the network's improvement effort problem-specific and user-centered?
2. In what ways was the network's improvement effort systems focused?
3. In what ways did the network's improvement effort use disciplined inquiry?

Research Design

This dissertation is a qualitative study of an improvement science implementation focused on secondary math instruction. I gathered data by attending network meetings and taking field notes. Interviews were conducted with nine teachers from the network. An additional interview with the primary network facilitators provided further insight into the improvement science effort. Interviews provided an opportunity for participants to explicate their thought process and to help me construct meaning around the practices in which they participated (Seidman, 2013).

Primary documents, such as training materials, agendas, and Plan-Do-Study-Act (PDSA) logs created by the teachers, were also analyzed as they were the artifacts generated by the teachers as they engaged in the improvement process. Their creation and use are vital to understanding the role of measures for improvement in the process of improving teacher work practices (Hannan, Russell, Takahashi, & Park, 2015). The network also made their measure for improvement data available. Finally, an anonymous survey was conducted asking participating teachers to rank the improvement science principles in order of value to their improvement effort.

Research Site and Population

Measures for improvement created by teachers and used to inform changes to specific practices are not prevalent in K-12 public education (Bryk et al., 2015; Yeager et al., 2013). However, some partnerships between research universities and K-12 systems have recently emerged. This study followed one such partnership between a local university and a five-school network whose focus was to improve secondary math student outcomes using improvement science principles. This provided an excellent opportunity to document how teachers, assisted by an expert in improvement science, could create and use measures for improvement to change their practice.

Research Significance

As noted earlier, most data in public K-12 education has been gathered with accountability, rather than improvement in mind. As such, it is difficult to link any findings to specific work practices that teachers can change to improve student outcomes. While teachers commonly create formative assessments to evaluate student understanding, these assessments are designed to report on student achievement, not on the effectiveness of a particular teaching practice. My research aimed to address this gap in knowledge documenting how a network can create measures for improvement designed to provide data on the effectiveness of specific teaching practices, with the ultimate goal of improving instruction.

Chapter 2. Literature Review

Introduction

In the information age, data has been a boon to many sectors as they seek greater efficiencies. While educators are awash in data, researchers are just now beginning to understand how educators engage with data. What data is found to be credible is an important factor in which data, if any, gets used, and there exists a gap in the literature concerning whether teachers perceive process data as valid and relevant when making decisions about instruction. This is important as educational organizations attempt to use data to improve classroom practices.

The quality of instruction students receive has a strong impact on their academic success. The first part of this literature review examines some characteristics of professional development that have led to more effective instruction. As classroom teachers work to improve their practice, data can be a source of valuable information. Therefore, I next discuss the role data plays in the instructional decisions made by schools. Researchers have outlined several contextual factors that influence the effectiveness of data use by teachers. Next, how the term data is defined by various researchers is explored along with some typical examples of educational data: formative and summative assessment. Next, the data characteristics teachers describe as most informative for instructional change are explored, which closely match the properties of a new category of data entering the field of education: measures for improvement. Currently, there is very little research with an explicit focus on how teachers use measures for improvement to inform instructional change. Next, how the characteristics of measures for improvement mirror several characteristics of interventions that have been shown to improve instruction is discussed. The synthesis concludes with a detailed description of improvement science principles, the field from which measures for improvement originates. What constitutes a measure for improvement is clearly defined. The use of data described by improvement science principles holds great promise as a pathway by which teachers can improve their instructional practice.

Improving Instruction in Schools

Given the important connection between teacher quality and student outcomes, school districts have devoted significant resources to improving classroom instruction. A clear, scalable, and consistently successful methodology for improving instruction has been frustratingly hard to produce.

Experts in improvement outline two types of knowledge necessary for improvement: subject matter knowledge and profound knowledge (Langley et al., 2009). Subject matter knowledge in education refers to the degree to which educators understand their subject. Math teachers can't be expected to effectively teach factoring if they do not understand the nuances of factoring. Therefore, one theory for improving education is to improve the subject matter knowledge of the teachers. Increased subject matter knowledge alone however, will have a limited effect on bettering classrooms without an accompanying form of information, which Deming (1994) termed *profound knowledge*.

Profound knowledge refers to the actual mechanisms by which an individual goes about implementing changes for improvement. In other words, once our teachers fully understand factoring, they need additional information on how to best provide opportunities for students to engage in activities that will lead to their own understanding of factoring. Profound knowledge, as defined by Deming (1994), consists of the interplay between theories of systems, variation, building knowledge, and psychology. Some interventions that show promise in building both subject matter knowledge and profound knowledge in educators currently include professional learning communities (PLC), individual coaching, and lesson study. These formats share several of the following characteristics.

Sustained Engagement with Colleagues . Both of these activities share some structural commonalities that have increased their effectiveness, with one example being their duration. Professional learning communities have been shown to be more effective when teachers attend to a particular developmental goal multiple times throughout the year (Desimone, Porter, Garet, Yoon, & Birman, 2002). Doing so allows teachers to try the new pedagogy in the classroom and conduct follow up discussions to improve their practice. In contrast, single session professional development has not been found to lead to persistent changes in instruction (Desimone et al., 2002). Additionally, teachers must be actively engaged during the process rather than passively absorbing information that is presented to them. Bringing together colleagues from the same department or school was also found to be more effective than bringing individuals from different schools together (Little, Gearhart, Curry, & Kafka, 2003).

Focus on student learning and teacher practice. Besides having a similar structure, effective interventions also require teachers to engage in activities with a clear focus on student learning and teacher practice (Thessin, 2010). Whether with colleagues in a professional learning community or in a one-on-one coaching session, teacher improvement begins with a discussion about how students are learning and what teachers are doing. Effective dialogue concerning student learning requires artifacts such as test scores and student work (Russell et al., 2017). There exists extensive research documenting how teachers use student assessment data to make sense of student learning (Datnow & Hubbard, 2015), but just as important, improvement in teacher instruction requires information about what teachers are doing in the classroom, such as observational feedback by a coach, or a professional learning community sharing research-based best practices (Desimone & Pak, 2017; Kraft & Blazar, 2017).

For example, Kraft and Blazer (2017) reported that the most common coaching practice was providing teachers with direct feedback about what they could do better or differently in future lessons, something that occurred in 78% of all sessions. Although a variety of effective activities centered on teacher practice have been recorded, little to no research exists documenting the formal collection of quantitative data concerning teacher practice for the purpose of informing instructional change (i.e., measures for improvement). Current research in this area reports only evidence that is not typically measured, such as conversations about lesson planning or discussing classroom observations.

Lesson study is a professional development practice originating in Japan in which the focus is on improving instruction by observing fellow teachers (Akiba, Murata, Howard, & Wilkinson, 2019; Dudley, 2013; Lewis, Perry, Friedkin, & Roth, 2012). Lesson study is an inquiry cycle by which a particular topic is first chosen for improvement, and then a research lesson is crafted. The participating teachers observe the lesson and gather data on student thinking and learning. Afterwards, they discuss specific teacher work practices that did or did not support greater student understanding. One of the main features of lesson study is its support of teacher learning through the enactment of the research lesson (Akiba et al., 2019).

Embedded in daily practice. Between managing classroom behavior, lesson planning, and assessing student understanding, educators are hard pressed to create time for additional responsibilities. Therefore, the degree to which interventions to improve instruction are integrated into the school day influences their effectiveness. Professional development, professional learning communities, and coaching are most effective when targeting the specific content and standards teachers are addressing in their lessons (Desimone & Pak, 2017; Desimone et al., 2002).

Professional development, professional learning communities, and coaching provide educators opportunities to develop greater subject matter knowledge and profound knowledge. Creatively combining these areas of knowledge can effectively lead to organizational improvement (Langley et al., 2009). Increasingly, data is being incorporated into professional development, coaching, and other activities that teachers engage in as they attempt to improve classroom instruction. The current rationale is that data can inform practice and help teachers to improve effectively and systematically (Coburn & Turner, 2011). However, data is generated from a variety of sources and comes in many forms, which has led to mixed outcomes when teachers attempt to use it to inform instructional change.

I contend that this inconsistency occurs when educators try to use student outcome data (such as test scores) rather than data about teacher practice (such as the number of times an instructional strategy was used) to inform a change in instruction. Due to their characteristics, accountability data is not conducive to discourse around either teacher subject matter knowledge or educational profound knowledge – the components necessary to increase the capability of an organization to improve (Langley et al., 2009). In order to better understand the credibility and relevance of different types of data for improvement purposes, it is necessary to first examine how data are often used in schools, including the culture and context in which data are collected.

The Role of Data in Improving Instruction

Improvement requires that teachers combine subject matter knowledge with profound knowledge in how to deliver quality instruction, and data play an important role in this process (Bryk et al., 2015). Data-driven decision making is a complex process that is just beginning to be understood and formalized in educational settings. School districts use a variety of data, such as attendance rates, student demographics, and state test scores, to make decisions ranging from staffing to discipline policies (Coborn & Turner, 2011). With the current focus on closing the achievement gap, much data use centers around improving student learning outcomes as measured by standardized tests (Datnow & Hubbard, 2015). Every day, superintendents, curriculum specialists, and others make decisions affecting the educational process. However, this review will focus on the use of data by classroom teachers for the purpose of improving instruction.

The decision to address classroom instruction generally begins with teachers being confronted with information showing a lack of understanding by students (Glover et al., 2016). That information could come from the end-of-the-year state test results, or a warm-up problem based on yesterday's content. Faced with student results showing weak comprehension on a topic, educators must decide whether to revisit the material or move on with new content (Farrell & Marsh, 2016b). If teachers decide to reteach, they can choose to use the same approach or change their instruction. Afterwards, how does an educator determine if these instructional choices improved student comprehension? Ideally, teachers would be making informed decisions based on measures for improvement.

Much research has been conducted on teacher use of data. The result is a growing consensus on some aspects. For example, researchers agree on the importance of dedicated time for teachers to engage with and reflect on data, and on the importance of school leaders in guiding data discussions to focus on teacher practice and not on student characteristics. Research has even linked specific data use routines to improved student outcomes as measured by test scores (Coburn & Turner, 2011). However, research is sparse on the exact process by which data can effectively lead to instructional change. What is lacking is a robust understanding of the characteristics of data that best inform educator efforts to build the profound knowledge needed to improve instruction. Most research acknowledges the importance of organizational context throughout the data-driven decision making process; this is examined further below (Coburn & Turner, 2011; Ho, 2016).

Context and perceptions matter. Broadly defined, data use can be characterized as the process by which individuals make meaning of information and determine future action (Coburn & Turner, 2011). Meaning making is a highly subjective process, and the context in which it occurs is extremely impactful (Ho, 2016). Organizational structures, such as who decides what data to analyze, how much time is set aside, and even the technology that is available for data collection, play an important role when educators attempt to use data to improve their instruction (Coburn & Turner, 2011; Ho, 2016). Educators are also affected by political context, such as whether the data is going to be used in decisions about student placement or teacher evaluations.

Beliefs about student assessment data play an important role in how teachers make meaning from data and decisions concerning whether to change their instructional practice (Horn, Kane, & Wilson, 2015). In the current high-stakes testing environment, it is difficult to untangle student assessment from teacher accountability. Datnow and Hubbard (2015) reported that teachers found student assessment data to be less credible when they believed decisions about student placement and teacher evaluations would be based on the results. This resulted in more superficial conversations by teachers around data, with less time spent on pedagogy and instructional practice.

When assessments are not created by the actual classroom teacher (as is the case for state tests and some district common assessments), they may be perceived as less relevant when teachers are asked to use the results to inform changes in instruction (Farrell & Marsh, 2016a). Teachers may perceive the results of these assessments as not providing an accurate reflection of the curriculum that is being taught in their classrooms, and therefore not helpful in discussions about improving instruction.

Characteristics of data. Educators report several structural aspects of data that contribute to their perceptions of its usefulness for improving instructions. Farrell and Marsh (2016b) found in their analysis of teachers' instructional responses to data that whether the instrument used to gather the data was developed internally on site or externally (e.g., state level) was a strong predictor of whether teachers actually changed their instruction, with more locally-designed instruments being reported as more credible. Yet another important characteristic was the perceived coherence between the data-capturing instrument and the curriculum being taught in the classroom (Jimerson & Wayman, 2015). Even formatting questions as multiple choice or free response can influence an educator's perception of the validity and credibility of the resulting data to be used to inform improvement (Spillane, 2012).

The perceived credibility and reliability of data has been reported by multiple studies as crucial to effective data use. While those studies acknowledge the different *uses* for data (accountability vs. improvement), there is often no differentiation by researchers of the different *types* of data presented to educators for the purpose of informing instructional change. Student formative and summative data is by far the most common data with which teachers engage, regardless of the intended use (Datnow & Hubbard, 2015; Gelderblom et al., 2016). More recent scholarship explicitly differentiates between data that can be used for accountability and data that can be used for instructional change (Bryk et al., 2015). As argued below, data for instructional change must be data about what students and teachers *do* in the classroom, not what students *understand* as demonstrated on assessments.

Data Can Be Used for Different Purposes

To further discuss how data can aid in the development of the profound knowledge needed to improve instruction, the term "data" must be clarified. Different people have different conceptions of what constitutes data. The most common perception involves rows and columns of numbers in a spreadsheet. An example of this in education would be achievement data, such as students' scores on an end-of-the-year state assessment (Mandinach, 2012). Some consider examples of student work to be a valid form of data. Even the observation of their peers teaching can be categorized as data.

Importantly, data can be used for different purposes, such as evaluation, accountability, and improvement (Yeager et al., 2013). Quite often the same piece of data is used for multiple purposes. End-of-year state tests are frequently used by states to hold schools accountable, and in some cases, teacher compensation can be tied to these scores. At the same time, school administrators often present these same test scores to their staff to initiate conversation around improving classroom instruction (Bryk et al., 2015).

As researchers have sought to understand what made for effective data use in schools, many have purposefully adopted a broad definition of data that included any information that was used in the educational decision making process (Coburn & Turner, 2011; Farrell & Marsh, 2016b; Ho, 2016). To better locate data about teacher work processes, it is useful to discuss three dimensions in which data can be characterized: intended use, what is being observed, and the nature of the data.

The intended use of data is often for summative or formative purposes. Summative use refers to analysis of past performance, while formative use would be in support of deciding what should be done next. The use of any particular piece of data is dependent on the desire of the user and is independent of the actual nature of the data, however some data are better suited to summative or formative use, as will be discussed later (Bryk et al., 2015, Datnow & Hubbard, 2015, Langley et al., 2009). Next is what is being observed, or the object on which the data was collected. Process data is data focused on implementation, while outcome data references data collected on the impact of a process (Bryk et al., 2015; Yeager et al., 2013). Finally, the methods of analysis and summarizing can be qualitative or quantitative. Qualitative data is data which is summarized with words, while quantitative data is data that is summarized with numbers.

Data from end-of-year state assessments (e.g., a statewide grade 8 ELA test) are generally intended for summative use, observe the impact of instruction, and are quantitative in nature. This is typical of the type of data that is most frequently collected in educational settings. In contrast, recording the number of times the teacher asked a student to think critically creates data that is more suited to informing future practice (formative use), whose object is a teacher work process (student questioning), and is also quantitative. This data is rarely collected by schools.

Teacher reflections on their instruction based on student work create data that can be characterized as formative (if the intended purpose is to inform their future practice) and process oriented (if the reflections are directed towards instructional implementation rather than student outcomes). However, and most importantly for the improvement principles discussed next, it is summarized with words, and therefore qualitative.

Improvement Science Principles

Given the complex interaction between data processes, organizational context, and desired outcomes for schools engaging with data, the importance of proven interventions to promote the effective use of data has become critical. Researchers consistently request a more detailed examination of the pathways by which data use leads to actual instructional change (Datnow & Hubbard, 2015; Gelderblom et al., 2016; Little, 2012). One such methodology for guiding organizational change through the use of data comes from the field of improvement science. Improvement science has its roots in the manufacturing sector, and the Plan-Do-Study-Act (PDSA) model is its most recognizable application (Langley et al., 2009). The PDSA cycle and the principles of improvement science are a promising structure because they address many of the issues brought up concerning effective educational data use, such as work processes, local context, and outcome goals.

Improvement science is a set of principles that an organization can employ to learn by doing (Lewis, 2015). Proponents of improvement science draw heavily upon Deming's (1994) system of profound knowledge when describing the information that is required for organizations to intentionally get better (Bryk et al., 2015; Langley et al., 2009), which will be highlighted below. Improvement science recognizes a fundamental difference between data designed for accountability (such as test scores) and data designed for improvement (such as information on whether students or the teacher talked the most during an activity). Therefore, improvement science offers a potential pathway by which data use in schools could lead to reasoned instructional change that results in better student outcomes (Yeager et al., 2013). Bryk et al. (2015) described improvement science in an educational context with the following six principles: 1) problem-specific and user-centered work, 2) a focus on individual variation, 3) recognizing systems produce outcomes, 4) scaling up requires measures for improvement, 5) disciplined inquiry drives learning, and 6) networks accelerate learning. These six principles represent practical solutions to the problems identified by researchers and educators as inhibitors to effective data use in schools (Bryk et al., 2015).

Problem-specific and user-centered. The first principle is to make the purpose of the improvement effort problem-specific and user-centered (Bryk et al., 2015). This means that sites should address locally-identified concerns generated by the actual practitioners – teachers. This orientation addresses the repeated concern by teachers that much of their conversations around data have been hard to connect to their actual practice. Deming (1994) wrote about the role human psychology plays when an organization tries to enact change. By gathering information about teacher practice in contrast to information about student understanding, problems around teacher motivation and attitudes to change are taken into account. Student assessment data often result in teachers making a fundamental attribution error, whereby they explain student performance in terms of the environment rather than holding themselves accountable (Langley et al., 2009). Additionally, when teachers are part of the decision-making process around classroom problem needing to be addressed, there will be more buy-in throughout the entire data use process.

Focus on variation in performance. The second principle is to concentrate on the variation in performance (Bryk et al., 2015), which matches the understanding variation component of Deming's (1994) profound knowledge. Many programs supported by research have been implemented at schools only to fail because of local complexities (Rohanna, 2017). Rather than determine if a program is working using schoolwide outcomes, improvement science focuses on how to determine if a particular program is working in individual classrooms under particular contextual conditions. This requires a different kind of data to answer (Langley et al., 2009). Student achievement data identify a deficient program, but such data are limited in the contextual information they provide around that deficiency (Datnow & Hubbard, 2015; Gelderblom et al., 2016). Recognizing that there are differences in the type of data that an organization can collect is fundamental to improvement science. The ability of data to inform as teachers attempt to change their instructional practice is highly dependent on the attributes of the data. This is discussed further in the fourth principle.

See the system that produces the current outcomes. As discussed above, the research literature consistently highlights the importance of political and organizational cultures to make data-driven instructional changes (Bryk et al., 2015; Coburn & Turner, 2011; Datnow & Hubbard, 2015; Gelderblom et al., 2016). The third principle of improvement science is to recognize the system that is producing the current outcomes. To begin an improvement effort, an organization first must develop a theory of change grounded in research (Langley et al., 2009). The theory of change stipulates that existing systems and work processes are the driving forces behind the current outcome, and improvement requires testing new or modified work processes in order to influence these drivers and eventually improve the desired outcome. A driver diagram (see Figure 2 for an example) outlining the primary drivers of a given outcome is vital tool in the improvement science process, as it guides the potential change ideas for teacher practice as well as informs what information should be gathered to evaluate the efficacy of those changes (Bryk et al,, 2015). Deming (1994) also outlined an appreciation for the system in which any attempted change was to occur as a necessary contextual starting point for learning.

We cannot improve at scale what we cannot measure. The fourth principle is arguably the most important for effective use of data to change instruction: an organization "cannot improve what it cannot measure" (Bryk et al., 2015). Deming (1994) described this more generally as building knowledge. In the context of improvement, the theory of change is a prediction: if a particular change in instruction is made, improvement in student outcomes will result. In the current climate of high-stakes testing, schools have increasingly turned to data to guide their decisions. Student achievement data, the most common data collected by schools, are essential for identifying areas of concern (Datnow & Hubbard, 2015); this, in itself, is a necessary step as schools decide where to focus their change efforts. However, once a need has been identified, many schools attempt to repurpose this same data to make decisions about instructional changes.

Student assessment data is limited in their ability to inform teachers about specific changes they can make to their practice to address student conceptual misunderstandings. Thus, data are limited in their ability to add to educators' profound knowledge supporting rational changes in instruction. It is at this point that teachers often report that "data" are perceived as either not credible or not relevant to their practice. The result is superficial change to instruction at best, mostly consisting of reteaching the material in the same manner (Gelderblom et al., 2016). To paraphrase Bryk et al. (2015), we can't improve teacher practice if we do not measure teacher practice!

