Continuous Improvement in Action: Educators' Data Use for School Improvement

Marisa Cannata, *Vanderbilt University*

Christopher Redding, *University of Florida*

Mollie Rubin, *Vanderbilt University*

Marisa Cannata (Corresponding author)
Vanderbilt University
PMB #414
230 Appleton Place
Nashville, TN 37203
Marisa.cannata@vanderbilt.edu
651-322-1746

Christopher Redding
School of Human Development and Organizational Studies in Education
College of Education, University of Florida
PO Box 117049
Gainesville, FL 32611-7046
c.redding@coe.ufl.edu
352-273-4472

Mollie Rubin
Vanderbilt University
PMB #414
230 Appleton Place
Nashville, TN 37203
mollie.rubin@vanderbilt.edu
651-343-3734

Acknowledgment: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E100030. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Continuous Improvement in Action: Educators' Data Use for School Improvement

Abstract

Continuous improvement approaches to school reform involve the use of data to understand a problem and test and refine potential solutions. In this paper, we describe how educators come to understand and use data in a continuous improvement approach to school improvement within a large, urban district. We find evidence that educators are likely to draw on a mix of evidence as well as evidence substitutes when refining the innovation at their schools. While teams considered outcome data, they gravitated towards perceptual evidence to gauge the level of teacher buy-in and make modifications that would better meet teachers' needs. Further, we find that the district culture of accountability shaped their use of data for improvement.

Keywords: school reform, continuous improvement, data use

Continuous Improvement in Action: Educators' Data Use for School Improvement

With the development of longitudinal data systems within a climate of high-stakes accountability, more and more school systems have been moving toward the use of data to inform decisions. A substantial amount of recent research has focused on how educators use data and data-use initiatives (Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006; Marsh, 2012; Spillane, 2012). Teachers report that they alter their instruction in response to analyzing student data (Stecher et al., 2008) and data-use interventions influence organizational culture (Marsh, 2012). Yet educators also encounter substantial obstacles in using data, including the lack of capacity to select relevant evidence and interpret results (Marsh, Pane, & Hamilton, 2006), lack of time (Ingram, Seashore Louis, & Schroeder, 2004), and insufficient access to timely data (Lachat & Smith, 2005; Lynch, Smith, Provost, & Madden, 2016).

Organizational and institutional contexts shape how educators interpret and use data (Coburn, Honig, & Stein, 2009; Marsh, 2012; Moss, 2012). Educators not only use data instrumentally to make decisions, but also conceptually to shape how they think about an issue or symbolically to mobilize support (Coburn, Honig, & Stein, 2009; Penuel et al., 2016; C. H. Weiss, 1977). District and school personnel's use of data is also shaped by their position in the organization, involvement in prior reform efforts, and opportunities to collectively interpret the data (Coburn & Talbert, 2006; Honig & Coburn, 2008). Central offices shape data use in several ways, such as setting organizational routines and expectations around data use, providing necessary professional development for the use of data, and carving out time for educators to learn to make sense of and apply the evidence (Grissom et al., 2017; Honig & Venkateswaran, 2012; Park & Datnow, 2009; Spillane, 2012; Wohlstetter, Datnow, & Park, 2008).

The recognition that how districts organize themselves is consequential for data use has led to calls for new ways to incorporate evidence use into school improvement efforts. The most notable of these is improvement science, which involves attention to both systemic forces that shape inter-organizational activities and micro-processes that can be tested through continuous improvement (Bryk, Gomez, Grunow, & LeMahieu, 2015; Cohen-Vogel et al., 2015). Within improvement science, continuous improvement research is based on the idea that taking effective practices to scale requires an understanding of the conditions and contexts that enable program success (Cohen-Vogel et al., 2015). With a greater sensitivity to how local contextual factors shape the impact of educational interventions, continuous improvement research uses multiple forms of evidence to not only understand whether or not an intervention is achieving positive results but also to understand how local conditions are producing this outcome (Lewis, 2015).

The defining features of continuous improvement research are that it is ongoing, iterative, and "fully integrated into the daily work of individuals within a system" (Cohen-Vogel et al., 2015, p. 265). In this way, continuous improvement cycles are a type of organizational routine that attempts to embed the collection and interpretation of data into educators' work. This paper explores how educators in one large urban district responded to the introduction of continuous improvement cycles as an organizational routine and the ways in which the existing institutional context of the district shaped that response. Given the importance of the organizational and political contexts, we use the new institutionalism in education to understand how the institutional context of schools and districts influences responses to organizational change.