Improvement science scholars describe improvement as the interplay between theory, measurement, and standard work processes (Bryk et al., 2015; Langley et al., 2009). A working theory of change describes how standard work processes influence desired outcomes (through existing systems). Change ideas are evaluated through measures for improvement. Standard work processes are the core practices of an organization. For math teachers, standard work practices would include the pedagogue used to teach content, grading processes, and the process by which formative checks for understanding occur. Improving sub-standard work processes is at the core of improvement science. Changes to standard work processes are based on a research-supported theory of change and codified into a driver diagram. Measures for improvement provide quantified data from which an organization can numerically, and objectively, evaluate the efficacy of changes to standard work processes to improve the desired outcome (Bryk et al., 2015; Langley et al., 2009).

Improvement science scholars recognize the importance of observation (Bryk et al., 2015; Langley et al., 2009). However, the subjective nature of observation renders it insufficient to support the continuous learning outlined in improvement science principles unless it is captured in a measurable way. Langley et al. (2009) defined data as "documented observations, including those that result from a measurement process" (p. 28). They go on to list five types of data conducive to improvement efforts: 1) continuous measurements; 2) counts or classifications of observations; 3) what people think, how they feel about something; 4) ratings; and 5) rankings. Importantly, four of the five types are clearly referring to quantitative data, and while the fifth is less clear, the example suggests Langley et al. have something quantitative in mind for this as well. The example they provided for 3) what people think, how they feel about something (Responses to the question, "Is this format for the newsletter easier to read than the current one?") is not open ended. The Y/N response options result in a quantifiable count. Improvement science literature is consistent in its espousal of quantified data to inform continuous learning (Bryk et al., 2015; Langley et al., 2009; Yeager et al., 2013).

Measures for improvement differ from measures for accountability in important ways. Measure for improvement are formative in that they focus on the work processes being done by teachers and students (Bryk et al., 2015). With this information, teachers can attend to specific aspects of their practice for change. Measures for accountability are more general, in order to facilitate comparison across teachers. An example from Bryk et al. (2015) involving teacher feedback to students is instructive. The standard teacher work process of guided student reading contains the "after reading" micro work process. A measure for improvement rubric recording if the "Teacher engages students in conversation about the meaning of the text" might have the descriptor that students are engaged in "some discussion, but talk is often unfocused or off topic." The data resulting from this rubric could be used formatively, as in a teacher receiving this score could improve her teaching work process by reminding students to stay on topic when she gives them feedback. The measure for accountability rubric recording on "Feedback Quality: Accurate, Substantive, Constructive and Specific" has the descriptor: "Feedback is inconsistent in quality; some elements of high quality are present; others are not." This rating easily allows comparisons amongst teachers but does not lend itself as well to a formative usage. Measures for improvement work best when they give teachers information about *how* to increase the quality of their feedback.

It is vital to the improvement science model of continuous learning that measures for improvement are quantitative rather than qualitative. While qualitative observation can inform the improvement process, in order to measure changes in work processes over time in connection with changes in outcomes, in addition to identifying variation in the work process quality between teachers, quantified data are required. This information is often captured on run charts, which facilitate the objective identification of patterns and trends. Run charts typically plot the operationalized dependent variable of interest as a function of the independent variable time (Bryk et al., 2015; Langley et al., 2009). Continuous learning requires frequent collection of measures for improvement; therefore, they must be practical (Yeager et al., 2013). This means the measure must be embedded in the regular work of teaching as seamlessly as possible. Bryk et al. (2015) defined four distinct measures for improvement: outcome measures, primary driver measures, process measures, and balancing measures.

Outcome measures “operationalize the aim statement in the driver diagram” (Bryk et al., 2015, p. 103). With this information, an organization can evaluate whether measurable progress is being made on the problem being addressed. An example of an educational outcome measure would be the percentage of students passing a class with a “C” or better. A shared outcome measure is necessary for an improvement science network improvement community to coordinate its efforts (Langley et al., 2009).

A theory of change posits that primary drivers are the major mechanism leading to the current outcomes. In order to facilitate the continuous learning needed for improvement, an organization needs to gather data on the ability of changes in standard work processes to eventually influence the chosen outcome. This is done through *primary driver measures*. According to Bryk et al. (2015), a primary driver measure should: 1) predict the ultimate outcomes of interest, 2) be sensitive to changes in standard work processes, and 3) provide guidance as to where subsequent improvement efforts might focus. In order to facilitate continuous learning, primary driver measurement data must be gathered and reviewed frequently.

While outcome measures report on the goal (end result), and primary driver measures report on the major mechanisms driving that outcome (intermediate feedback), process measures report on whether specific micro work processes are performing as planned (immediate feedback) (Bryk et al., 2015; Langley et al., 2009). The most important characteristic of process data is to provide guidance on where subsequent improvement efforts might focus. At small scale, qualitative observation might suffice as a process measurement. An example would be several teachers making a change to a standard work process and deciding, based on observation and experience, whether the change resulted in an improvement. However, the literature on improvement science is clear that in order to drive continuous learning, to identify variation in individual performance, and to improve at scale, process measure data must be quantified (Bryk et al., 2015; Langley et al., 2009; Yeager et al., 2013). Process measurement most frequently takes place in the context of PDSA cycles (discussed more in the next principle) and is frequently plotted on a run chart (Bryk et al., 2015; Langley et al., 2009).

Finally, *balancing measures* refer to data collected about other components of the system to ensure that any improvements in the identified outcome/primary driver/process measures do not lead to a decline elsewhere in the system (Bryk et al., 2015; Langley et al., 2009). For example, if a network's aim was to increase student achievement as measured by grades, a balancing measure might be student attendance rates. This measure would enable the network to determine if their efforts to improve grades was having an unintended, detrimental effect on attendance. Currently, balancing measures are infrequently considered in educational improvement efforts; therefore, they are beyond the scope of this report.

In applying improvement science principles to education, it is necessary to collect more than student achievement outcome data. Data on the *primary drivers and work processes* that have been identified for change must be analyzed to determine whether a particular instructional change contributes to improved student outcomes (Bryk et al., 2015 and Yeager et al., 2013). This type of data is not often collected at school sites, creating a logistical challenge for schools trying to effectively use data to change instructional practices. Again, there exists a gap in the research concerning the creation and use of these measures by educators.

Use disciplined inquiry to drive improvement. The fifth principle directs practitioners to use disciplined inquiry to drive change (Bryk et al., 2015). This process is described as using small, quick experimentations in instruction, while regularly using measures for improvement to inform change efforts. This principle relies heavily on the measures for improvement described above. PDSA cycles are the typical format for conducting improvement science disciplined inquiry (Bryk et al., 2015; Langley et al., 2009). As discussed, observational data can suffice for small scale efforts; however, quantified process measure data are essential to identifying effective change ideas, as well as for attending to individual variation among teachers. Process measure data (typically organized in run charts) must be regularly used during the PDSA cycle to guide change efforts. This is the essence of improvement science – "learning by doing."

Comparing the use of process data with student achievement data is instructional. Student achievement data are often used for high stakes accountability decisions such as student placement, teacher evaluation, and state sanctions. This creates strong incentives for discussions about data to center on the test scores of low performing students, rather than the substantive conversations around teacher practices that help all students, which have been observed at high-performing schools (Datnow & Hubbard, 2015). If student achievement data were supplemented with process measure data, informing which practices work in the classroom, this profound knowledge could be leveraged alongside subject matter knowledge to consistently move more teachers towards improved instruction. The frequent nature of process measures, and their focus on instructional work practices rather than on student achievement, encourage teachers to investigate which variations in their practice have better addressed student misconceptions.

Accelerate learning through networked communities. The sixth principle of improvement science is that learning is accelerated when organizations share their progress as they work towards a common goal (Bryk et al., 2015). For example, this could occur when teachers within a department take time to interpret data gathered from utilizing common measures as they modify their practice. Since this process data reports on what teachers are actually doing in the classroom, there is less potential for dismissing the results as due to a strong or weak group of students, which has been reported when teachers examine outcome data (Datnow & Hubbard, 2015). The community of practitioners build meaning around success of different changes to instruction together. This shared knowledge about what actual routines teachers can employ to change their practice to improve student outcomes reduces the variability that is so often seen when a school attempts to use data to determine the success of an intervention addressing an identified weakness (Bryk et al., 2015). The speed and depth of actionable knowledge that can be constructed increases as the size of the network improvement community (NIC) increases.

An Educational Model of Improvement Science

In an educational improvement science model, such as the one described by Bryk et al. (2015), we would expect the network to first identify a specific problem to be solved that is within the scope of the organization. The problem should consist of relevant work processes, and ideally should be centered on the experience of the user. In a school, the network would examine student experiences and address the largest impediments to success.

In the idealized model, we would expect to see a network identify how work is actually carried out in classrooms and how larger institutional forces shape this. Teachers would be engaged in the process of conceptualizing the problem, examining possible change ideas, and collecting information about improvement as these ideas are tried out. This guides the creation of change ideas to be tested (Bryk et al., 2015).

Measurement is an integral part of a high-fidelity implementation of improvement science. In the model, we expect to see specific, measurable aims the network seeks to accomplish. Regular data collection on the identified drivers disciplines the network and keeps it accountable. Frequent, quantitative data about specific work processes targeted for change would inform any network decisions on the efficacy of their efforts. Ideally, intermediate outcomes would be identified and measured, as student academic outcomes may not change for quite some time. Finally, other measures would track unintended consequences of any changes (Bryk et al., 2015).

Without measures for improvement, it is difficult to objectively and quantifiably connect the work of the network to its drivers or its aims. In the model improvement science implementation, measures for improvement are used to regularly collect data on a common driver of focus, and this data inform decisions about the efficacy of the change idea. Through this process, a successful work process is identified and expanded to more classrooms. As more data are gathered, the network learns how to make the new instructional processes work for a variety of students, when employed by a variety of teachers in a variety of contexts (Bryk et al., 2015).

The frontline work of an improvement network is to generate learnings by rapidly testing change ideas and gathering evidence in a minimally intrusive manner. This is typically done through Plan-Do-Study-Act cycles in conjunction with process measures. The data collected through process measurement identify new, effective work processes that are then iterated upon and scaled up to more classrooms. Data continue to be regularly gathered, building institutional knowledge on how a common work process can succeed in multiple contexts (Bryk et al., 2015).

Network learning is an important concept in the improvement science model. We would expect to see organizational resources, formal agreements, and normative understandings shared among the participating teachers. This allows the network to learn faster than an individual teacher. This is exemplified by the network putting a priority on solving a problem together (Bryk et al., 2015).

Outcome and Process Data are Complementary

While outcome data (e.g., student achievement data) are vital for identifying deficiencies or gains in student understanding, primary driver and process measures are necessary for educators to connect change in practice to improvements. Summative achievement data can be effective for highlighting what students do or do not understand. Summative achievement data can provide valuable information about what students have learned when compared over time. This data help district and school leaders to notice where an intervention needs to take place in order to improve student outcomes. This type of data has been particularly helpful in highlighting the achievement gap between White and Asian American students and their Black, Latino, and Asian Pacific Islander counterparts.

However, the effectiveness of this data when repurposed by teachers to inform changes in instructional practice has been limited because they come at the end of a unit or school year or the following fall, much too late to inform constructive changes in instruction (Jennings, 2012). Formative assessment data are also good for highlighting what students understand. Teachers can’t improve student learning if they don’t measure student learning; therefore, frequent formative assessments are often used to measure student learning. Unfortunately, the kind of data typically collected under the guise of "formative assessment" turn out to be not so helpful in informing teacher practice, making it difficult for teachers to use the data to inform instructional change in a data-driven way.

The use of formative assessment is a critical part of quality instruction. Through formative assessment, teachers attempt to ascertain what students do or do not comprehend and what conceptual or procedural misunderstandings led to the results observed on the assessment (Oláh, Lawrence, & Riggan, 2010). Next, teachers must confirm their appraisal of student understanding and choose an appropriate instructional strategy to address misunderstandings. It is in this last part, the attempt to link their understanding about where students are struggling to the best way to scaffold their learning, that teachers consistently report the greatest challenge (Datnow & Hubbard, 2015). The difficulty, again, is the mechanism by which teachers use student achievement data to select an appropriate instructional response, even when data are as immediate and relevant as a daily exit ticket.

If teachers want to improve teacher practice, schools must measure and assess teacher practice. Primary driver and process measures are much more informative to instructional improvement efforts than assessments of student understanding. Primary driver and process measures provide timely data clearly connected to classroom work processes that allow teachers to make informed decisions about instructional change. This is precisely the type of information necessary to build Deming's (1994) profound knowledge, information needed to implement effective change in instruction. Similarly, these attributes match those perceived by teachers to make data more valid according to research (Coburn & Turner, 2011; Datnow & Hubbard, 2015; Farrell & Marsh, 2016a; Ho, 2016).

Earlier, I outlined some of the characteristics of data that teachers describe as important, particularly the perceived relevance of the data to their practice and the standards being addressed. Interim assessment data, such as quarterly benchmarks, have become the most common data used by educators wishing to change their instructional practice to improve academic outcomes. An examination of this data by Datnow and Hubbard (2015) highlighted the deficiencies when student assessment data in general are used to inform decisions concerning instructional change.

When benchmark assessment data are not closely tied to the content, standards, and skills covered in the classroom, they lose their value regarding decisions about instruction (Jennings, 2012). Student achievement data use was found to lead to fewer instructional changes in situations where teachers perceived a misalignment between the assessments and their classroom curriculum. Even with proper alignment however, there is not always a clear connection between benchmark scores and specific actions teachers can take to address the weaknesses uncovered by the data. Assessments consisting of multiple-choice questions provide only limited insights concerning student cognition that teachers can use to change their practice in a meaningful way. Research shows that data from multiple choice benchmark assessments led to less instructional change and was found less credible by teachers compared to free response question items (Datnow & Hubbard, 2015).

The interpretation and use of data take place within the political and organizational context of a given school. Therefore, in sites where the pressure of accountability is high (for example, schools that are labeled as “failing” by the state), assessment data conversations centered around student motivation and the identification of “bubble” students who could most easily be moved into the next content mastery bucket, rather than substantive conversations about how to improve instruction for all students (Colyvas, 2012).

Teachers also reported a tendency to view student assessment data as a one-way communication with parents and the community and that it was “dissociated from their teacher duties” (Datnow & Hubbard, 2015, p. 22). This effect became more pronounced from the elementary to the high school level. Lack of buy-in for student achievement data was associated with teachers’ beliefs in an “external locus of control” (Datnow & Hubbard, 2015, p. 21), such as attributing data to simply a good or bad group of students.

Student achievement data are created, by definition, to capture information about student understanding (Black & Wiliam, 2009). Teachers use this information first and foremost to evaluate student comprehension. There may be a subsequent intent to use the same data to inform instruction, but the content of the data remains information about what students do or do not know. The same logic would indicate that if teachers want to evaluate their practice, a necessary step in improving instruction, information needs to be gathered about teacher practice (Bryk et al., 2015). Teachers require formative data on their own practice in order to evaluate the effectiveness of their instructional choices. This is exactly the data provided by primary driver and process measures.

Unfortunately, most teachers have been making instructional decisions without the benefit of the profound knowledge primary driver and process measures could provide. When interventions that are designed to support instructional improvement, such as lesson study, coaching, and professional learning communities, are rated as effective by teachers, the activities they engage in are described as being relevant to teacher work processes as well as cohesive with curriculum and content (Desimone et al., 2002). Once student achievement data has been used to identify student misconceptions, I contend that primary driver and process measures are the closest analogue to what is reported by teachers as effective about interventions to support teacher decisions about instruction.

The improvement process can be framed as the use of measures for improvement to build profound knowledge on how best to implement change. However, there is a paucity of research examining the explicit role of primary driver and process measures in instructional improvement efforts. Many researchers have expressly called for a more detailed examination of the pathways by which teachers make meaning of and use data to make instructional change (Little, 2012). It is hoped that by documenting teachers who have received professional development in the use of primary driver and process measures, their successes in making effective, informed choices regarding instruction can be reported.

Because of my interest in the use of data to inform instructional decisions, the focus of this study will be on the improvement science principles most relevant to data use that are observable in the early stages of the chosen network. These are the principles described by Bryk et al. (2015) of problem-specific and user-centered work, seeing the system that produces the current outcomes, and using disciplined inquiry to drive improvement.

Chapter 3. Research Design and Method

Introduction

This chapter presents the research methods and analyses used to examine the improvement science principles as practiced by a network improvement community. First, this chapter provides an overview of the study procedures, which is followed by information about the network's setting, participant selection and recruitment, and data collection. Lastly, the analytic procedures are described. The specific questions for this study were:

1. In what ways was the network's improvement effort problem-specific and user-centered?
2. In what ways was the network's improvement effort systems focused?
3. In what ways did the network's improvement effort use disciplined inquiry?

Research Design and Rationale

This dissertation is a qualitative study of how public secondary school math teachers implemented improvement science principles to improve classroom instruction as part of a two-year professional development network facilitated by a major research university. According to Merriam and Tisdell (2015), "Qualitative researchers are interested in understanding the meaning people have constructed" (p. 15). This description is appropriate as I sought to understand how teachers experienced the network's improvement science implementation, as well as the factors that supported or hindered the network's efforts. Interview responses, which are subjective, were the primary evidence used to understand the network's decisions about what data were chosen, how those data were interpreted, and what conclusions could be drawn from the data. This points to a constructivist approach to the study of this network's improvement science data use. Creswell and Creswell (2018) described social constructivism as rooted in the idea that "individuals develop subjective meanings of their experiences" (p. 8), and they recommended that researchers "rely as much as possible on the participants' views of the situation being studied" (p. 8). As such, qualitative research methods enabled a deep exploration of this purposefully chosen setting, including the necessary context for individual teachers to report on their engagement with improvement science principles.

Though the research was guided by theory, I was not seeking to test a predictive model. Quantitative instruments could not easily capture the lived experiences of teachers engaging in an improvement effort modeled after improvement science principles (Creswell & Creswell, 2018). Qualitative research methodology allowed for an emergent understanding of the implementation process based on the reported experiences of the participating teachers. Interviews permitted collection of the rich data needed to describe the process by which teachers involved in an improvement science professional development use data about teacher practices to make decisions about their efforts (Seidman, 2013).

Strategies of Inquiry

Site selection and participants. The network of five secondary schools chosen for this report was created in 2017 and initially was part of a larger, existing 16 high school network of schools from the same large, metropolitan, urban school district. This larger network had chosen student mastery of algebra, viewing this course as foundational to other math courses, as their problem of focus. During the larger network meetings in September, October, and November 2017, all schools worked to define a common problem of practice, establish an aim statement, and develop a driver diagram. During this time, facilitators for the smaller five-school network also worked on site with their schools once or twice a month (Rohanna, 2018).

In December 2017, the primary facilitators for the five-school network (referred to as "the network" from here on out), decided to split off from the larger group and continue their improvement efforts separately. The primary facilitators consisted of Sarah, an educational Ph.D. candidate with the university, and Tom, a university math specialist with a doctorate in educational leadership. Sarah had considerable previous experience in program evaluation, while Tom's expertise was in math pedagogy.

The network included 23 teachers in 2017 across five secondary schools. By October 2018 it had grown to 29 teachers. The five schools represented an ethnically diverse group of students (Table 1). Permission from the network facilitators to observe their network in January 2018 was obtained. The network met in January, March, April, and May 2018, and conducted professional development in three areas: 1) strengthening teacher content knowledge through pedagogical demonstrations and activities, 2) instructing teachers in improvement science principles, and 3) supporting teachers in their improvement science implementation.

Table 1. *Network School Characteristics, 2017-2018*

School	Number of Network Teachers	Total Students	Demographics (%)					
			Low SES	African-American	Asian	Hispanic/Latinx	White	Other
1	7	1,004	91.8	1.7	9.7	80.8	1.6	6.2
2	6	613	56.3	20.6	5.2	40.6	28.1	5.5
3	6	394	7.9	51.5	0.0	46.7	0.3	1.5
4	5	660	75.3	13.0	3.5	69.2	10.9	3.4
5	5	1,564	75.6	24.6	7.0	52.9	11.6	3.9
Combined	29	4,235	70.3	19.3	6.2	59.7	10.5	4.4

Note: The source is the California Department of Education. The number of teachers represents those participating in the network as of October 2018.

For each of the 11 network meetings, teachers convened on the university campus for a full day of professional development (Table 2), split between math content and improvement science activities. Substitute coverage was provided for the participants. Additionally, the facilitators visited the school sites, participating in math department meetings. The frequency of these visits varied by site, due to different levels of access granted by each site administration. The teachers completed one to six PDSA logs from January 2018 to May 2018; each log represented a two-week PDSA cycle.

Table 2. Network Schedule

Year One 2018	Year Two 2018-2019
Jan 31	Sep 13
Mar 1	Oct 18
Apr 12	Nov 13
May 17	Mar 21
July 23/25	May 16

In October 2018 the network expanded to 29 participating teachers. The network continued to meet at the local research university campus for guided professional development centered on improving instruction through improvement science principles. The participants were encouraged to incorporate the improvement science tools they developed to participate in further discussion on instruction during their respective math department meetings. Every two weeks the participants were asked to complete a PDSA cycle around a change idea. Therefore, the network represented a fertile context for understanding the implementation of improvement science principles.

This network was purposely chosen for study because of the potentially rich data use environment. The network schools conducted regular PDSA cycles, which were expected to involve the creation and use of primary driver and process measures. The chosen focus of their improvement effort was the math success of their students, as secondary math grades are a well-documented barrier to college admissions. The selection of this focus ensured a robust opportunity for teachers to repeatedly engage with multiple forms of data with the express purpose of informing instructional change. Specifically, teachers in the network were encouraged to include data on their practice alongside any existing student academic outcome data, such as formative and summative assessments when making decisions about instruction.

This context is far from the typical K-12 instructional improvement effort. However, the focused attention on instructional change and improvement science methods in this setting made this an ideal opportunity to document how teachers experience an improvement science implementation and how they use data in a situation where they are strongly encouraged to engage with it and are supported in doing so.

By attending and observing the monthly university-led professional development meetings from January 2018 to October 2018, I developed a level of familiarity with the participating teachers who became the interview subjects for this qualitative research. Specific teachers were recruited for interviews in October based on discussions with the facilitators and my own observations of the participants during their professional development. There was a purposeful selection of teachers with substantial variation in years of teaching experience. In so doing, I allowed for multiple perspectives of the improvement science process. These were important distinctions to account for as I built a deep and rich picture of teachers' experiences with improvement science measures for improvement and the effectiveness of these measures to inform instructional change.

I interviewed nine teachers individually and interviewed the two university facilitators together. This number was enough to provide for some variety of experience and predilection for data use, while still limited enough to allow for the completion of an in-depth profile of each participant's thought process. The teachers were selected from four of the five schools participating in the network.

Data collection methods. I first analyzed PDSA logs created by the participating teachers, Interviews were conducted with each targeted teacher, questioning teacher beliefs on the characteristics of data that they perceive to be valid and relevant for informing instructional change, and what factors influenced their efforts. Completed PDSA logs (a blank is included in Appendix A) were used during these interviews, providing an opportunity for each teacher to reference specific aspects of their instructional change effort to their reflection on data use. Additionally, a short survey was presented during a network meeting (Appendix B), where they were asked to anonymously rank the six improvement science principles in which the network had engaged them, according to their value in the network's improvement effort. Finally, I was granted access by the facilitators to the measurement data they collected (Appendix C).

Document analysis was conducted on the 99 PDSA logs that the 29 teachers completed for the network professional development sessions. Individual teachers completed between one and six PDSA logs. Access to the PDSA logs that were submitted from January 2018 through May 2019 was given. These documents served as an important additional source of evidence for triangulation of the emergent themes around improvement science principles and implementation. The PDSA log explicitly asked teachers to identify an idea for a change in instruction and then to answer the following questions:

Predictions: What improvement do we think will happen?	**Questions:** What do we want to learn from this cycle?	**Data:** What information will we collect to answer our questions and test our prediction?	**Results and Next Steps:** Did you meet your prediction? What were the results? What did we learn? What will we do next? *(completed after implementation)*

Seven teacher interviews were conducted in December 2018. The facilitators were interviewed in March 2019. Later that same month, four of the seven initial teachers were interviewed again, in addition to two new teacher interviews. The interview protocols

(Appendix D) revolved around the PDSA logs that each teacher completed as part of the network professional development. The PDSA logs specifically asked teachers to document an instructional change idea, identify what data they would collect, and reflect on the results of their instructional change. The interview questions focused on the validity and relevance of process data in their instructional decision-making. The participant teachers were provided the opportunity to more fully explain their reflections on their PDSA cycle, which provided evidence to address the research questions. Participants were also provided the opportunity to reference changes in their instruction, if any, that were made and to reflect on the role of process data on their instructional choices. Teachers were given the opportunity to discuss the credibility and relevance of any data that were collected. The responses from this protocol were used to answer the research questions. The interviews were conducted in person, at a location of the participants' choosing to facilitate comfort.

Field notes were created from my observations of the network's professional development (Appendix E), as teachers were asked to capture data on their practice (through PDSA logs) and discuss any data driven decisions around instruction. Observations during these meetings were focused on how the participating teachers described any changes to their instruction in response to their improvement science efforts with the network. This evidence was used to further answer the research questions.

As this research was part of a larger study being conducted on networked improvement communities, I was fortunate to have access to the network facilitators' primary driver measurement data. These data provided further information on the process by which educators evaluate a communal area of focus using the principles of improvement science, of which data collection and analysis feature prominently. Finally, surveying the participating teachers generated additional evidence about how teachers experienced the improvement science effort in which they were engaged.

Data analysis methods. This study focused on the three data centric improvement science principles that were observable from the chosen network. Therefore, interview transcripts and documents were first reviewed for their relevance to the following principles: problem-specific and user-centered, systems thinking, and disciplined inquiry. Evidence for each principle was further separated into that which reported on teacher perceptions and that which reported on the network applying the principles in practice. As the bulk of the evidence concerned the disciplined inquiry principle, further subcategories were created such as measures for improvement, PDSA cycles, unquantified data, and supports or hindrances to implementation.

The interview data gathered was analyzed based on transcribed audio recordings of each teacher interviewed. The transcriptions were studied for patterns in how the network conducted its improvement efforts, with particular attention to aspects of implementation that pertained to the three improvement science principles referenced in the research questions. Additionally, transcriptions were examined for patterns around teachers' perceptions of the improvement science professional development, with particular attention to the data used by the network. Direct quotes illustrating the subcategories described above were captured through Quirkos, a CAQDAS software package for the qualitative analysis of text data, commonly used in social science.

Over the course of several months, teachers were prompted by the facilitators to engage in various practices based on the improvement science principles. The field notes collected from my observations of these meetings were analyzed for evidence of the network's implementation of improvement science principles and teacher perceptions of these efforts. Any statements made by the participating teachers concerning their experience with the improvement science effort were coded along the same themes as interview responses.

The PDSA logs were examined for a connection between the driver diagram and change idea. Teacher responses to the PDSA log question, which asked teachers to identify a data collection method, were analyzed for their suitability as measures for improvement. Once the PDSA logs were sorted by the type of data that was collected, teacher reflections were coded and compared to the themes generated from teacher interviews. Evidence related to the selected improvement science principles provided supporting documentation concerning the practical implementation of these principles by the network. The information gathered through field notes and document analysis provided information helpful in answering each research question.

Interview transcripts and documents provided different views into the network's improvement science implementation. Therefore, I was able to triangulate these sources to strengthen corroborated results or further analyze disconfirming results. This led to the emergence of this study's 10 key findings.

Ethical Issues

All participants in the study were advised on the purpose and nature of the research. A problem statement was presented, which clearly advised them on the process by which the study would proceed. All participants received consent forms for interviews and were reminded of the non-evaluative nature of all data collection conducted in the study. This report consists of pseudonyms for participants and school sites to protect confidentiality. All files containing the real names were password protected and then deleted once all the interviews had been transcribed.

Since I do not work directly or indirectly with any of the schools in the network, there are no conflicts of interest, and I have no authority over any of the participant teachers. However, given my position as a doctoral student at the university conducting the professional development, the potential existed for that association to influence participant responses. I continued to remind the participants that the research I conducted would avoid identifying individual sites and participants and would not be used for evaluation or accountability. The purpose of my research was to understand their improvement science effort in an attempt to support measures that improve classroom instruction. A summary of the findings was shared with the participants and participating schools to support their efforts to improve classroom instruction through their improvement science professional development. All interviews were conducted in a space of the teachers' choosing (since the professional development was hosted on a university campus) in order to increase comfort and minimize the influence of the university setting.

Ensuring Credibility

As an experienced K-12 math educator I came into this research with strong opinions about the effectiveness of different types of data to inform and lead to instructional change. I attempted to limit any bias in my data collection and reporting by following a semi-structured protocol, which allowed the participants to share their own experiences and reactions to the data resulting from their improvement science effort. The use of standardized protocols and coding procedures allowed themes to emerge from the participants' responses as free from researcher bias as possible. During each interview, I summarized important responses back to the participant. These member checks were conducted with participants throughout the interview process so that they could comment on and affirm the accuracy of any themes. Descriptive validity was ensured by the use of direct quotes from the rich data that were gathered.

Because the participant teachers were actively engaging in professional development around improvement science principles, there existed the potential for response bias. Participants may have been more inclined to report these principles as useful since I was a part of the university community that conducted their training. By embedding myself in their professional development months in advance of my data collection, I tried to establish a level of trust that supported greater candor in participant responses. Still, participants may have been biased towards referencing improvement science positively considering the significant time they were being asked to spend on it. Therefore, interviews included direct reference to the complete PDSA logs where participants were asked to comment on specific decisions and actions undertaken.

Finally, interview questions described potential data items rather than the terms *measures for improvement* or *process measures,* which may have been considered leading. As the purpose of this study was to examine the usefulness of different types of data to inform instructional change rather than to identify student misunderstandings, participants were guided to direct their responses to this explicit aspect of data use.

As addressed above, this study only attempted to generalize as far as other sites that have received explicit improvement science professional development. Further extrapolation was reserved as a suggestion for additional exploration.

Limitations

It is important to address my positionality to the subject matter. I have taught high school mathematics for 15 years, serving a similar population of students to the schools I observed. As such, I cannot help but to have formed opinions on what information could best serve teachers looking to improve their instruction. Additionally, in the course of observing this professional development, I assisted the facilitators logistically at times, such as leading a group of teachers reflecting on a PDSA cycle. At no point did I address the group or share with individual teachers my personal opinion of the improvement science effort in which they were engaged. I attempted to limit any bias in my interviews by following a scripted interview protocol that was screened by my committee co-chairs. I conducted member checks, whereby I asked participants to confirm that I had indeed understood them correctly. I attempted to confirm any findings through the triangulation of multiple data sources, including interviews, PDSA logs, observations of network meetings, and survey data.

This study attempted to bring attention to how teachers can use improvement science data principles to inform decisions concerning instruction. Asking teachers directly about their beliefs was the most straightforward way to begin this exploration, but there existed the potential for self-reporting to incorrectly reflect actual actions. With more time and resources, teachers could be observed teaching over an extended period as they participate in improvement science PDSA cycles, and their claims could be verified through direct observation. Additionally, the ultimate goal was to change instruction in order to improve student outcomes. The data collected through interviews and document analysis of PDSA logs, while pertinent to teacher use of data to inform instruction, are difficult to connect to student outcomes. Therefore, the effectiveness of any changes in instruction can be inferred only from the existing literature on the influence of strong mathematical pedagogy on student achievement. Future studies could attempt to measure the impact of improvement science principles on student outcomes.

Summary

This qualitative study aimed to understand how teachers use improvement science data principles to make changes in their instruction. Through repeated interviews with several teachers engaged in improvement science PDSA cycles, supplemented with observations and analysis of PDSA logs, this study contributes to the existing literature by illuminating an improvement science implementation in practice. In particular, an extended look at how one network applied the data centric principles to support an instructional improvement effort is provided. This study may serve as a reference for future sites looking to include improvement science principles and measures for improvement into their instructional decision-making process.

Chapter 4. Findings

As this study's research question was rooted in the implementation of the data centric improvement science principles by a school network, this chapter begins with a brief description of the network that served as the focus of this study. The decisions and actions undertaken by the network provide necessary context to understanding the findings that follow.

Improvement Initiative Overview and Timeline

In January 2018, the facilitators, Sarah and Tom, created the following initial network problem of practice and global aim statements based on the reported needs of the school sites:

Problem of Practice*: our current practices are not aligned with students' learning needs today. Many students are failing math.*

Global Aim Statement*: Our practices actively engage students in math and meet their variety of learning needs. We know we're improving when more students are learning and passing math classes with a C grade or better.*

Sarah's stated goals for the network were to build capacity and process, defined as teachers being able to:

- understand and use a driver diagram,
- develop aligned change ideas,
- gather meaningful evidence, and
- engage in inquiry, dialogue, and reflection.

The facilitators decided that the network's efforts would be focused on instruction, and the classroom would be considered their system. With input from the network schools, in addition to their own personal expertise and experience with school improvement efforts, the facilitators also created a driver diagram (Rohanna, 2018). Table 3 catalogs the data-related activities in which the network engaged.

Table 3. Network Data Activities

Date	Description
	Year One (2018)
Jan 31	Problem of practice/Aim statement/Driver diagram introduced PDSA cycles begin
Mar 1	Secondary driver measure attempted
Apr 12	No data specific agenda item
May 17	PDSA cycles end
July 23/25	Network data working group
	Year Two (2018-2019)
Sep 13	No data specific agenda item
Oct 18	Digital primary driver measurement begins PDSA cycles begin
Nov 13	No data specific agenda item
Mar 21	Digital primary driver measurement ends Improvement science survey conducted PDSA cycles end
May 16	Primary driver data analyzed by network teachers

Plan-Do-Study-Act (PDSA) cycles formed the core of the improvement science activities. Some school sites chose a common driver and common change ideas with which to focus their PDSA cycles, while other sites allowed individual teachers to choose their own drivers and change ideas. Every participant tested multiple change ideas, and the majority of teachers chose multiple drivers to focus on throughout the course of the professional development. Appendix B provides a summary of the changes made to the driver of focus and work process change idea by the teachers interviewed for this study.

Additionally, a measure was created by the facilitators to track potential growth in their secondary drivers. The facilitators rolled out their measure in March 2018, but it was discontinued shortly thereafter. In July 2018, four teachers attended a one-day paid meeting, and the facilitators created a digital measure encompassing all three primary drivers based on teacher input. The digital survey was pushed out weekly and allowed teachers to choose one of the three primary drivers to report on. This tool was implemented from October 2018 to March 2019. The facilitators collected this data through a Qualtrix survey, and the teachers did not have access to their response data.

From September 2018 to May 2019, the network continued its efforts to build capacity and processes and included the new goal of "developing practical measures aligned to the driver diagram in order to evaluate whether [the network] was making progress towards [its] network aim." (Rohanna, 2018, p. 86). Documenting this effort constitutes a considerable focus of this study. The network continued to hold meetings on the university campus in September, October, and November 2018. In December 2018 I interviewed seven teacher participants. A teacher strike occurred in January 2019, which delayed the start of the spring semester by six days. Because of this, the network pushed back their scheduled February meeting until March 2019.

In March 2019, a joint interview of the network facilitators, Sarah and Tom was conducted. Based on the responses collected in this interview, it was necessary to revise my teacher interview protocol. The initial protocol was created with the understanding that the participating teachers would be involved in the creation and use of primary driver and process measures to evaluate the efficacy of their change efforts. The facilitators explained that they would be responsible for the creation of driver measures, that the teachers would not see the measurement data until the end of the school year, and that teachers would be asked to evaluate their change ideas qualitatively rather than with process measures. Therefore, the teacher interview protocol was revised to ask teachers about their perceptions of the driver measures that were created for them.

Four of the seven teachers interviewed in December 2018 participated in a follow-up interview in March 2019, and two additional teachers were interviewed at that time as well. In total, nine teachers were interviewed for this report (henceforth referred to as Participant 1, 2, 3, etc.) This sample included teachers from four of the five school sites, and three of the four teachers who participated in the July 2018 primary driver measure collaborations. Table 4 provides a brief description of the teachers who were interviewed.

Table 4. Interviewed Teacher Characteristics

Characteristic	Interviewed Teachers
# years of experience	4 - 26 1 teacher National Board Certified
gender	5 male/4 female
race/ethnicity	Caucasian, Asian American, Latinx, African American
course(s) taught	Math 6, 7, 8, Alg I, Alg II, Geo, PreCalculus, AP Calculus Math 6 ELL, Math 6/7 Accel, Alg1 ELL, Geo ELL Math only: 2 Math as well as Science or Elective: 7
grade level(s) taught	6-8 Five teachers 9-12 Four teachers
which school (1-5)	4 of 5 participating schools

At the March 2019 meeting I asked the 17 teachers in attendance to complete an anonymous survey. Teachers were asked to rank each improvement science principle in order of its usefulness to their instructional improvement efforts. The network held their final meeting of the school year in May of 2019. At this time, the participating teachers were shown and asked to reflect on the primary driver data the network had gathered from the online survey. Additionally, Sarah and Tom informed the participating teachers that funding had been secured for the university to continue to support their improvement science effort.

Organization of Findings

This study explores the practical implementation of improvement science principles in a network focused on secondary math instruction. The primary concerns of this study are the principles most directly connected to data use that were observable. Therefore, I will first present findings that answer research question 1: In what ways was the network's improvement effort problem-specific and user-centered? Next, I will present findings that answer research question 2: In what ways was the network's improvement effort systems focused? Finally, I will present findings that answer research question 3: In what ways did the network's improvement effort use disciplined inquiry? For each question, I report on the participating teachers' perceptions of each principle. Additionally, findings will be included for each question which report factors that helped or hindered the improvement science implementation in relation to the data centric principles observed for this study. These findings reflect factors that influenced the network's improvement effort in terms of its own stated goal of higher student academic outcomes through teacher practices that are more actively engaging.

Findings on RQ 1: In what ways was the network's improvement effort problem-specific and user-centered?

This is an important step in the improvement science process of data use. Primary driver and process measures seek to quantify the effect of network actions on drivers of focus and work processes. According to the improvement science theory of change, this will be effective only if the network has chosen a problem-specific and user-centered aim statement and work processes on which to focus its efforts.

Finding #1: The network adopted an approach that was problem-specific and user-centered.

With respect to RQ 1, I found that the observed network emphasized a problem-specific, user-centered approach to their improvement efforts. The aim of the network was to increase the number of students passing their math classes with a C or better. As evidenced by their driver diagram (Figure 2), the network was expected to achieve their goals through the problem-specific work process of increased teacher practices that actively engage students and meet their variety of learning needs. The network clearly focused on user (student) needs and how to best support teachers in attending to them.

Further evidence of the network's problem-specific, user-centered approach comes from the activities in which the participating teachers were engaged during their professional development time. At least one hour of every network meeting (see Appendix B for sample agenda) was devoted to pedagogical work processes, modeled by the math specialist Tom. Even the fact that a math specialist was involved in the network's improvement effort at all showed adherence to the problem-specific, user-centered improvement science principle.

As would be expected according to the problem-specific, user-centered principle, the network devoted several hours of each professional development day to modeling and engaging the teachers in specific teaching practices they could bring to their classrooms immediately. In fact, 11 of 29 (37.9%) teachers chose at least one change idea that was a direct replication of these activities:

> It was actually directly from one of our workshops, where this true-false activity was proposed, so it's actually in the workshop that I created this lesson and this change idea, and brought it out that way. (Participant 14)

> This is something that I believe [the network] came up with and that's what we've been using, I just adapted it for my classroom. (Participant 17)
>
> Another thing is I really do enjoy seeing some of the teaching strategies that sometimes Tom would throw in there. Those were really fun and I've tried a lot of them with my kids and they just love them. (Participant 7)

In keeping with a user-focused practice, the network centered much of its professional development on pedagogy that connected to daily teaching practices. Participant 17 recognized network activities that supported the Common Core instructional initiative at his site, stating:

> I don't know if it's what the [network's] goal was, but I guess it's the whole idea of the Common Core. How we need to have different ways to see a problem, and how are we going to be able to solve it? It's not just one way. I think that's one thing that I really appreciate about the number of activities that Tom does.