We begin by describing this framework of institutional theory and the role of organizational routines in explaining individual behavior inside institutions. We organize prior research on educator data use within this framework. We then describe the context of the study

(including the focus of the reform efforts and the district in which the work was situated), the data and methods, followed by the main findings in two sections. In the first section, we demonstrate that educators initially resisted efforts to discuss data due to concerns about it being misused for accountability, but became more open as they focused on gathering feedback for improvement. Then, in the second section, we describe the types of data the educators cited as evidence and how they often distinguished data that was useful for improvement from data that was quantifiable and used for accountability. Across both sections, we show how the state and district's institutional setting and emphasis on high-stakes accountability shaped responses to continuous improvement.

New Institutionalism in Education

Institutional theory emphasizes that people in organizations are often motivated by perceived legitimacy, rather than efficiency, and that shared beliefs or assumptions may guide behavior as much as technical expertise (J. W. Meyer & Rowan, 1977). Individuals within institutions operate based on shared assumptions about how things "should" be done as a way to maintain their reputation as legitimate actors. This is particularly true in education, where actions between different organizational layers have been argued to be loosely-coupled (Weick, 1976). The increasing focus on accountability and instructional improvement over the past few decades has changed the institutional environment of schools and districts, leading to the new institutionalism marked by tighter coupling of institutional behavior (Hallett, 2010; Rowan, 2006; Spillane, Parise, & Sherer, 2011). We focus on two main themes from this new institutionalism in education: the role of cognitive schema, such as organizational routines, in socially constructing the institution; and the need to pay attention to power in organizational change (H.D. Meyer & Rowan, 2006).

A specific aspect of the institutional environment is the presence of organizational routines, which are repetitive, recognized patterns of action that involve multiple participants and interdependent actors (Feldman & Pentland, 2003). Organizational routines are both part of a formal structure and are enacted in practice (Feldman & Pentland, 2003). That is, they exist both as an idea for how the process works and as specific enactments of practices which contribute to the routine. For instance, organizational routines to improve instruction act as a coupling mechanism by which classroom practice is coupled to state standards and assessments (Spillane et al., 2011). In this way, organizational routines can link the micro and macro aspects of institutions and the individuals within them (Hallett, 2010). Such routines are not static; they can be changed and can change organizations (Feldman & Pentland, 2003; Sherer & Spillane, 2011). The introduction of new organizational routines can build instructional coherence and change how teachers talk about teaching and learning (Sherer & Spillane, 2011) and build social connections between educators (Hatch, Hill, & Roegman, 2016). New organizational routines can also encounter resistance and bring turmoil (Hallett, 2010; Spillane et al., 2011).

Spillane (2012) outlines three benefits of using organizational routines as a conceptual framework to examine institutional behavior. First, a framework of organizational routines focuses the inquiry on interactions between individuals. Second, organizational routines emphasize patterns of interaction rather than isolated events. Third, organizational routines, with their emphasis on both micro and macro practices, highlight the relationship between organizational structure and individual agency. Scholarship on continuous improvement approaches suggest these are important aspects of the work. For example, continuous improvement emphasizes a networked approach (LeMahieu, Grunow, Baker, Nordstrum, & Gomez, 2017), with collaboration of multiple individuals around a shared problem of practice.

This requires attention to the interactions of individual as they negotiate meaning around shared problems. The focus on patterns of interactions rather than isolated events is relevant to understanding how, if at all, continuous improvement research is integrated into the daily work of educators. Attention to the relationship between the micro and macro practices reflect the need to consider not only the specific innovation that is being implemented, but the system and context in which it is embedded (Fishman, Penuel, Allen, Cheng, & Sabelli, 2013).

The Institutional Environment of Data Use

With the increased focus on accountability, districts and schools are increasingly pushed to be data-driven; data use practices serve as an organizational response to the changing accountability context (Spillane et al., 2011). There is substantial research describing how the social and institutional contexts of schools shape the use of various forms of data (Cannata, Rubin, et al., 2017; Coburn & Talbert, 2006; Farrell, 2015; Ingram et al., 2004; Marsh, 2012; Moss, 2012). Districts influence data use by setting organizational routines, establishing norms of data use, ensuring a flow of information to schools, creating opportunities for collective sensemaking around data, and providing training in data-use processes (Grissom et al., 2017; Marsh, 2012; Park & Datnow, 2009; Spillane, 2012). School organizational culture also shapes how data informs improvement initiatives (Anderson, Leithwood, & Strauss, 2010; Marsh et al., 2006). Administrators can create school-wide norms for data use (Halverson, Grigg, Prichett, & Thomas, 2007) or breed mistrust in data among teachers when data is seen as being distorted (Ingram et al., 2004).