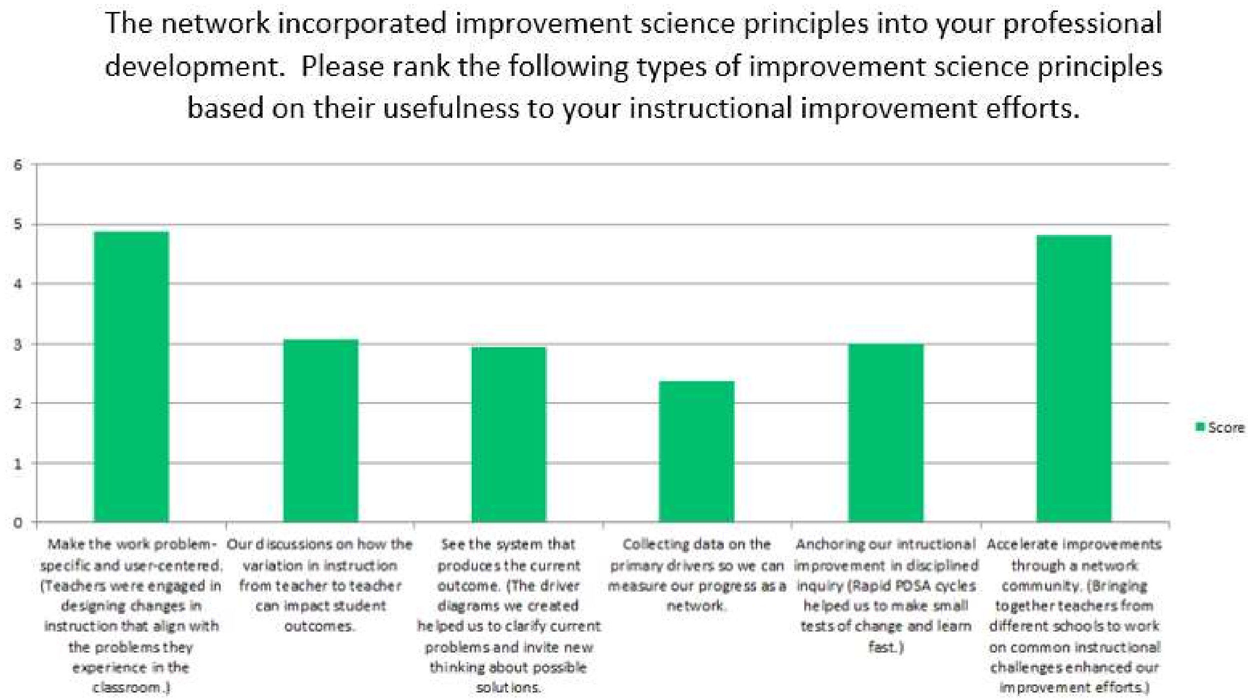
As Participant 11 succinctly put it, "I was most likely to use something in my classroom if it directly applied to what I was doing at the time."

In particular, the ability to choose their improvement driver of focus allowed for a personalized improvement effort. Participant 7 shared, "The reason why we chose these drivers around engaging mathematical discourse is that was always one of my weaknesses, being able to facilitate whole group conversation. During the transition from small group to whole group discussion." Participant 4 also chose her change idea because, "I feel like it's still one of the things that I need to work on, in my practice." The decision by the network to allow teachers to choose their own drivers of focus and change ideas allowed teachers to personalize their improvement efforts to more closely attend to their students. This is indicative of a strong problem-specific and user-centered network approach. However, this decision had negative implications for the network regarding network learning, as will be examined later.

Finding #2: The participating teachers valued the network's problem-specific and user-centered approach.

The teachers themselves also provided evidence of the close adherence by this network to a problem-specific, user-centered network approach. In the comments section of the March 2018 anonymous survey (Appendix F), teachers described the network's focus on teacher actions that supported student learning, sharing, "I liked having the opportunity to make changes that my students needed to directly impact their current learning," and "I think the most important aspect of all the network stuff we've done is that we are able to gather ways to change our instruction to better engage our students and fit their needs."

The results of the March 2018 survey are displayed in Figure 1. Eight of 17 (47%) surveyed teachers ranked the "problem-specific and user-centered" principle as the most beneficial of the six improvement science principles to their improvement efforts.



Improvement Science Principle RANK	1		2		3		4		5		6		Total	Score
Make the work problem-specific and user-centered. (Teachers were engaged in designing changes in instruction that align with the problems they experience in the classroom.)	50.00%	8	6.25%	1	25.00%	4	18.75%	3	0.00%	0	0.00%	0	16	4.88
Our discussions on how the variation in instruction from teacher to teacher can impact student outcomes.	6.25%	1	18.75%	3	18.75%	3	12.50%	2	18.75%	3	25.00%	4	16	3.06
See the system that produces the current outcome. (The driver diagrams we created helped us to clarify current problems and invite new thinking about possible solutions.	6.25%	1	6.25%	1	18.75%	3	31.25%	5	18.75%	3	18.75%	3	16	2.94
Collecting data on the primary drivers so we can measure our progress as a network.	0.00%	0	18.75%	3	0.00%	0	18.75%	3	25.00%	4	37.50%	6	16	2.38
Anchoring our intructional improvement in disciplined inquiry (Rapid PDSA cycles helped us to make small tests of change and learn fast.)	0.00%	0	17.65%	3	23.53%	4	17.65%	3	23.53%	4	17.65%	3	17	3
Accelerate improvements through a network community. (Bringing together teachers from different schools to work on common instructional challenges enhanced our improvement efforts.)	35.29%	6	35.29%	6	17.65%	3	0.00%	0	11.76%	2	0.00%	0	17	4.82

Figure 1. March 21, 2018 Improvement Science survey results.

Further evidence of the value teachers placed in the improvement science "problem-specific, user-centered" principle is seen in that five of 29 (17.2%) teachers changed their PDSA cycle driver of focus, and 21 of 29 (72.4%) teachers attempted two or more separate PDSA cycle change ideas over a five-month period. Teachers reported positively about the flexibility to change their drivers and choose a change idea that supported work they were doing individually in the classroom. One teacher, whose school site had previous experience with PDSA cycles, outlined the importance of the "user-specific" improvement science principle in their anonymous survey comment, manifested through teacher autonomy:

> Given that we have been doing this process at our school site prior to participating in the network, we have gotten more done since the network started because it clarified that we can all work on our own change ideas while still grounding them in common learning. In the past, our school site had pressed upon us to do the SAME change idea and buy in was difficult, if not impossible and group in-fighting was common. By allowing for the differentiation, we could have focused and interesting dialogues that had different starting points but often led to similar conclusions/realizations.

This sentiment was echoed by another anonymous survey response:

> Being responsive to the problems that individual schools are experiencing by allowing time to address them through PDSA cycles ensures more participation from the school. It gives us ownership and time to reflect on issues.

During network meetings I observed teachers focusing on problems that were within the scope of their work processes. They recognized the role instruction had in student learning and enthusiastically engaged in discussions on their change ideas in practice and the observed effects on student comprehension. The majority of the teachers who were individually interviewed for this report also expressed value in the network focusing on topics that were within their control as teachers.

Findings on RQ 2: In what ways was the networks improvement effort systems focused?

This is another important step in the improvement science process of data use. According to improvement science theory of change, to improve a chosen outcome an organization should focus its change efforts on high leverage work processes. This can only occur if a driver diagram of the systemic processes influencing that outcome is created and used to direct improvement efforts.

Finding #3: The network adopted an approach that was systems focused.

With respect to RQ 2, the network in this study displayed a strong adherence to the improvement science principle of seeing the system that caused the current outcomes. Although the initial driver diagram was presented to teachers in January 2018, the final driver diagram

(Figure 2) was the result of several rounds of conversations between the facilitators and the participating teachers. I observed teachers and facilitators discussing the classroom processes that constituted the final driver diagram. In fact, the third primary driver was the result of explicit teacher feedback concerning the number of times students were afforded the opportunity to demonstrate their learning, a work process the teachers identified as influencing the success of their students. The change ideas are "To Be Determined" because teachers continually created new ones throughout the professional development.

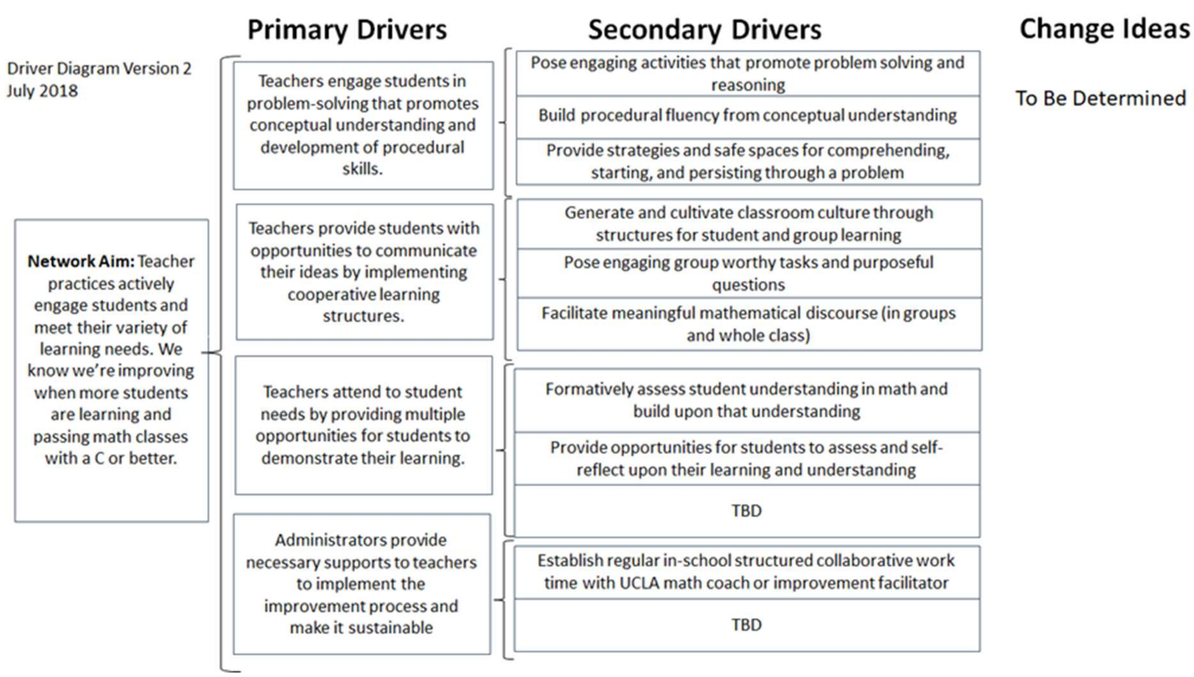


Figure 2. Network driver diagram.

Beginning in January 2018 and in subsequent meetings, I observed the driver diagram guiding the selection of PDSA change ideas. Organized by school site, teachers were encouraged to generate change ideas for their classrooms that directly addressed one of the drivers. Each of the nine teachers interviewed was able to recall the specific primary and secondary driver of focus that their change idea attempted to address. This is strong evidence of the network attending to systems thinking with high fidelity.

During interviews, teachers consistently referenced the driver diagram in connection with their change ideas. As Participant 14 put it, the "three primary drivers that the network has, just kind of outline really precise ways that we can improve instruction." Describing the role of the driver diagram for her site, Participant 4 explained, "I think without that, we would all just be kind of going in scattered directions and wouldn't necessarily be working on moving toward the same kind of goal." A response from Participant 23 was typical of how teachers used the driver diagram to keep their change ideas focused on high leverage work processes, "So for example, driver one is going to be engaging students. So, no matter what activity it is, that's the focus, to try and keep that engagement up."

These statements were indicative of the network's emphasis on systems thinking, which is expected of an improvement effort modeled on improvement science principles. Every teacher referenced decisions they made about their change idea in terms of the particular driver they were hoping to influence, rather than its effect on student grades. In this way, teachers demonstrated an understanding and appreciation for the connection between the drivers of a goal outcome that had been identified by the network, and the choices they were making in their teaching practice.

Finding #4: The participating teachers valued the network's systems focus.

When asked explicitly about the role the driver diagram played in choosing what change to try in their instruction, every teacher reported positively about how it influenced their selection. Participant 3 would "look at the drivers and see how I can use them." Participant 9 said she liked the drivers because they "give me a focus. I feel like that's what the drivers provide." Participant 4 also referenced the benefit of a driver of focus, saying, "I think what it does is it helps you kind of focus your target." Participant 7 went on to explain how the driver diagram could support future teachers that may join the network's change effort:

> Having the driver diagram already set up, I think that part was really useful, because we didn't have to reinvent the wheel every single time. That was kind of cool. Aside from being cool, it was really helpful, particularly when new colleagues come into the fold, or when colleagues leave. So, there's this institutional memory of the work that we're doing and its placement in education. I think that was really cool.

Findings on RQ 3: In what ways did the network's improvement effort use disciplined inquiry?

PDSA cycles are the core of the improvement science principle of disciplined inquiry. What differentiates this principle from other professional development improvement efforts is the use of quantified primary driver and process measures to regularly inform decisions. The majority of the data centric work done by the network observed for this study occurred during the PDSA process.

Finding #5 - The network's implementation of primary driver and process measures were limited.

When asked how primary driver and process measures factored into the network's vision of disciplined inquiry, the primary facilitator described their plan as such:

> I envisioned that there would be data that we can look at as a group and look at over time. I think one driver was around problem solving tasks. "How many did you work on this week?" I envisioned over time, we would see it get higher and higher. There's not a lot of variation. But I had a vision that was my goal that we will be able to have conversations about, "Oh look, we're more consistently using these practices in our classrooms," and there would be some conversation around that. I don't know that we're going to get there on that.

The facilitators recognized that the principle of disciplined inquiry relies heavily on continuous feedback from primary driver and process measures. She explained how the use of these measures to collect data on their drivers undergirded their theory of change for the network.

> So, the idea behind why we're trying to measure those primary drivers is before we to get to the student outcomes, we want to understand is it even related to anything that we're doing? Are we changing drivers one, two, and three? It basically tells us are we moving the needle on our drivers, that should change the student outcome? Or are we changing these student outcomes without moving the needle on our drivers? Then maybe we need to rethink our drivers. So, they basically inform the network as to how we're doing as a whole.

Despite the recognized importance of primary driver measures, I observed limited evidence of the network's use of quantified primary driver and process measures as described in the literature review. There was an attempt in March 2019 to measure secondary drivers. However, the facilitators encountered immediate difficulties. When asked about the creation of a driver measure that reported on such a large number (nine) of secondary drivers of classroom instruction, the facilitators cited the difficulty in creating one common measure on which the five schools would agree, across such a wide variety of drivers, that would be quick and easy for teachers to complete. According to one of the facilitators:

> Maybe they're not practical enough. But at the same time, how do you make them timely? You can't oversimplify it. So how do you create these practical [primary driver and process] measures that are easy for teachers to use and collect on their own? But I don't want them to go straight to student outcome data. The drivers were more around their practices and what they were doing in the classroom. I mean I'm struggling with this still, it's the idea of like we're trying to create some, particularly at the primary driver level, was creating some common measures across schools.

The rubric (Appendix C) they presented to teachers was designed to capture information about the network's secondary drivers. The facilitators asked teachers to fill out the paper by the following week, but the measure was discontinued at the next meeting as no teacher had used the rubric. The primary facilitator described it thusly:

> Nobody did it. Even the people who were comfortable with data, they didn't do it. ...I gave them a template, but well they told me they had too many other things going on. As soon as they left that day, they kind of put that piece of paper away and didn't think much about it.

Participant 23 described the low priority placed on primary driver and process measures by the network in the first year, saying he felt the secondary driver rubric "got swamped with so much other stuff that they [the network] had no time to focus in on it. So, I think that a lot of people just kind of felt 'Out of sight, out of mind.'" With no primary driver and process measures being collected, they could not be used to regularly inform the network's disciplined inquiry as per the improvement science principle.

The following year, a primary driver measure (Appendix C) was created and used by the facilitators to collect data from October 2018 to March 2019. This survey resulted in quantified data that could have been used to inform the efficacy of the network's disciplined inquiry in an ongoing manner. However, the facilitators did not share the results of this measure with teachers until May 2019, their final meeting of the year. This decision by the facilitators runs counter to the intended use of primary driver and process measures to regularly inform the network of the effect PDSA cycles had on work processes and focus drivers.

The facilitators felt quantified instructional process measures would be too difficult and time consuming to create and use during the PDSA cycles. Because of this, it would hamper the ability of the participating teachers to learn by doing. The primary facilitator explained their decision thusly:

> ...in order for [PDSA cycles] to be informative to the teacher, we wanted it to be a little bit more qualitative. We didn't want them to feel like they needed to collect student outcome data or anything like that for their PDSA. So, the use for the PDSA's is for the teacher. It's for it to be informative to them, it's to be timely to them, it's to be easy for them to collect. We made a decision early on that we were going to think differently about data at the PDSA level.
>
> And I wanted fast learning cycles. So, where other schools may use a PDSA that's aligned with some type of testing cycle, those aren't fast. Those might be once every nine weeks. That to me, takes away what the goal of a PDSA is. It's a learning cycle. So, you can keep the cycle fast and collect data that's informative and really easy for them to get. Or you can align it to some type of measure that you might be able to collect in every

> three weeks, six weeks, whatever, nine weeks. That to me takes away the purpose of the whole PDSA cycle.

Another complication in the network's use of primary driver and process measures to inform disciplined inquiry resulted from allowing teachers to change their driver of focus from PDSA cycle to PDSA cycle. Typically, a network would agree on a single driver of focus for its PDSA cycles. The primary facilitator described the purpose of a primary driver measure thusly, "It's a common measure, right? It's something that different schools could all be collecting the same information. Those are measures that you're going to be able to aggregate up from because everybody's using a common measure." A driver of focus allows the network to learn by comparing data across schools and teachers.

As reported previously however, teachers in this network were allowed to change the driver of focus on their PDSA logs, and five of 29 (17.2%) teachers did so. While facilitators and teachers both reported the benefit that allowing teachers to change their driver of focus had on the problem-specific user-centered principle, this decision hindered the network's ability to learn through the disciplined inquiry principle. Teacher autonomy in focus drivers resulted in great variability in the number of teachers reporting on any given primary driver. Additionally, as the reporting was done anonymously, the network lacked the ability to analyze data by teacher.

This digital measure took teachers less than five minutes to fill out, yet still had a highly variable completion rate each week. Even when the number of teachers who reported on any of the three primary drivers was *combined,* the percentage of responses ranged from a high of 20 of 29 (69%) teachers to a low of one of 29 (3.4%) teachers (Figure 3). Even if the facilitators had shared their primary driver information during the course of the improvement effort, it would not have been robust enough to support disciplined inquiry for specific teachers working on specific primary drivers.

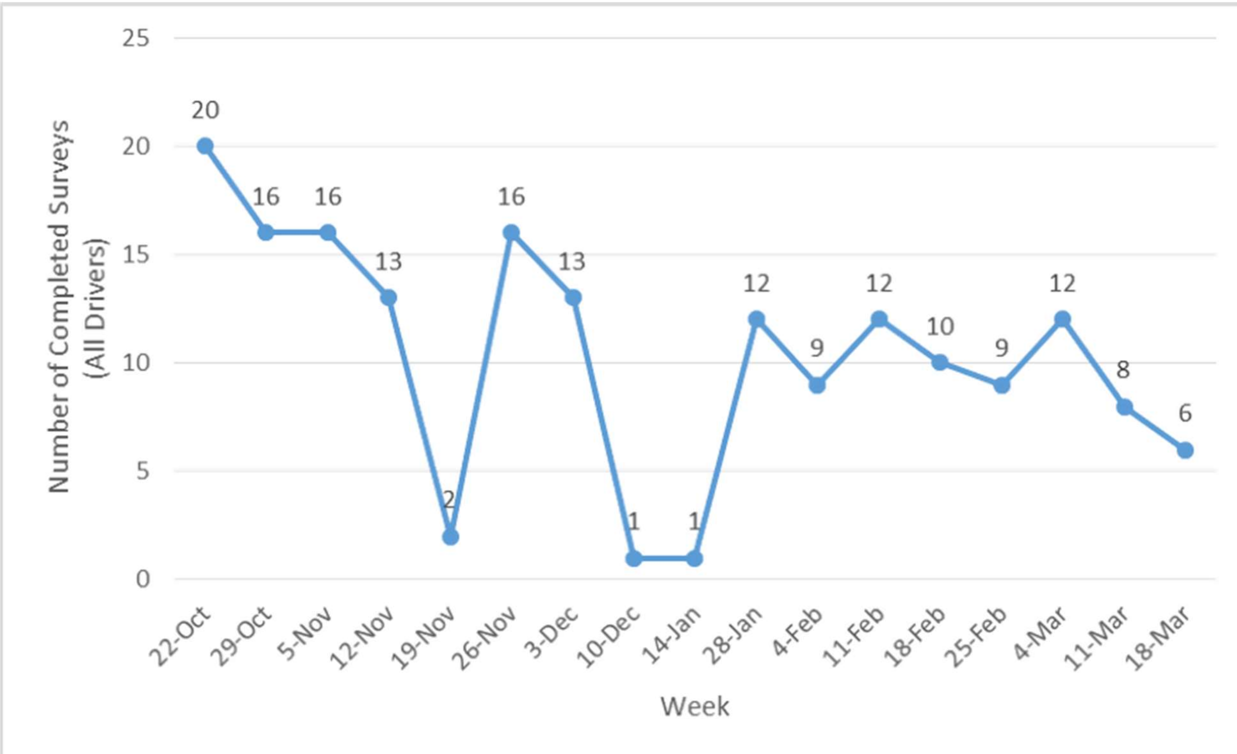


Figure 3. Total # of participating teachers who completed the primary driver measure by week.