Within this institutional environment, educators must grapple with what types of data should count as evidence. The data-rich culture of schools ushered in by high-stakes accountability policies has created access to a variety of sources of data, including data related to

student or school outcomes (passing rates, disciplinary incidents), inputs (student demographic information), school processes (data on quality of instruction or program implementation), and perceptions (surveys or focus group discussions) (Marsh, 2012). Data use is hampered by educators' perceived validity of data (Marsh, et al., 2006), a lack of consensus on how to measure effectiveness (Ingram et al., 2004), and disagreements about what constitutes evidence in the first place (Coburn & Talbert, 2006). Coburn and Talbert (2006) find norms of how data should be used vary across district stakeholders' social location and role, creating conflicting standards of what counts as evidence. Examining educator data use requires understanding the interaction between the type of data and their purpose of using it, as different types of data are used in different ways (Farrell & Marsh, 2016).

Further, educators may rely on professional judgment rather than systematically collected data (Ingram et al., 2004; Supovitz & Klein, 2003). Dormann and colleagues (2016) identify three main types of information used: external evidence, which comes from rigorous research and external experts; internal evidence, which comes from the local context but is still systematically collected and analyzed; and evidence substitutes, which come from personal experience, intuition, and beliefs. Other research confirms that professional judgement may be informed by personal experience, intuition, ideological preference, and customary practice (Bryk & Gomez, 2008; Ingram et al., 2004). In their review of evidence-based decision-making among district central office staff, Honig and Coburn (2008) find that district leaders base their decisions on multiple forms of evidence, including perceptual data, such as feedback from teachers and input from students, parents, and community members. In reference to this local knowledge, they write, "These forms of evidence seem important to central office support for school improvement efforts" (Honig & Coburn, 2008, p. 586). Thus, perceptual evidence and

evidence substitutes may be just as important as outcome data for decision-making.

The organizational routine of data use was established as the institutional environment of schooling evolved to include high-stakes accountability as a key practice and not just a symbolic myth (Hallett, 2010). With this in mind, it is important to heed the call from institutional theorists to attend to how power is embedded in institutions (H.-D. Meyer & Rowan, 2006). Any change effort that challenges embedded organizational routines—such as data use routines within a climate of high-stakes accountability—are likely to bump up against them. In the era of accountability, there is a hierarchical power structure where state and federal policies create performance mandates that are filtered through districts to schools and eventually down into individual classrooms. As the curriculum is increasingly controlled by district decision-makers, teachers' power over their classroom is being usurped by expectations of how time, space, and resources are to be used (Au, 2011). Organizational routines around data are another mechanism of control as teachers allocate their instructional resources in ways consistent with the logic of data-driven decision-making (Booher-Jennings, 2005). Thus power and cognitive schema are inherently linked; teachers' actions are mediated by these power relations (Horn, 2016; Louis, Febey, & Schroeder, 2005; Palmer & Snodgrass Rangel, 2010). This is clearly seen in the proliferation of data use and more specifically how, why, and even which data are used most regularly within educational settings.

Improvement Science and Educator Evidence Use

Improvement science is an emerging approach to school improvement that focuses on using continuous improvement processes to address problems of practice in ways that contribute to organizational learning (Bryk et al., 2015). Key to improvement science is the creation of networked improvement communities that engage in disciplined cycles of inquiry where data is

used to understand the problem and test potential solutions (Bryk et al., 2015). The specific approach to data use in improvement science is the Plan, Do, Study, Act (PDSA) cycle. PDSA is drawn from improvement science as a tool for organizational improvement (Langley, 2009). PDSA links a change idea with easily measured outcomes to answer the question: How do we know a change led to an improvement? In addition to outcome data, PDSA may involve the collection of feedback data about the usability of specific practices. These different forms of evidence are used to make organizational improvements that align with existing system processes, before testing and evaluating these improvements again. In this way, improvement science represents a particular type of organizational routine focused around the use of data.

When teachers use data to inform their instructional decisions, they usually draw on data available from the district's accountability systems. The data teachers are traditionally asked to draw on for improvement purposes may also be used by administrators or central office staff to hold them accountable. This approach is at odds with improvement science's distinction between data for accountability and data for improvement (Yeager, Bryk, Muhich, Hausman, & Morales, n.d.). While data for accountability focuses on outcomes to identify whether expectations are being met, data for improvement is more practical; it provides evidence about organizational change and is grounded in daily practice (Yeager et al., n.d.). Distinguishing whether data is used for accountability or improvement is important as educators' understanding of the purposes of data shape their behavior (J. A. Weiss, 2012). Yet if district structures and the logic of high-stakes accountability shape teacher data use (Marsh, Farrell, & Bartrand, 2014), teachers may not always have the ability to develop data use practices more appropriate for improvement.