By choosing to present the primary driver measure data (Appendix G) at the end of the year rather than at each meeting, the continuous learning of the network was hindered, as a yearly measure cannot inform the two-week PDSA learning cycles the teachers conducted. Improvement science literature often depicts this information displayed as a run chart, with time as the independent variable and the primary driver measure as the dependent. Similarly, the non-longitudinal, combined totals of each primary driver (Appendix G) are not detailed enough to inform specific work processes or to identify individual teachers in relation to a given implementation time. This undermined the efficacy of the PDSA disciplined inquiry, and also precluded the network from attending to the improvement science principle of identifying and addressing the variability in individual teachers attempting those work processes.

Finding #6 – Teachers reported mixed perceptions of the relevance and credibility of the instructional driver measures that were created by the network.

Given this network's focus on improving classroom instruction, teacher buy-in is essential, and teachers must be equal partners in all aspects of the change effort. Driver measures play a central role in the disciplined inquiry principle; therefore, teacher perception of driver measurement influences a network's ability to use them to inform its improvement progress. Important evidence emerged during participant interviews with respect to teachers' understanding of the relevance and credibility of the driver measures used by this network. In order to conduct disciplined inquiry, teachers need to understand the theory behind how driver measures can inform network efforts.

Teacher perceptions of the relevance of driver measures. When asked about the need for driver measures, each teacher was able to clearly articulate their purpose. They understood that the network needed to gather data to conclude whether the work they were doing was having any measurable effect on their drivers. Participant 9 stated, "The network is able to use these protocols, to see how effective is it going. Are they seeing the growth?" Participant 4 framed it,

> Just like we have to collect data on our own process, it's almost like the collection of data of the bigger idea, which is for the network, their improvement cycle, so they're collecting data on whether or not teachers are being more methodical about making improvements in their classroom.

Participant 23's understanding was that "it was a way for the network to gather data and look at how much were we doing as far as implementation of strategies. And then I guess kind of using that to determine the effectiveness of our efforts." Two teachers even referenced the fact that the facilitators needed data in order to justify the grant being used to conduct this improvement science professional development, with Participant 17 saying, "I know it's what keeps the party going basically, is being able to gather evidence on being able to see if things are moving in the right direction towards that primary change."

Teacher perceptions of the credibility of the driver measures that were created.

While teachers understood the practice of using driver measures to evaluate their improvement efforts, they disagreed on the credibility of the two driver measures that were created by this network. In particular, teachers varied in their perception of the ability of the secondary driver measure (Appendix C) that was attempted in March 2018 to accurately report on drivers connected to student engagement. This is important, as 15 of the 29 (51.7%) teachers conducted PDSA cycles on drivers connected to engagement. The fact that none of the teachers participating in the network completed the rubric is also indicative of a lack of credibility in its ability to inform disciplined inquiry.

The teachers interviewed for this study expressed various degrees of confidence in the March 2018 rubric. Participant 3 brought up the subjectivity in measuring engagement by sharing, "The one thing I would say is there are different ways kids can get engaged, that I don't know if that rubric addresses them all," referencing his use of an online discussion board. Participant 7 added, "You think that people are engaged, and then you kind of see, oh, actually that person was just on their phone the whole time. I think teachers have a tendency to overestimate levels of engagement among students."

Participant 14 first expressed confidence in the rubric to measure engagement, stating, "I mean, I think this is a clear rubric. It's got good points, and I feel like people can find themselves and their experience within this rubric," but later qualified his position, adding "When you're talking about engagement, and getting people to talk to each other in groups and things like that, it's hard to measure that with data and numbers." Finally, Participant 4 brought up the conflict between quantity and quality when measuring an instructional driver such as student engagement:

> So, for example, I can say that I am setting up collaborative structures. It's easier for me to prove that I'm setting up collaborative structures so that my students can engage in mathematical discourse. It's another thing to say that the students actually did do that with any quality.

Teachers also reported their credibility as being influenced by the potential for the results to be used by administrators for evaluation. Participant 13 explained,

> Whenever you're dealing with self-reporting, and depending on trust issues that teachers often have because of previous experiences with administrators and things like that, I don't know what the inclinations are to sort of self-report in a way that's inaccurate or false.

Participant 7 expressed confidence in the self-reported data, while also acknowledging the ability of evaluation to alter fidelity:

> I don't think that when people are putting their responses in, it's at all like misinformation, or distortion of the truth or anything like that. I don't know. I don't think so, because I don't think that any of this is being tied to a professional evaluation or anything like that.

Finding #7 - The network placed a strong emphasis on disciplined inquiry.

With respect to improvement science disciplined inquiry, we see evidence of the divergence in implementation fidelity due to the lack of a driver of focus and common change idea. However, in accordance with the disciplined inquiry principle, I observed the participating teachers pick change ideas based on their work processes, make predictions about possible results, and suggest sources of data to test the efficacy of their change ideas. Through these practices, the network strongly emphasized the Planning stage of improvement science PDSA cycles.

The disciplined inquiry of improvement science is designed to generate network learnings through rapid PDSA implementations. Each PDSA cycle was conducted over a two-week period, and a PDSA log was digitally filed with the network. From October 2018 through March 2018 the 29 participating teachers generated 99 PDSA logs, for an average of 3.4 logs per teacher. The number of logs completed by an individual teacher ranged from one to six (Figure 4). Excluding Thanksgiving and winter breaks, and the six-day lockout, the average teacher completed approximately one PDSA log a month. This is less than what the network requested, but was within the facilitators' expected output. If three logs were expected over this time, 20 of 29 (69%) teachers completed the expected amount.

Several teachers described testing change ideas without formally completing a log; however, each cycle was to be documented by a PDSA log. Teachers discussed the results of the PDSA logs in the monthly network meetings as well as at their school sites.

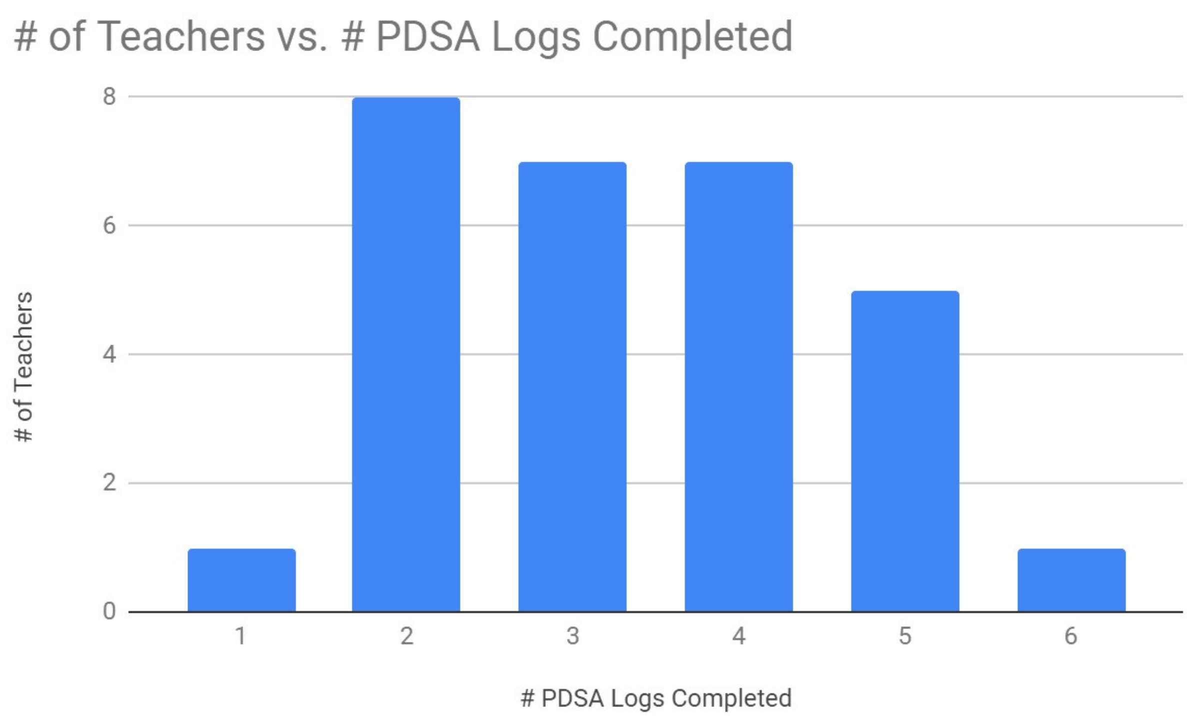


Figure 4. Amount of PDSA logs completed by participating teachers.

The PDSA log asked teachers to think very specifically about their work processes and what information they would use to evaluate the effectiveness of any changes. The log asked teachers to choose a driver-aligned change idea and make a hypothesis about how their change idea would impact student learning. Teachers were explicitly asked about the information they would gather to evaluate their prediction and to reflect on their change idea once implemented. Teachers were given at least one hour each network meeting to discuss the results of their previous cycle and also to plan their next cycle. During this time, I observed teachers discussing their change ideas in terms of their work processes and the driver diagram. The robust PDSA logs and the network time devoted to disciplined inquiry provided evidence of the network's strong attendance to improvement science disciplined inquiry.

Early on, the facilitators made the decision to allow teachers to substitute unquantified data in place of quantified primary driver and process measures. When asked why this decision was made Sarah referenced how School 1 had previously wrestled with using primary driver and process measures to regularly guide their PDSA cycles:

> A lot of their teachers were getting really stuck on how do they measure their change idea. And what was that measure supposed to be and are they supposed to do a run chart for their change idea. The work around the PDSA's wasn't really happening very well. I heard a lot of people saying, "I don't know what data I'm supposed to collect." And they got the impression if they didn't collect the right data, then they shouldn't even do the change idea because how do they know whether it was effective? And so we made a decision early on that we were going to think differently about data at the PDSA level.

The network's decision represented a break of sorts from the process measures needed for disciplined inquiry. While the network did push teachers to think about what information they could collect to evaluate their change ideas, the facilitators' decision not to quantify PDSA data collection constitutes a serious divergence from the improvement science principle of using process measures to regularly inform network learning through PDSA cycles.

Though the network did not use process measures, the facilitators did encourage teachers to regularly inform their PDSA cycles with other data. Sarah strongly believed that the teachers could use unquantified information such as samples of student work and their personal observations to decide whether their change idea had been a success. When asked if there was any concern about the subjectivity brought about from this divergence from the improvement science model, Sarah elaborated:

> So, this is one of the things that people were bringing up to me and it was this idea of, "Well, how do I know if it's effective if I don't have this measure and I don't have this pre and post data?" The role of the change idea is just to experiment with a new practice. It's not to determine whether it was causal or not, whether it worked or didn't work. But I feel pretty strongly if the teacher tries an idea in their classroom and they collect some evidence around it and they said, "You know what? I think that was an improvement. I do feel like my students were more engaged. I do like how this works with my students. I'm going to continue doing that." To me if it's subjective, I'm okay with that because I think most teachers have a good understanding in their own classroom with something that's having the effect that they want it to have or if it's not.

This network chose to replace process measure data with unquantified data, which the facilitators believed would increase teacher buy-in for the improvement effort. Although process measures are most consistent with the principle of disciplined inquiry, the literature does describe a role for unquantified data such as that collected here to inform small scale improvement efforts, a proof of concept so to speak. Overall, the network in this study devoted much time and effort to a sustained disciplined inquiry effort.

Finding #8 - Teachers experienced difficulty with conceptualizing process measures that could regularly inform their PDSA cycles.

The network did not gather primary driver and process measurements in year one, and teachers were not shown their primary driver measure data in year two until May 2019. The participating teachers were asked to use the unquantified information they gathered during their PDSA cycles to determine the success of their change ideas on their drivers of focus. Teachers highly prized student work and classroom observations for their ability to provide insight into student learning, which will be discussed next. The network's ability to learn through disciplined inquiry was hindered by the lack of process measures to drive decisions about the efficacy of change ideas. This was in part due to the difficulty teachers had in conceptualizing data about their work processes separate and apart from student work.

After describing process measures in interviews (such as timing teacher talk versus student talk), teachers were supportive and were able to explain how they would use such information to modify their instruction, but had trouble coming up with process measures themselves. When teachers were asked about what information they wanted to evaluate their change ideas, none of the teachers reported information about their work processes. Even when their secondary driver described teacher centric actions, such as "*pose* engaging group worthy tasks and purposeful questions" or "*provide opportunities* for students to assess and self-reflect upon their learning and understanding," they chose to gather information about student comprehension and engagement instead of the process measures needed for disciplined inquiry.

The exchange below demonstrates how a difficulty in conceptualizing process measurement hindered the teachers' ability to learn about the efficacy of their new work processes. In our interview, Participant 17 identified a potential change idea based on a presentation by the network math specialist, Tom.

> Participant 17: Tom had a really cool lecture/presentation about the learning zone and the performing zone, and it just made me realize I'm not giving the kids enough time to reflect on their work, to be able to do better in the future and when they do their performance, the performing zone.
>
> Interviewer: The kids need some more time to reflect. That's a change idea. So, let's say you were to be doing a PDSA log with me verbally and that was your next change idea, and you were filling out this, and you were getting to the part with data. What kind of information would you want to collect, for you to be able to look back and say "is this change idea being successful?" What information would you as an instructor want to know, to evaluate whether that was a positive change in your instruction?

> Participant 17: I guess just the reflection itself, because I think part of being an instructor is being able to take their input into how I could tailor the next segment or lesson, based on what their reflections are.

The change idea identified by the Participant 17 is the *amount of time* he is giving his students to reflect. Process measurements he could have gathered include recording the number of times he asked students to reflect in a week, or recording the amount of time he gave students to reflect. He could have created a rubric to self-evaluate the quality of the reflective questioning he employed. Either data could be compared week over week to track the change in his work process through disciplined inquiry. However, Participant 17 chose to examine the student work instead.

While certainly informative, student reflections do not represent a direct, quantified measure of the amount of time provided for student reflection. Student work lacks the specific actionable information that could inform disciplined inquiry like a run chart of rubric scores would allow. In the majority of the interviews that were conducted, teachers were unable to describe an appropriate process measure to inform the PDSA change idea they had chosen. When similarly questioned how student work and observations would help her to evaluate her driver, Participant 13 recognized the difficulty in using unquantified information in that fashion, responding, "I mean, I honestly don't know how to answer that."

Finding #9 – Teachers used the visible learning from observation and student work to evaluate their work processes.

The teachers in this study collected qualitative data on students rather than quantified process measures, which is to say that they collected data on the student reaction to their change in work processes rather than data on the work process itself. With the understanding that the observed network took an alternate approach to the information it asked teachers to gather to inform the efficacy of their change ideas, an ancillary finding emerged.

Every teacher interviewed reported that the most important feature of the information they used to evaluate their instruction was its ability to make student learning visible. It did not matter what type of data were discussed, when asked why it was chosen, without fail teachers stated the purpose was to shed further light on student understanding of content. This laser focus by teachers on student comprehension contributes to an earlier finding, the difficulty teachers had in conceptualizing information they could collect about their own work processes.

Overwhelmingly, teachers chose student work and their own observations as the information to evaluate their PDSA cycle change ideas. All 29 (100%) teachers used student work for at least one PDSA cycle, while 22 of 29 (75.9%) teachers used observations of students for at least one PDSA cycle. When interviewed, teachers reported choosing this information because student work and their own observations and interactions with students best allowed them to assess student comprehension. Participant 14's response, "I was looking to see, mostly, their content understanding and procedural understanding" was typical.

Examples of visible student learning. If we examine a PDSA log from three participating teachers, the type of information typically chosen by the teachers in this study is seen and their rationale was probed during the interview.

For one of Participant 3's PDSA logs (Figure 5), we see his chosen secondary driver was to "Provide opportunities for students to assess and self-reflect upon their learning and understanding," and his change idea was to engage the students in an error analysis activity. During his interview, when asked about the success of his change idea, he chose to evaluate its efficacy in terms of information gleaned from student work,

> The activity didn't work, because the way they were explaining it didn't really show they knew what they did wrong. For example, they would fix their signs or something on the work part. But the explanation part it would just be like "I would need to check my work" or "I would need to read the question" or "I need to just try harder."

He proceeded to further use student work to evaluate his change idea, “So those explanations that they were giving were very vague and they didn't give me enough information still of like what did they learn from it? What are they going to do different from that?” deeming the error analysis activity ineffective precisely because it did not increase his visibility into student learning.

PDSA Form: Participant 03

Date	9/14/18 - 9/28/18
Change Idea	**Describe:** I will incorporate an error analysis component when students receive their assessments back **What is the change:** The process of error analysis for retaking **Aligns to:** Multiple Opportunities to demonstrate learning (Primary Driver) and Opportunities to self-reflect and re-assess Secondary Driver
Plan	After administering an exit slip and test, I will implement the change of the error analysis. I will provide time in class to model the process and explain what the new analysis will look like

Predictions: What improvement do we think will happen?	**Questions:** What do we want to learn from this cycle?	**Data:** What information will we collect to answer our questions and test our prediction?	**Results and Next Steps:** Did you meet your prediction? What were the results? What did we learn? What will we do next? *(completed after implementation)*
I think that having students analyze and writing it down their thought process will help students recognize what they know and what they don't (their strengths). By figuring out their mistakes and correcting them, they should demonstrate a higher level of mastery when retaking	What are students really thinking in their thought process? How are students with lower proficiency able to analyze their work? How can we get higher proficiency students to find another method?	Test/Exit Slips with error analysis form Schoology grades to compare the learning target Teacher Observations	-Some of the students had a hard time explaining in words what they can improve on or what they did well -Lower proficiency students had a hard time finding something that they did well in the problem in the error analysis -I will have to compare the retakes score with their original score -Most of the explanations were not full sentences.

Figure 5. Example PDSA log - Participant 3.

Participant 4’s log (Figure 6) relied on observations of her students to evaluate her change idea, which was linked to the secondary driver: “facilitate meaningful mathematical discourse (in groups and whole) class.” In her activity, some students were trained to lead their classmates in a discussion. When asked about the success of her change idea, Participant 4 replied:

> I do remember doing this. I remember listening for the questions. I remember hearing them so sporadically that it was very obvious, from the very beginning, this isn't working for some reason. It sucks because I know, technically, we should be all scientific about it

> and have a tally sheet. But I feel like a lot of times, when I collect "data," I put in quotations. It ends up being sort of anecdotal.

Participant 4 acknowledged the recommendation of the facilitators to use a process measure, even indicating it as the information to be collected on her PDSA log. Unable to follow through on the tallies, she instead made a decision based on student observation that her change idea was not, in fact, "facilitating meaningful mathematical discourse."

PDSA Form – Participant 04

Date	**One- or two-week cycle:** September 17 through October 1
Change Idea	**Describe:** Implement the "Pairs Check" (Rally Coach) strategy for problem #1-90. **What is the change:** In the past, I have had students work individually and then share as a group. **Aligns to:** (1) Provide students with opportunities to communicate their ideas by implementing cooperative learning structures and (2) facilitate meaningful mathematical discourse in groups and whole class.
Plan	I will utilize the "Pairs Check" (Rally Coach) strategy for Geometry problem #1-90 during Period 3 on Monday, September 17. Students will be separated into Partners A and B; Partner A will tell their partner how to create a new shape from one of the four given triangles using either reflection or rotation; Partner B will draw the shape according to their partner's instructions. The process will continue back and forth for five minutes. The pairs will then share their shapes with the other pair at the table and group them into "same" and "unique" categories.

Predictions: What improvement do we think will happen?	**Questions:** What do we want to learn from this cycle?	**Data:** What information will we collect to answer our questions and test our prediction?	**Results and Next Steps:** Did you meet your prediction? What were the results? What did we learn? What will we do next? *(completed after implementation)*
I predict that students will: Increase their use of the related academic vocabulary. Increase the proportion of productive talk to total length of discussion.	Will a more structured talking protocol make students more willing to engage in discussion? Will more practice in structured eventually lead to self-directed productive talk?	Tally number of times I hear related academic vocabulary words. (e.x. Reflect, rotate, symmetry, triangle, square, rhombus, etc.) Record one group with my phone and determine length of on-task discussion.	At first, it took a few minutes for students to get started using the protocol. It's possible students were confused by my instructions and weren't sure what to do at first. After student A finished, however, student B was more likely to begin speaking immediately. The protocol took longer than I expected, however, and we ran out of time after each duo created, on average, three shapes. Next time, I might be more strategic choosing which student goes first so that a more confident student can model the process for a less confident student.

Figure 6. Example PDSA log - Participant 4.

Participant 17's secondary driver (Figure 7) was to "provide strategies and safe spaces for comprehending, starting, and persisting through a problem." In the following exchange he described how he used student work from his change idea, which was a graphic organizer with a box for student work and written explanations, to make adjustments to his work process.

> Participant 17: That's where I would try and look, to see through student work. The score sheet will tell me they didn't do so well, but as a teacher I have to dig deep and see. Prime example actually, is just talking about different denominators, I saw that with the multi digit multiplication. The kids were not getting it one particular way, so I had to introduce another way, which is the area model way.

Interviewer: How did you decide to try that second technique, versus the one you tried the first time?

Participant 17: I checked their work. I said okay. We're not getting it this way, and then I saw, for example, when you're multiplying for second and third digit, they don't put additional lines or placeholder, and they usually just start in the first digit. I address that with them…In this case I saw almost close to half were not doing the standard way. I showed you guys one way, and I'm going to show you another way that you could do this, and then I pose it to them as whichever you feel like is easier to you.

Interviewer: That makes sense.

Participant 17: Trying to identify wherever the misconception is, and why.

Here, the graphic organizer made student learning visible, allowing the teacher to attend directly to his chosen secondary driver by providing his students with a new strategy in which to comprehend, start, and persist through the content.

Date	One- or two-week cycle: 9/17/18 to 9/21/18
Change Idea	**Using the graphic organizer for story problems will increase the number of students who cas successfully solve a story problem without the help of a teacher.** **Primary driver: Teachers engage students in problem solving that promotes conceptual understanding and development of procedural skills.** **Secondary driver: Provide strategies and safe spaces for comprehending, starting, and persisting through a problem.**
Plan	Introduced the story problem graphic organizer to all students. By Friday, September 21st, we will collect and score the grade level story problem we created as a department.

Predictions: What improvement do we think will happen?	**Questions:** What do we want to learn from this cycle?	**Data:** What information will we collect to answer our questions and test our prediction?	**Results and Next Steps:** Did you meet your prediction? What were the results? What did we learn? What will we do next? *(completed after implementation)*
Students will need guided support to start, but will be able to complete a second problem of a similar structure with the graphic organizers in small groups without teacher support.	Will the graphic organizer help students solve story problems? How can we better support students in solving word problems?	We will collect the students' story problem, and score each assessment using the rubric. We will also look for how students have articulated their thinking on paper.	Students were able to do a similar problem without teacher support at around 80% completion. The graphic organizer definitely helped students feel more successful because they were able to start the problem and helped other students be more analytical in solving the problem. This was also a great way to build on prior knowledge of the students because I was able to check what they have Furthermore, writing the sentence at the end helped bring a conclusion and full understanding of the problem. Next steps would be challenging the students more with multi-step and accommodating for student needs.

Figure 7. Example PDSA log – Participant 17.

When describing student work and observations of students, teachers frequently used them as the basis to probe students to further make their learning visible. Questions such as "Why do you think it's that?" (Participant 11), "So what did you do? What step did you take?"

(Participant 13), and “How did you deconstruct the problem?” (Participant 9) are indicative of the way in which teachers used student work or observations to begin a conversation designed to uncover what students knew.

Every teacher commented on the complexity of the interaction between their practice and student comprehension. They used student work and observations to make on-the-fly and later, reflective, decisions about their change in work processes. When asked if they had unlimited resources and what information would best help them evaluate the effectiveness of their instruction, they consistently referred to video, peer teacher/administrator observations, or an assistant to help with student observations as the preferred types of information to assist them in evaluating the efficacy of their change idea. Participant 7 described the value in a more accurate record of his questioning of students to make learning visible.

> Interviewer: What kind of information would ... If someone were to observe you, not for evaluation, but for the purpose of improving your instruction, what kind of feedback would you find helpful?
>
> Participant 7: Definitely the range of questions, and these are things that we worked on in prior years, was alright, say, “Well, what is your line of questioning? What are the questions that you propose to ask, and did you ask?” Because the classroom dynamics are ever-changing, and so kids are bringing up something, and you're like, “Oh, now I've got to ask a different kind of question,” because the old question was going to be different... And the type of question that's being asked, either all student to teacher, teacher to student, student to student.

When deciding how to inform their PDSA cycle disciplined inquiry, teachers looked to student work, classroom observations, and context to help them build a picture of student understanding to inform their improvement efforts. Here is where successful work processes, built through reflection and pedagogical professional development (content and questioning techniques), were frequently mentioned by the teachers. A majority of teachers referenced turning to their "toolkit" of teaching strategies (tacit knowledge) that they have routinized in response to the chaos of a typical classroom.

Finding #10 - Teachers reported a lack of time and bandwidth as barriers to disciplined inquiry.

When discussing the practices involved in improvement science disciplined inquiry, teachers consistently brought up concerns about the lack of time and the broad array of tasks they are required to attend to in their limited time. By far the longest and most passionate response in every interview occurred when teachers were asked if they had adequate time to collect information, reflect on their practice, and make changes to their instruction. Often, teachers would shake their heads "no" before the question had even been finished. Each teacher reported that lack of time and competing obligations negatively impacted their ability to conduct disciplined inquiry with fidelity.

Time and bandwidth hindered the collection of process measure data. While teachers recognized the value in using process measures (when described to them) about their work practices to help with evaluating the effect of their change idea, they expressed doubts about the ability to capture it efficiently themselves. In the following exchange this teacher recognized the value of collecting process measure data, but highlighted the tension that many teachers cited between using process measures and attending to student learning in the moment.

> Participant 4: I was, let's see what I wrote, was that I was going to just do a quick count of the pairs that were able to graph their equations within a given time limit.
>
> Interviewer: And do you remember, were you able to actually go ahead and count the pairs?
>
> Participant 4: No.
>
> Interviewer: Okay. Maybe you could talk about why you weren't able to do that?
>
> Participant 4: A lot of it just has to do with all the different balls that you have up in the air when you're live in the classroom. And between making sure that all the students

> understand the instructions, and the task, and then managing any groups, or pairs, that might be stuck just with the content material. Dealing, then, with potential disengagement, and how to get students back on track. And then, you've got 36 students. I find it very difficult. It's like one more thing that I just can't- Add on top of all the other stuff. And my primary concern is always, do they know what to do? Are they able to do it? And then, what kind of support do they get when they're stuck? I mean, it [process measure data] would be incredibly helpful for me. But, given the amount of work sometimes, that can be involved just with doing that, it's ... To have one more thing to add onto the plate. If there was someone to do that for me…

When Participant 9 was asked about collecting process measure data she replied,

> I would probably just add something more to the worksheet. I wouldn't want to add ... my thing is I don't really necessarily want to add anymore for me, like any kind of checklist because like I said, I want to be present and really paying attention to what they're doing in answering these questions.

Despite this tension, she still acknowledged the value of quantified information on teacher work processes, saying she thought "it would definitely drive my instruction."

Participant 9 described the choice teachers often face between supporting student learning in the classroom and attending to improvement science data collection:

> I've been in a situation where you have the checklists. And sometimes you spend more time checking than engaging. You have a 50-minute period and maybe this instructional activity takes 30 minutes. So, I don't want to be spending my time checking off boxes.

Participant 14 shared a similar decision at times to abandon quantified data collection:

> That's a struggle with exit tickets too, is that at the end of the day, you have 100 little pieces of paper, that you have to grade, and input, and put in. And so, I switch to informal exit tickets oftentimes, where I'm just doing a little problem and then an eyeball check, and sometimes a little check on a notebook or something like that. But to have things formally graded, and assessed... That would be amazing.

Participant 17 also described keeping up with multiple data collection demands:

> Definitely data collection, outside of my assignments, has been hard. I keep up with the network's PDSA log, and I keep up with the data that we have to provide for our department, and I keep up with my grades. But I feel like already, I'm bringing that heart attack closer.

Time and bandwidth hindered individual disciplined inquiry. Teachers frequently described tension between the disciplined inquiry required of PDSA cycles and their other classroom obligations. Participant 23 explained his difficulty in finding time for PDSA cycles:

> Just the different things that I'm doing on a daily, weekly basis. Sometimes my periods run where I'm literally flying around as much as I can for 50 minutes and I don't have that time to really put that quantifiable aspect into PDSA's.

When asked if the number of tasks she was asked to attend to in the classroom was prohibitive to disciplined inquiry, Participant 4 stated, "I would say 95% of the time. It is rare when I get an opportunity during class to think about anything but the immediacy of what's going on." Participant 13 emphasized the premium teachers placed on even a small period of instructional time:

> You know, five, 10 minutes, 15 minutes in the classroom, it's really precious. If you're gonna put something else on the plate, which I now have to think about, it becomes a little more difficult. I feel like if I'm already getting pulled out of class for like assemblies and other stuff, so it just becomes an issue of like, oh my God, I'm not really utilizing the time. So, I think it's a matter of efficiency really. And a lot of times it becomes difficult.

Participant 7 referenced tension between the disciplined inquiry obligations of the network and his school's focus on mandated state testing,

> The network is more of like, this is something that really supports us and helps us out. But then the other things are more of obligations. I just feel that a lot of what we end up having to focus on is the state test. The test is how our school is evaluated, and that is the thing that we need to be focusing on.

Participant 11 also brought up the amount of teacher time that is consumed by state testing,

> All of the teachers at our school are all spread out pretty thin. I just feel that a lot of what we end up having to focus on is SBAC. If we can find some ways to all have the network activities aligned with that, then I feel that would be amazing.

Time and bandwidth hindered collaborative disciplined inquiry. An important component to improvement science is the idea that networks learn faster than individuals. Therefore, time to share information from PDSA cycles and discuss future practices with colleagues is an essential component of disciplined inquiry.

Eleven of 17 (64.7%) teachers surveyed at the March 2019 meeting chose the ability of the network to accelerate learning as either the first or second most important of the six improvement science principles in supporting their improvement efforts. This was confirmed in interviews as well. Participant 3 enjoyed “working with other teachers to get different perspectives.” Additional teachers shared similar sentiments, such as Participant 4: “The other benefit that I've really enjoyed is being able to spend some time with people from other schools” and Participant 9: “The sharing because I think the more tools you have in your tool kit, the more effective you could be.”

Unfortunately, time and bandwidth limitations frequently impinged on the network’s ability to attend to collaborative disciplined inquiry with fidelity. When asked if enough time had been provided to engage in these discussions, Participant 17 replied, “Structured time, I would say no.” Participant 4 explained how limited department time meant no time for their department to share any learnings from their PDSA cycles,

> We'd been given very little time to do any kind of department work. And so, then when we do meet, we tend to have to discuss operational issues. So, we have a meeting on Monday, but we're not talking about PDSA, we have to talk about our schedule for next year, and things like that.

Participant 4 went on to explain that because the contractual department time was not enough, that their “department meets every other week, as well, after school, and that's uncompensated.”

When asked how much time factors into whether they were able to attend to the improvement science activities with fidelity, Participant 23 replied, “That's probably, I'd say 85% of the reason why things are done or not done.” When asked in particular about whether he felt there was sufficient time to have conversations with his colleagues concerning their PDSA results, he further elaborated:

> So, there's two problems. One of them is on an individual level, which is your individual time in your specific classroom and doing it. The other thing is the systematic one, which is your department. We all want to work together. I think if you gave us the opportunity to plan and put something together, we could implement it. So, I think the department time, we don't have enough of it.

While teachers consistently reported appreciation for the monthly pull out days for this professional development, even that required extra work as teachers described having to prepare plans and then refocus the class after a day with a substitute teacher. This was reflected in the attendance of the participating teachers. Of the 29 teachers from the five school sites, approximately 20 would be in attendance during a given session. I also observed teachers who arrived late and left early, due to other professional obligations.

During the six weeks before the first professional development meeting in March 2019, only two of the nine interviewed teachers conducted any PDSA cycles at all. Participant 4 explained, “This semester, we've been given very little to no time during our Wednesday professional development time, and so I know my peers and I all came in cold. In other words, we didn't have our PDSA forms completed.” Participant 11 shared:

> I did a few of them that were not necessarily written in. Because, like I said, it was a little bit time consuming. I'll be honest. I know that was one of the biggest complaints from my department, the time that it took up to complete.
>
> She went on to say that she completed PDSA work during the network meetings, but:
>
> I don't think other than the times when we're [here] or we're working with the network when they come out there's no structured time to talk about it. It really is just during the

> [network] time that people are present. Other than that, it's just mind over matter. We don't see it. It's not there.

As a theory of change, improvement science requires PDSA cycles and network discussions to lead to network learning, but these activities can tax teachers' limited non-student time. Participant 14 best summed up the tension between time and the disciplined inquiry required to improve work processes through the PDSA cycle:

> Because that PDSA lesson did take me a lot longer than my other lessons would take, as far as the planning and the process and all of the work that I put into creating it. And that's a very interesting thing that I find is, the more energy and the more effort I put into the [lesson], there is generally a direct correlation with the amount of success that the lesson has. So that's definitely one, a really big takeaway for me.

Conclusion

The findings outlined here have important implications for future networks attempting to employ improvement science principles to support an instructional improvement effort. Networks will need to address several potential barriers to implementation, including the concerns teachers reported about their time/bandwidth, and the difficulty of creating primary driver and process measures for the instructional drivers of student learning, such as quality engagement and meaningful discourse. Given this report's finding that teachers most valued information that made student learning visible, future improvement efforts would benefit from structuring limited teacher time around activities that best highlight student comprehension, such as analyzing student work, developing effective questioning, and teacher observations.

No discussion on variation in teacher practices. One of the six core tenets of improvement science is that much of the gain in an improvement effort can be realized by uncovering differences in effectiveness between teachers or schools in moving the systemic drivers. Because teachers and schools were not required to work on a common primary driver, and in fact, many teachers moved fluidly between the three options, it was not possible for the network to examine the data for variation among teachers. Quantified primary driver and process measure data is essential to this process, and without it no discussion by the network about variation was possible.

Chapter 5. Discussion

Discussion of Findings

This study reported on observations of a network improvement community working to improve secondary math instruction. The particular focus was on how the data centric improvement science principles could support this effort. Therefore, the findings presented a narrower picture of the improvement effort than intended. Improvement science as outlined by Bryk et al. (2015) is comprised of six principles, and network improvement communities take time to set up and mature. It is not expected that a network would fully attend to each principle in the first two years of its existence. These findings merely report on the current state of one particular ongoing improvement effort in an attempt to provide insight for future implementations. The teachers interviewed reported very positively about the network's attention to community learning, one of the most important improvement science principles for an early network. As growing the norms necessary for network learning was a major goal of this young network, this should be viewed as a strong start.

With respect to the improvement science principles of focus for this study, teacher interviews and my observations and document analysis support the finding of a clear problem-specific and user-centered network approach (research question 1). The work of the network was clearly centered on teacher and student experience, and teacher work processes were emphasized throughout network meetings. Additionally, this study documented substantial evidence that the observed network was very systems oriented (research question 2). The driver diagram was a central part of every PDSA log, and teachers referred back to it frequently. During interviews, teachers expressed strong support for the network's systems focus and its problem-specific and user-centered approach.

This study's findings concerning the network's disciplined inquiry (research question 3) are best viewed in two parts. First, several findings support a robust network effort to conduct repeated PDSA cycles based on mainly on student work and teacher observations. Second, several findings report on the ability of the network to use that information (in addition to the networks' driver measures) to regularly inform their PDSA cycles. The evidence documented in this study depicts a network clearly dedicated to disciplined inquiry. They expended time and effort on information-driven PDSA cycles, which often involved work processes based on pedagogy presented by the network itself. Teachers greatly appreciated the ability to change their driver of focus and change idea.

However, these same decisions greatly hindered the network's ability to generate collective learnings from quantified primary driver and process measures as outlined by the improvement science literature. PDSA cycles were not informed by process measures, nor was PDSA information captured in a way that allowed for run charts or comparison across teachers and schools. Information from the primary driver measure that was employed was not shared regularly with the participating teachers and may not have been sensitive enough to work process changes. These challenges may be due, in part, to the difficulty of accurately and efficiently measuring the drivers of classroom instruction amenable to changes in teacher work processes (such as quality discussion or productive engagement). Therefore, PDSA cycles were informed strictly by unmeasured, informal data on work processes.

Finally, the participants in this network highlighted the well-known, but often ignored, reality of teacher time and bandwidth constraints. This was found to impede all aspects of the improvement implementation, but given the regularity with which primary driver and process measures are to be gathered and referenced, it is impossible to ignore the tension between the idea of "practical" measurement and the hustle and bustle with which classroom teachers endure daily.

Implications for Future Implementation of Improvement Networks

Time and bandwidth concerns must be addressed. The most important implication for schools attempting to better classroom instruction through an improvement science implementation is the lack of time and bandwidth available to most teachers to attend to tasks that are not instruction. The teachers in this study cited the lack of time and bandwidth as major concerns when discussing the improvement science professional development in which they were engaged. Whether it was creating the driver diagram, consistently using primary driver and process measures, conducting PDSA cycles, or sharing learnings with their colleagues, teachers felt rushed and overburdened. Measures for improvement are often referred to as *practical* measures in improvement science literature, an acknowledgement of the necessity to embed as much work as possible into the daily routine of practitioners. Yet today's public-school teachers are tasked with more work and less non-student time with which to attend to it than ever before.

In light of this fact, schools implementing improvement science principles must purposely schedule time for teachers to work explicitly on improvement efforts, separate and apart from the usual educator duties. This might take the form of paid professional development *outside* the contractual school day. Even better, districts could renegotiate teacher contracts to lengthen the school day to include more paid, non-student time in which instructional improvement efforts could be conducted. Alternately, districts could hire more teachers, enabling schools to accommodate the same number of students but giving teachers more non-student time within the current school day.

Given the difficulty of increasing public school funding in most parts of the country, every conceivable effort should be undertaken to embed the improvement principles into the existing work structure. As discussed earlier, current literature documents the increased effectiveness of professional development that is integrated in such a way (Desimone & Pak, 2017; Desimone et al., 2002).

Department chairs should be supported in regularly, and substantively attending to PDSA cycles during department meetings. Schools must decide if they have the time and resources to properly attend to an instructional improvement science effort, given the documented difficulty in measuring the drivers of classroom instruction amenable to changes in teacher work processes and the difficulty teachers exhibited in creating instructional process measures. At this junction, perhaps the data centric improvement science principles are best applied to aspects of education such as teacher feedback, or grading policies, areas more conducive to easily quantifiable process measures.

Systems thinking can focus an instructional improvement effort. The fact that teachers expressed value in systems thinking demonstrates the potential that exists for an improvement science implementation to focus a school's improvement effort on key, high leverage work processes. Teacher support is essential to a successful instructional change effort, given classroom teachers are the front-line practitioners in every school. This study affirmed the literature supporting the effectiveness of including teachers in the creation and use of a driver diagram. The power of systems thinking to keep a large number of educators focused on the same problem of practice cannot be overstated, given the frequently solitary nature of teaching.

Systems thinking, and the driver diagram in particular, constitutes a paradigmatic shift in the approach to instructional improvement. Too often, instructional professional development entails an information dump of best practices. This one-size-fits-all approach assumes that tight fidelity to the new instructional package will automatically result in higher student achievement. Learning is a highly contextual process however, and without systems thinking most schools stay oblivious to the local context that is undermining their improvement efforts. In fact, the existing literature frequently report professional development is most effective when it is grounded in the content and practices teachers are addressing in their daily classroom lessons (Desimone & Pak, 2017; Desimone et al., 2002). Systems thinking guides professional development to explicitly focus on these work processes and the context in which they occur.

Future improvement science professional development efforts would do well not to rush the initial work of creating an aim statement and a well-researched driver diagram, incorporating teacher feedback throughout the process. Focusing network efforts on the highest leverage work processes will result in the most efficient use of scarce teacher time.

Process measures are necessary for objective decision making. Another implication of this study with which future improvement efforts must attend to is the necessity of regular, quantified process measures to inform PDSA cycles. Quality instruction is difficult to define. One-size-fits-all solutions have eluded educators because of the strong influence context has on learning. Educators currently rely almost exclusively on anecdotal experience and student achievement data to inform their instructional choices. However, experience is subjective, and outcome data does not provide actionable information teachers can use to improve specific work processes. As a result, most schools repeatedly adopt and abandon (Rohanna, 2017) instructional change efforts because they lack objective, quantified data about the work processes they attempted to put into place; hence, the necessity of regular, quantified process measures outlined in the improvement science principles.

The network observed in this study most definitely engaged in disciplined inquiry, yet process measures were not created, and the PDSA cycles were not informed by the primary driver data collected. The PDSA data gathered by the teachers was not measured and quantified as defined by the improvement science principles. This hindered the ability of the network to collectively learn which work processes were and were not working, as well as to identify variation among its participating teachers. This led to a sight familiar to anyone who has observed a group of teachers discussing instruction – highly subjective, anecdotal story-telling provided as evidence that a particular work process was a success. The implication being, if you do what I do in my classroom with my students, it will work for you in your classroom with your students.

Network improvement communities take years to build properly, and in this study the facilitators chose to hold off on PDSA cycle process measures in order to focus on network norms for collaboration and a first attempt at a primary driver measure. The teachers in this study demonstrated difficulty in conceptualizing instructional process measures for themselves. The facilitators struggled to create a primary driver measure that was sensitive to work process changes. The experience of this network suggests that creating primary driver and process measures for the drivers of classroom instruction that are addressable by teacher work practices presents a difficult challenge.

All of this was taken into consideration when the network observed in this study made a deliberate choice to limit the role of quantified process measures. Future implementations may also make this choice, but without a transition to the regular use of primary driver and process measures to inform disciplined inquiry, systemic improvement efforts will be hindered.

Best practices for network learning and teachers desire for autonomy can create tension. The results of this study indicate that future instructional improvement science implementations will have to deal with the persistent tension between teachers' desire for autonomy and practices that support the disciplined inquiry and network learning principles. A common problem of practice, focus driver, and change idea allow a network to create common primary driver and process measures that are used to inform disciplined inquiry and accelerate network learning. This is the theory of change that improvement science is built upon (Bryk et al., 2015; Langley, 2009).

These instruments are what separate improvement science from other professional development methodologies. The specific, quantitative nature of these instruments, as defined in the improvement science literature, are what allow for the objective analysis of effective practices and the identification of variation between teachers.

However, when interviewed, teachers consistently described their ability to change drivers of focus and PDSA change ideas as highly valued. As described in the literature review, the credibility teachers had in data was influenced by their proximity to the design of the data instrument (Farrell and March, 2016b). Teacher perception of credibility can even be affected by the formatting of the instrument. Therefore, the less input teachers have in the creation of primary driver and process measures, the less likely they are to use the results. It is necessary for data to be viewed as credible in order for teachers to be open to using it to change their practice.

This idea is consistent with the improvement science principle of keeping the network problem-specific and user-centered, and contributed to teacher buy-in. In fact, several of those interviewed for this study shared previous experiences where a commitment to a common driver and change idea was resisted by teachers and led to an ineffective instructional improvement effort.

Future improvement science implementations will need to balance the teachers' need to adjust for context by exercising autonomy and the network's need for common driver, change idea, and primary driver and process measures in order to objectively inform PDSA cycles. This is no easy feat.

Topics for Future Research

Time and bandwidth concerns. Time and bandwidth concerns should be investigated to better understand the intensification of teacher. It was unsurprising that lack of time and bandwidth were frequently mentioned as hindering the improvement science implementation undertaken by the network in this study. What was surprising however, was the lack of literature addressing teacher workload and the intensification of the profession. Organizational learning cannot happen without dedicated time and attention, and it remains an unanswered question as to how much non-student time is currently available to public school teachers. The Organisation for Economic Co-operation and Development (OECD) *Education at a Glance* report attempted to document the state of education in countries around the world. According to Indicator D4 "How much time do teachers spend teaching?" from the 2018 report, American educators teach more hours and have less non-student time than the OECD average. The United States' ranking stands in stark contrast with nations such as Japan and the Nordic countries, which are often recognized for their academic success. Educators from these countries teach fewer hours and have more non-student time than the average OECD country (OECD, 2018).

Therefore, American educational research would benefit from a comprehensive study into teacher workload. Possible questions include: What is the length of the average teacher's contractual workday, and how much of that day is required to be direct instruction? How much time does the average teacher spend directly preparing for those lessons or grading? How much time does the average teacher spend on tasks unrelated to instruction? How much non-student contractual time does the average teacher have? Quantitative survey analysis would enable policy makers to compare across school districts and grade levels and look for correlations, such as between teacher workload and teacher retention or student outcomes. Learning is highly contextual however, so researchers would have to carefully account for other variables such as students' socio-economic status.

However, a qualitative approach such as journaling, would also yield extraordinary insight into the myriad and disparate tasks required of teachers. Excerpts of the interviews conducted for this study themselves revealed teachers working on a multitude of tasks that did not concern direct instruction or grading, work that was frequently unpaid and done outside of the workday. The answers to the above questions are vital as the field of education continues to move towards data-driven decision making. As teachers comprise the vast majority of the educational workforce, the work of instructional improvement will inevitably fall to them.

Primary driver and Process measure informed instructional PDSA cycles. Research still needs to be done on an instructional improvement effort using quantified primary driver and process measures to inform PDSA cycles. The network in this study created an instructional primary driver measurement; however, that information was not used by participants to guide the improvement effort. Additionally, no process measures were created for the new instructional work processes undertaken. As a result, PDSA cycles were not informed by measures for improvement as outlined by improvement science principles. At the outset of this dissertation I had hoped to report on the practicality of using primary driver and process measures to support classroom instruction. Despite the observed network's strong adherence to disciplined inquiry, the data that were collected did not allow for a quantitative evaluation of primary driver movement and the efficacy of new work processes. Therefore, future research is needed to examine the practicality of creating *instructional* process measures and the ability of these measures to support PDSA cycles around classroom instruction.

Given the difficulty in finding schools undertaking this specific type of work, this question may best be answered through action research. Then, the researcher could be sure of the creation and use of instructional primary driver and process measures. It would be best to coordinate with school administration to ensure that the appropriate time would be made available and that a critical mass of teachers would participate. Improvement science can be quite technical, and improved instructional work processes depend on solid pedagogical knowledge. Therefore, it would behoove potential researchers to team up with improvement science and pedagogical experts. The network observed for this study provides an excellent model.

Connections to Prior Research

Unmeasured disciplined inquiry most closely matched lesson study practices. The network facilitators in this study made a deliberate decision to organize PDSA cycles around qualitative information such as observations and student work. As a result, teachers were able to speak at length during their interviews about how this information provided valuable insight into student thinking. The desire of teachers for information that makes student learning visible most closely resembles the qualities of lesson study discussed in the literature review (Akiba et al., 2019; Dudley, 2013; Lewis et al., 2012). In the practice of lesson study, teachers observe a lesson in action. Afterward, they debrief with their peers, with a specific focus on improving the lesson to make learning even more visible.

The teachers in this study reported great appreciation for the opportunity to talk with their peers about the similar work processes with which they were experimenting. Lesson study asks teachers to build and spread high impact instructional routines (Akiba et al., 2019; Dudley, 2013; Lewis et al., 2012). This is very similar to the systems and disciplined inquiry principles of improvement science. Absent primary driver measures and process measures, the activities of the network observed in this dissertation had much in common with lesson study methodology.

Given the difficulty in creating primary driver and process measures, lesson study may be a more appropriate professional development methodology for instructional improvement efforts. The information used in lesson study frequently consists of student work and qualitative data gathered through observation and role playing (Akiba et al., 2019; Dudley, 2013; Lewis et al., 2012). This is precisely the information that was used by the network in this study or requested by the interviewed teachers. Teachers reported this information as helpful in making student learning visible. In addition, this information contributed to productive conversation about changes in teacher practices that would increase learning.

Use of survey data. Previous instructional improvement science reporting has been limited to examples of survey data as the primary quantified measurement (Gelderblom et al., 2016; Hannan et al., 2015; Jackson et al., 2016; Yeager et al., 2013). The sparse literature on instructional improvement science implementations document primary driver and process measures consisting entirely of student surveys. Student feedback can be informative, but it is only one measure of the effectiveness of a given work process. The existing literature frequently acknowledges the difficulty of creating practical primary driver and process measures for instruction, but is often lacking in examples beyond student surveys (Bryk et al., 2015; Rohanna, 2018). The network in this study faced many of the same challenges, including how to create driver measures that were sensitive to change, were predictive of changes in the outcome measure, and were not intrusive or time consuming.

Limitations

As with all research, there are limitations to the scope and capacity of what can be addressed. It is possible that the teachers who agreed to be interviewed held strong opinions about the efficacy of improvement science principles that were not representative of the group. I tried to mitigate this by requesting interviews with teachers from each school. However, there was one school from which I was unable to secure a teacher interview. Additionally, given that I altered my teacher interview protocol in March 2019, ideally I would have liked to re-interview all seven of the teachers whom I interviewed in December 2018 instead of only four. Despite this, the consistency of the interview responses across teachers and schools supported the conclusion that I had achieved saturation, no new themes or information were being observed in my data.

Finally, and most importantly, I urge caution in any attempt to generalize these findings to a broader context. My research reflects the experiences of a particular group of educators being led through a particular improvement science exercise with a specific focus on instructional improvement. While my research questions could be applied to other improvement efforts, one might expect that the resulting answers would be site-specific. It has been noted here that previous research has documented the successful implementation of primary driver and process measures in other applications of education, such as effective teacher feedback and the sequencing of community college math courses. Also, the ability of the teachers participating in the network I observed to report on primary driver and process measures was, in part, constrained by decisions the facilitators made about how much time and effort would be expended on their creation and use, and their decision not to share with teachers the results of their measurements until the end of the year. In my reporting, I have tried to share a factual accounting of the lived experience of teachers engaging in an improvement science effort focused on instruction, something that is lacking in the current literature.

Reflection on the Research Process

In part, due to my personal background in engineering, I am inclined to look for quantifiable evidence when evaluating instructional efficacy. However, my time in the classroom awakened me to the reality that teaching is an art as much as a science because of the highly contextual nature of learning. When I came across the improvement science principles, I was excited to find a blend of these two philosophies. A practical, objective methodology that took into account local context. Learning by doing, practical measurement, disciplined inquiry.

I set out to discover if the theory of improvement science could survive the reality of the classroom when applied to an instructional improvement effort. The network facilitators I observed went to great lengths to apply the improvement science principles. The work was problem-specific and user-centered. They worked hard to see the system, creating and repeating, referencing their driver diagram. The participating teachers engaged in ongoing disciplined inquiry through the PDSA cycles. The network encouraged discourse between math teachers across five schools, accelerating the network's learning. Finally, the facilitators created a primary driver measure, with the aim of using measures for improvement to guide their improvement efforts.

As it turned out, the drivers of quality classroom instruction are difficult to measure, and the benefits the network experienced by allowing teacher autonomy conflicted with the improvement science principles for network learning guided by quantified process measures. It is still early in the life cycle of the network observed in this study, and I am confident they will continue to refine and expand their measures for improvement. I remain optimistic in the potential for improvement science principles to support teachers in their efforts to improve classroom instruction. However, due to the difficulty in creating accurate instructional primary driver and process measures and the realities of teacher workload intensification, educators must continue to search out the most efficient methodologies.

Appendix A – Network PDSA Log

PDSA Log:

PDSA Form Template: Name__________________________

Date	One- or two-week cycle:
Change Idea	**Describe:** **What is the change:** **Aligns to:** ______________Primary Driver and _________________________Secondary Driver
Plan	

Predictions: What improvement do we think will happen?	**Questions:** What do we want to learn from this cycle?	**Data:** What information will we collect to answer our questions and test our prediction?	**Results and Next Steps:** Did you meet your prediction? What were the results? What did we learn? What will we do next? *(completed after implementation)*

Appendix B – Sample Network Agendas

October 18, 2018 Network Agenda

How do we work together?
Be Open to New Ideas and Embrace "Learning by Doing" – Don't let perfection get in the way of learning and trying new practices.

Seek to Understand –Share your perspective. Consider other perspectives. Ask questions rather than just offering suggestions. Understand that conflict is normal. Push without pushing buttons.

Respect Colleagues – Be on time. One conversation at a time. Monitor air time. Respect confidentiality. Have fun.

Oct. 18, 2018
8:00 a.m. – 3:00 p.m.

Why are we here?
Shared Problem of Practice: Our current practices are not aligned with students' learning needs today. Many students are failing math.

Network Aim: Our practices actively engage students in math and meet their variety of learning needs. We know we're improving when more students are learning and passing math classes with a C or better.

#	Time	Topic	Outcomes	Speaker/ Facilitator
1	8:00 (30)	▪ Breakfast	▪ Learn about others in the network and build relationships	NA
2	8:30 (15)	▪ Welcome and Purpose for the Day	▪ Shared understanding of network's purpose and meeting objectives	Sarah
3	8:45 (120)	▪ Eliciting, and Leveraging Student Thinking (Provide break here)	▪ Leveraging student mathematical thinking to build on mathematical understanding	Tom
4	10:45 (75)	▪ PDSA Inquiry & Dialogue	▪ Learn about, ask questions, and support new practices of teachers and from other schools	NA
5	12:00 (15)	▪ Measuring Our Network's Progress (Part 1)	▪ Provide network tool for collecting data regarding the consistent use of practices aligned to drivers	Sarah
6	12:15 (45)	▪ Lunch	▪ Lunch	NA
7	12:50 (25)	▪ Measuring Our Network's Progress (Part 2)	▪ Practice using new tool to collect data about consistent use of practices	NA
8	1:15 (45)	▪ Planning Our Next PDSA: Work Time	▪ Completed PDSA form for next PDSA Cycle	NA
9	2:00 (30)	▪ Share Out Change Idea + Map to Driver Diagram	▪ Learn about others' ideas, and provide feedback	NA
10	2:30 (10)	▪ Network Survey	▪ Collect data about the state of the network, gather feedback	NA
11	2:40 (20)	▪ Reflection on the Day and Next Steps	▪ Concrete next steps	Sarah

Next Meeting Nov. 13th: Continued math instructional professional development and sharing our PDSA cycles.

May16th 2018 Network Agenda

How do we work together?
Be Open to New Ideas and Embrace "Learning by Doing" – Don't let perfection get in the way of learning and trying new practices.

Seek to Understand –Share your perspective. Consider other perspectives. Ask questions rather than just offering suggestions. Understand that conflict is normal. Push without pushing buttons.

Respect Colleagues – Be on time. One conversation at a time. Monitor air time. Respect confidentiality. Have fun.

May 16, 2019
8:00 a.m. – 3:00 p.m.

Why are we here?
Shared Problem of Practice: Our current practices are not aligned with students' learning needs today. Many students are failing math.

Network Aim: Our practices actively engage students in math and meet their variety of learning needs. We know we're improving when more students are learning and passing math classes with a C or better.

#	Time	Topic	Outcomes	Speaker/ Facilitator
1	8:00 (30)	■ Breakfast	■ Learn about others in the network and build relationships	NA
2	8:30 (30)	■ Welcome and Purpose for the Day	■ Shared understanding of network's purpose and meeting objectives	Sarah
3	9:00 (45)	■ Our Shared Math Experience	■ Learn more about applying choral counting to higher level math classes.	Tom
4	9:45 (10)	■ Break	■ NA	NA
5	9:55 (65)	■ PDSA Showcase	■ Learn from colleagues and celebrate some of their change ideas!	NA
6	11:00 (45)	■ Primary Driver and Aim Data	■ Celebrate and discuss our continued progress. How did we do as a network?	NA
7	11:45 (45)	■ Lunch	■ Nourishment and get to know colleagues	NA
8	12:30 (60)	■ Building Your Team Vision	■ University shares their vision for network. School teams discuss their own purpose and vision for this work.	NA
9	1:30 (60)	■ Team Planning for Next Year: How will you meet your vision?	■ Teams identify steps, resources, and support needed to meet their own visions.	NA
10	2:30 (15)	■ Survey	■ Provide feedback regarding network activities and supports	NA
11	2:45 (15)	■ Cake and stipends!	■ Show our appreciation for our teachers!	NA

Have a great summer!

Appendix C – Network Driver Measures

March 2018 Secondary Driver Measure for Improvement

Write in your Focus Secondary Driver: **Write in your Focus Indicator:**

Definition of Rubric Terms

Active participation: Observable evidence of participation, such as raising hand, body language demonstrating that they're listening to teacher or others, on task, writing, engaging in dialogue in a small group.

Indicator	0	1	2	3	4
Engagement of All Students	No or almost no time spent on a whole or small group activity or task **designed to actively engage more students.**	The activity only engages the same students. Active participation is demonstrated by those who usually participate, not those who do not usually participate.	The activity begins to engage more students. Active participation is demonstrated by those who usually participate, plus 1 to 3 who don't usually participate.	The activity engages more students. Active participation is demonstrated by those who usually participate, plus more than 3 who don't usually participate, but not all students	The activity engages all students. Active participation is demonstrated by those who usually participate, plus all of those who don't usually participate.
Indicator	**0**	**1**	**2**	**3**	**4**
Opportunities to Explain Thinking	Teacher primarily talks during class with very little or no opportunity for students to explain or demonstrate their thinking to the teacher.	Few students (0% to 25%) explain or demonstrate their thinking to the teacher (either verbally or in writing).	Some students (26% to 50%) explain or demonstrate their thinking to the teacher (either verbally or in writing).	Most students (51% to 75%) explain or demonstrate their thinking to the teacher (either verbally or in writing).	All or almost all students (76% to 100%) explain or demonstrate their thinking to the teacher (either verbally or in writing).
Indicator	**0**	**1**	**2**	**3**	**4**
Student Questioning	Teacher asks all questions. Student do not ask each other questions.	Teacher asks most of the questions. A little student-to-student questioning occurs with teacher prompting. None or almost no student-to-student questioning occurs without teacher prompting.	Teacher and students ask questions. Some student-to-student questioning occurs with teacher prompting. A little student-to-student questioning without teacher prompting.	Teacher and students ask questions. A lot of student-to-student questioning occurs with teacher prompting. Some student-to-student questioning without teacher prompting.	Teacher and students ask questions. A lot of student-to-student questioning occurs with teacher prompting. A lot of student-to-student questioning without teacher prompting.
Notes					
What is the evidence for my ratings? How did I collect evidence? What did I learn about my students thinking?					

2018-2019 Primary Driver Measure for Improvement

What Primary Driver are you focusing on this year?

- ◯ **Primary Driver 1:** Engage students in problem solving that promotes conceptual understanding and development of procedural skills.
- ◯ **Primary Driver 2:** Provide students with opportunities to communicate their ideas by implementing cooperative learning structures.
- ◯ **Primary Driver 3:** Attend to student needs by providing multiple opportunities for students to demonstrate their learning.

Primary Driver 1

The goal is to implement one problem-solving task or activity a week. We understand that this will take time and you may not be able to do any tasks some weeks.

Please complete the following items truthfully. Your answers are anonymous. There is no right or wrong answer. We're all learning together.

How many problem-solving tasks or activities did you conduct with your students this week?

(Please do not count the same activity more than once if you did with different class periods. For example, if you did choral counting on Tuesday with all class periods, you would count as 1. If you did choral counting on Tuesday and Which One Doesn't Belong on Thursday, you would count as 2.)

[▾]

Please select the problem-solving task(s) or activities that you implemented this week? Select all that apply.

- ☐ Which One Doesn't Belong
- ☐ Reversals
- ☐ Choral Counting
- ☐ Other (please describe) [ ]
- ☐ Notice and Wonder
- ☐ Other (please describe) [ ]
- ☐ True False Number Sentences
- ☐ Other (please describe) [ ]

2018-2019 Primary Driver Measure for Improvement (continued)

Primary Driver 2

The goal is to provide students with multiple opportunities each week to communicate their ideas to others through a cooperative learning activity. We understand that this will take time and you may not be able to do any activity some weeks.

Please complete the following items truthfully. Your answers are anonymous. There is no right or wrong answer. We're all learning together.

How many collaborative activities did you conduct with your students this week, with the intent of them communicating their ideas to each other?

(Please do not count the same activity more than once if you did with different class periods. For example, if you did Pair Checks on Tuesday with all class periods, you would count as 1. If you did Pair Checks on Tuesday and another type of group activity on Thursday, you would count as 2.)

[ ]

Please describe the collaborative activities that you implemented this week? (Describe up to 3 activities.)

	Purpose of Activity (Select all that apply)			Purpose of Activity (Select all that apply)
	Facilitate mathematical discourse and vocabulary	Students learn from each other	Students articulate reasoning	Other
Describe Group Activity [ ]	☐	☐	☐	[ ]
Describe Group Activity [ ]	☐	☐	☐	[ ]
Describe Group Activity [ ]	☐	☐	☐	[ ]

2018-2019 Primary Driver Measure for Improvement (continued)

Primary Driver 3

The goal is to implement one formative assessment task or activity each week where students can demonstrate their learning. We understand that this will take time and you may not be able to do any tasks/activities some weeks.

Please complete the following items truthfully. Your answers are anonymous. There is no right or wrong answer. We're all learning together.

How many formative assessment tasks or activities did you conduct with your students this week where students could demonstrate their learning?

(Please do not count the same activity more than once if you did with different class periods. For example, if you did exit tickets on Tuesday with all class periods, you would count as 1. If you did exit tickets on Tuesday and Which One Doesn't Belong on Thursday, you would count as 2.)

[ ]

Please select the formative assessment task(s) or activities that you implemented this week, and answer corresponding questions. Select all that apply, but only answer for the ones that you did this week.

	Was it a formal or informal assessment?		How informative was assessment re: students' thinking?		
	Formal	Informal	1 - Assessment did not elicit student reasoning. (I know what they got right and wrong, but not why.)	2 - Assessment did elicit some student reasoning but I was still unclear of how to build upon the students' thinking.	3 - Assessment elicited student reasoning and it was clear how to build upon students' thinking
Number sense routines (WODB, choral counting, notice and wonder, etc.)	○	○	○	○	○
Student work from other problem-solving tasks/activities	○	○	○	○	○
Exit tickets	○	○	○	○	○
Group closure activitiy	○	○	○	○	○
Written reflection activity	○	○	○	○	○
Other (please describe) [ ]	○	○	○	○	○

	Was it a formal or informal assessment?		How informative was assessment re: students' thinking?		
	Formal	Informal	1 - Assessment did not elicit student reasoning. (I know what they got right and wrong, but not why.)	2 - Assessment did elicit some student reasoning but I was still unclear of how to build upon the students' thinking.	3 - Assessment elicited student reasoning and it was clear how to build upon students' thinking
Other (please describe)	○	○	○	○	○
Other (please describe)	○	○	○	○	○

Appendix D – Interview Protocols

Teacher Interview Protocol #1

1. What kind of information would help you understand if students had mastered a concept?
   a. What about this information allows you to conclude that students did or did not master a concept?
2. Could you describe some of the information you use when planning your instruction? (RQ1)
3. How comfortable are you with analyzing data?
   . Could you describe a time that you used data?
4. What kind of information would you want to inform any changes in your instruction if you determined you needed to reteach a lesson?

. How would you use this data to inform any changes in your instruction?
   a. How would you know that any difference in student outcomes was the result of your change?
5. Could you please describe your PDSA change idea?
   . What prompted you to choose this change idea?
6. What data did you select to test your change idea?
   . Why did you feel this data was relevant to your change idea?
   a. What did you hope to learn from this data?
   b. What insight into how your students make sense of the material do you hope to gain?
   c. How will this data inform your next steps regarding instruction?
   d. QUESTION about how this data was discussed in any collaboration?
   e. Is there any additional data you would have liked to inform your instruction?
7. Were the students informally assessed on this material?
   . Please describe your informal assessment process.
   a. How did you use the results of this assessment to inform your instruction?
8. Were students formally tested on this material?
   . Could you briefly describe the assessment?
   a. How did you use the results of this assessment to inform your instruction, if at all?
      i. How valid would it be to use the results of this assessment to evaluate the effectiveness of your instruction?

Teacher Interview Protocol #2:

1. Describe your involvement in creating the Driver Diagram.
   i. How, if at all, did discussion around the Partner School Network driver diagram influence your understanding of how to improve instruction?
2. How, if at all, did the Partner School Network driver diagram influence the selection of your change ideas?
3. What was your understanding of the purpose of the Primary Measure weekly survey (show them the survey)?
   . How, if at all, would you have used the results of that data to make decisions about your instruction?
4. Describe your understanding of the secondary driver measure (March rubric) that was created? (show them the rubic)
   .To your understanding, why didn't teachers use it?
   i.How, if at all, would you have used the results of that data to make decisions about your instruction?
   ii.How much of a factor does time/cognitive load play in your ability to use information (such as PDSA cycles) to improve your instruction? (in the moment? Over time?)
5. What factors led to you not completing all the PDSA logs? (If applicable).
   a. How much of a factor does time/cognitive load play in your ability to use information (such as PDSA cycles) to improve your instruction? (in the moment? Over time?)
6. In what way, if at all, did the PDSA process influence your understanding of how to improve instruction? (Reflection on your practice? Focus on student work? The idea of evaluating small changes for their impact on a primary driver? Communication with colleagues?)
7. What factors do you take into consideration when you are evaluating the effectiveness of your instruction?
8. How much of a factor does time/cognitive load play in your ability to use information (such as PDSA cycles) to improve your instruction? (in the moment? Over time?)
9. Could you describe a change idea that you have worked on/are working on?
   . Why did you choose this change idea?
10. What information did you decide would be helpful in determining if your change idea was successful?
    . How will this information help you determine if your change idea was successful?
11. Could you describe the process of how you gathered that information?
    . How did you collect and record the data?
12. How did you decide that this information could credibly help you to determine the effectiveness of your change idea?
    . ASK FOR MORE DETAIL HERE
13. Please describe how you used this PDSA cycle to inform any instructional changes?
    . Could you please describe an example?
14. In what way, if any, has your involvement in this UCLA led improvement science professional development shaped your use of information to inform changes in instruction?
15. Thank you for taking the time to participate in my research. Are there any additional insights you would like to share with me before we conclude?

Network Facilitator Interview Protocol:

1. What was the purpose of the Network professional development?
2. What is your understanding of the role that measuring primary drivers plays in an instructional improvement effort?
3. What did you intend for teachers to understand about measuring primary drivers?
   a. How did you explain what information teachers should gather to measure their primary drivers?
   b. How much time/focus was directed on the idea of using information to measure primary drivers?
4. How did you instruct teachers to use the information they gathered?
   . How much time/focus was directed to using the information they collected to measure their primary driver?
      a. What were teachers asked to do with the information they collected?
      b. What were teachers asked to do after a single PDSA cycle?
5. How did expectations for collecting information to measure their primary driver change throughout the course of the PD, if at all
   . What led to these changes?
6. Which sites did you visit and how often?
   . In your observations, how much time were teachers given to discuss the PDSA logs?
   a. What activities did you observe around the PDSA logs?

   b. Who led these activities?
   c. To your knowledge, what was their understanding of how instruction could be improved by connecting the information collected during a PDSA cycle to a primary driver?
   d. What were teachers asked to do instructionally after a discussion on a specific PDSA log?
7. How could future improvement science PD best support teachers in measuring primary drivers to improve instruction?

8. What barriers did you encounter in developing teacher understanding of the goal-driver-measure-instruction relationship?

Appendix E –Field Notes Instrument

Network Meeting Field Note

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Date:	Network Location																	
2																			
3	11/15/14 10:21 AM	Evidence	PROBLEM-SPECIFIC/USER-CENTERED	teacher work practice		SYSTEMS THINKING	Driver Diagram		Disciplined Inquiry	DISCIPLINED INQUIRY	Primary Driver Measure	Process Measure	Unquantified Data (student work/observations)						
4	Time Stam	Evidence	1A	1A	1A	1A	1A	1A	1B	1B	1B3	1	1B	1B6	1C1	1C	1D	1	1
5																			
6																			
7																			
8																			
9																			
10																			
11																			
12																			
13																			
14																			
15																			
16																			
17																			
18																			
19																			
20																			
21																			
22																			

Appendix F – Survey Instrument

1. The network team incorporated improvement science principles into the network's work. Please rank the following types of improvement science activities based on their usefulness to your instructional improvement efforts since you started participating in the network, with 1 being the MOST useful and 6 being the LEAST useful.

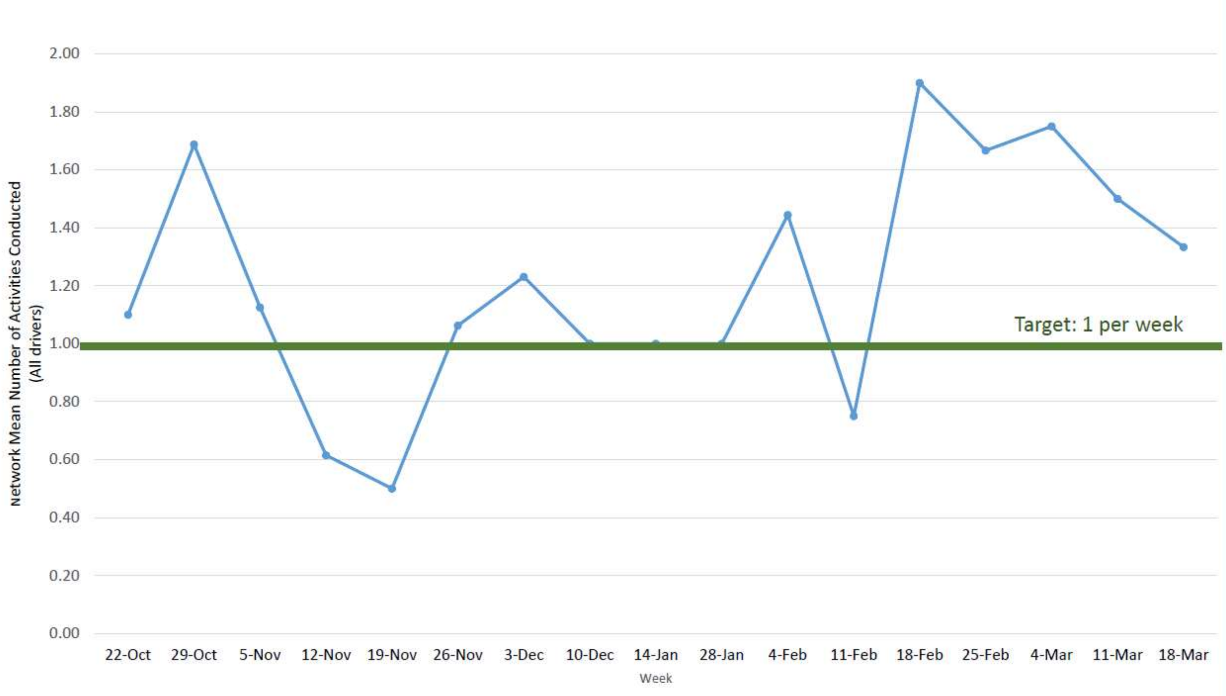
- Make the work problem-specific and user-centered. (Teachers were engaged in designing changes in instruction that align with the problems they experience in the classroom.)
- Our discussions on how the variation in instruction from teacher to teacher can impact student outcomes.
- See the system that produces the current outcome. (The driver diagrams we created helped us to clarify current problems and invite new thinking about possible solutions.
- Collecting data on the primary drivers so we can measure our progress as a network.
- Anchoring our intructional improvement in disciplined inquiry (Rapid PDSA cycles helped us to make small tests of change and learn fast.)
- Accelerate improvements through a network community. (Bringing together teachers from different schools to work on common instructional challenges enhanced our improvement efforts.)

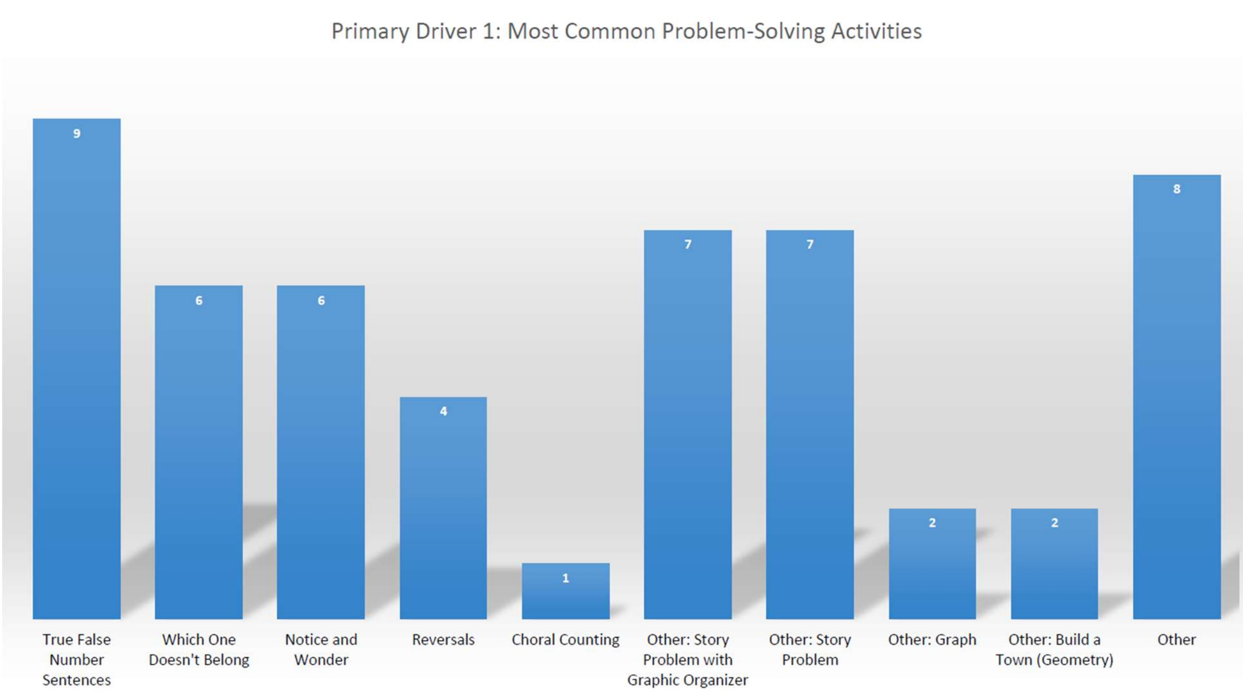
2. Briefly share why you chose the statement you did as the most useful.

3. Briefly share why you chose the statement you did as the least useful.

Appendix G – Driver Measurement Data

Number of primary driver activities (see Appendix E) conducted by date.





References

Akiba, M., Murata, A., Howard, C. C., & Wilkinson, B. (2019). Lesson study design features for supporting collaborative teacher learning. *Teaching and Teacher Education*, *77*, 352–365. https://doi.org/10.1016/j.tate.2018.10.012

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5. https://doi.org/10.1007/s11092-008-9068-5

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Colyvas, J. A. (2012). Performance metrics as formal structures and through the lens of social mechanisms: When do they work and how do they influence? *American Journal of Education*, *118*(2), 167–197. https://doi.org/10.1086/663270

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage Publications.

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*(4), 1.

Deming, W. E. (1994). *The new economics for industry, government, education*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Desimone, L. M., & Pak, K. (2017). Instructional coaching as high-quality professional development. *Theory into Practice*, *56*(1), 3–12. https://doi.org/10.1080/00405841.2016.1241947

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

Dudley, P. (2013). Teacher learning in lesson study: What interaction-level discourse analysis revealed about how teachers utilised imagination, tacit knowledge of teaching and fresh evidence of pupils learning, to develop practice knowledge and so enhance their pupils' learning. *Teaching and Teacher Education*, *34*, 107–121. https://doi.org/10.1016/j.tate.2013.04.006

Farrell, C. C., & Marsh, J. A. (2016a). Contributing conditions: A qualitative comparative analysis of teachers' instructional responses to data. *Teaching and Teacher Education*, *60*, 398–412. https://doi.org/10.1016/j.tate.2016.07.010

Farrell, C. C., & Marsh, J. A. (2016b). Metrics matter: How properties and perceptions of data shape teachers' instructional responses. *Educational Administration Quarterly*, *52*(3), 423–462. https://doi.org/10.1177/0013161X16638429

Gelderblom, G., Schildkamp, K., Pieters, J., & Ehren, M. (2016). Data-based decision making for instructional improvement in primary education. *International Journal of Educational Research*, *80*, 1–14. https://doi.org/10.1016/j.ijer.2016.07.004

Glover, T. A., Reddy, L. A., Kettler, R. J., Kurz, A., & Lekwa, A. J. (2016). Improving high-stakes decisions via formative assessment, professional development, and comprehensive educator evaluation: The school system improvement project. *Teachers College Record*, *118*(14), 1-26.

Gomez, K., Gomez, L. M., Rodela, K. C., Horton, E. S., Cunningham, J., & Ambrocio, R. (2015). Embedding language support in developmental mathematics lessons: Exploring the value of design as professional development for community college mathematics instructors. *Journal of Teacher Education, 66*(5), 450–465. https://doi.org/10.1177/0022487115602127

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using improvement science to better support beginning Teachers. *Journal of Teacher Education, 66*(5), 494-508. doi:10.1177/0022487115602126

Ho, J. E. (2016, January 1). *Cultures and contexts of data-based decision-making in schools*. UCLA. Retrieved from https://escholarship.org/uc/item/9nn9z1g0

Horn, I. S., Kane, B. D., & Wilson, J. (2015). Making sense of student performance data: Data use logics and mathematics teachers' learning opportunities. *American Educational Research Journal*, *52*(2), 208–242. https://doi.org/10.3102/0002831215573773

Jackson, K., Henrick, E., Cobb, P., Kochmanski, N., & Nieman, H. (2016). Practical measures to improve the quality of small-group and whole-class discussion [White Paper]. Retrieved from http:www.education.uw.edu/pmr/files/2016/09/White-Paper.pdf

Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record, 114*(11), [110304].

Jimerson, J. B., & Wayman, J. C. (2015). Professional learning for using data: Examining teacher needs and supports. *Teachers College Record, 117*(4), 1-36.

Kraft, M. A., & Blazar, D. (2017). Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Educational Policy*, *31*(7), 1033–1068. https://doi.org/10.1177/0895904816631099

Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: a practical approach to enhancing organizational performance*. San Francisco, CA: Jossey-Bass.

Lewis, C. (2015). What is improvement science? Do we need it in education? *Educational Researcher*, *44*(1), 54–61. https://doi.org/10.3102/0013189X15570388

Lewis, C. C., Perry, R. R., Friedkin, S., & Roth, J. R. (2012). Improving teaching does improve teachers: Evidence from lesson study. *Journal of Teacher Education*, *63*(5), 368–375. https://doi.org/10.1177/0022487112446633

Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*(2), 143–166. https://doi.org/10.1086/663271

Little, J. W., Gearhart, M., Curry, M., & Kafka, J. (2003). Looking at student work for teacher learning, teacher community, and school reform. *Phi Delta Kappan*, *85*(3), 184–192. https://doi.org/10.1177/003172170308500305

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, *47*(2), 71–85. https://doi.org/10.1080/00461520.2012.667064

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.

OECD. (2018), Education at a glance 2018: OECD indicators, OECD publishing, Paris, France. http://dx.doi.org/10.1787/eag-2018-en

Oláh, L. N., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education*, *85*(2), 226–245. https://doi.org/10.1080/01619561003688688

Rohanna, K. (2017). Breaking the "adopt, attack, abandon" cycle: A case for improvement science in K-12 education. *New Directions for Evaluation, 2017*(153), 65-77. doi:10.1002/ev.20233

Rohanna, K. (2018). *Preparing schools to successfully participate in networked improvement communities: A case study of year 1 of a math instructional network* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (10936491)

Russell, J., Stein, M. K., Correnti, R., Bill, V., Booker, L., & Schwartz, N. (2017). Tennessee scales up improved math instruction through coaching. *State Education Standard*, *17*(2), 23–27.

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). New York, NY: Teachers College Press.

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, *118*(2), 113-141.

Stecher, Brian M., Deborah J. Holtzman, Michael S. Garet, Laura S. Hamilton, John Engberg, Elizabeth D. Steiner, Abby Robyn, Matthew D. Baird, Italo A. Gutierrez, Evan D. Peet, Iliana Brodziak de los Reyes, Kaitlin Fronberg, Gabriel Weinberger, Gerald Paul Hunter, & Jay Chambers (2018). *Improving Teaching Effectiveness: Final Report: The Intensive Partnerships for Effective Teaching Through 2015–2016*. Santa Monica, CA: RAND Corporation, 2018. https://www.rand.org/pubs/research_reports/RR2242.html. Also available in print form.

Thessin, R. A. (2010). *A qualitative study of teachers' work in professional learning communities in a small urban district* (Doctoral dissertation). Retrieved from https://search.proquest.com/docview/857923368/abstract/41E0741771244A32PQ/1

Wayman, J. C., Wilkerson, S. B., Cho, V., Mandinach, E. B., & Supovitz, J. A. (2016). *Guide to using the Teacher Data Use Survey* (REL 2017–166). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from http://ies.ed.gov/ncee/edlabs.

Yeager, D. S., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Carnegie Foundation for the Advancement of Teaching. Available from https://www.carnegiefoundation.org/resources/publications/practical-measurement/