At present, there is little research on how data is used in the context of continuous improvement research. In one study on the Building a Teaching Effectiveness Network (BTEN),

Hannan and colleagues (2015) allude to the challenges faced by the majority of schools in the network when using data for improvement efforts. Eight of the ten schools in the network struggled with documenting and acting upon what they learned. The other two were more capable of applying improvement methods. They write, “What set these two schools apart was their use of data and disciplined inquiry to amass knowledge of their school systems and their teachers’ needs, coupled with their strategic application of this knowledge in their support efforts” (p. 502). Although educators may be able to use evidence to ground local improvement efforts, they face a number of barriers. Another study of educator engagement in PDSA found that they saw value in PDSA, but also expressed frustrations in trying to carry it out, such as finding time and capacity constraints regarding data collection and analysis (Tichnor-Wagner, Wachen, Cannata, & Cohen-Vogel, 2017).

Current research on educator engagement in improvement science has been agnostic to the institutional context of data use. While improvement science begins with a recognition that teaching and learning occurs within interdependent, complex systems, and that obtaining substantial and sustained improvement in student outcomes requires significant changes in how these school systems operate (Bryk et al., 2015), there is little attention to the role of the institutional environment in how school systems engage in improvement science. Yet, research on the introduction of high-stakes accountability suggest that such large changes will face challenges as they are potentially incompatible with the current culture of schools (Ingram et al., 2004; Supovitz & Klein, 2003). Over time, accountability pressures and local data systems have increased the availability of data and educators’ capacity to work with it (Marsh et al., 2006). Accountability pressures have been found to motivate the use of data, although often in unintended ways (Bertrand & Marsh, 2015; Booher-Jennings, 2005). The literature has also

describes a tension between using data for improvement versus accountability purposes (Ingram et al., 2004; J. A. Weiss, 2012).

We hypothesize that the dominant institutional environment of schooling, which includes a culture of high-stakes accountability and organizational routines around using student outcome data, may undercut improvement science's goal to use data for school improvement. This paper provides a case study of using the PDSA process as a new organizational routine within a district with a history of reforms oriented towards high-stakes accountability. Within the new institutional environment of education, we explore how educators in this district responded to this new organizational routine. Specifically, we seek to understand how they used evidence for the improvement of a school-wide reform and what types of evidence they cited to make claims about what they were learning. We provide a picture of what it looks like when a new organizational routine is introduced that asks people to think, speak, and act in ways that are starkly different from prevailing institutional norms, routines, and expectations.

Context

The site of this study is a large, urban district in the southwestern United States. The district serves approximately 80,000 students, the majority of whom are low-income or from traditionally underserved racial/ethnic groups. Approximately a quarter of all students in the district are classified as Limited English Proficient (LEP). The district is located in a state with a long history of high-stakes test-based accountability. In response to the state accountability framework, the district has long-emphasized school accountability, instructional reform, and data-driven decision-making. For core subjects tested by state assessments, six-week benchmark assessments form the primary basis of conversations around data within schools, as the district has trained administrators and teachers in strategies for reviewing these assessment results and

other forms of student data, including student passing rates, attendance, and discipline.

This current study emerged from a seven-year study on new approaches to scaling effective practices in high schools. During the 2011-2012 school year, researchers conducted case studies of two higher and two lower value-added high schools within the district. A key differentiating characteristic of the higher and lower performing schools was the presence of systemic practices that helped students take greater ownership and responsibility for their own learning. The Student Ownership and Responsibility (SOAR) innovation was defined as building efficacy and engagement among students (Cannata, Smith, & Taylor Haynes, 2017). In Fall 2012, in collaboration with district personnel and school administrators, three "innovation" high schools were selected to be the first sites of implementation in the district. This purposive sample aimed to select moderately performing schools in the district that had the capacity to benefit from the continuous improvement reform model.

In Spring 2013, a district design team was established that included researchers, central office administrators, representatives from the innovation schools, and representatives from other high schools. The team was facilitated by external program developers. Meeting for two days every month, this district design team studied the research from the district case study schools, reviewed evidence-based practices related to the SOAR innovation, engaged in additional needs-analysis, and designed an initial approach to developing SOAR in the innovation schools. This approach focused on the interrelated needs of building growth mindsets, problem-solving, and goal-setting skills in students. In Fall 2013, each of the three innovation schools established school design teams, comprised of 6-8 individuals, most of whom were teachers. These school design teams were led by their representatives on the district design team and were expected to adapt the initial SOAR design for their local school context, use PDSA to test and refine specific

SOAR practices, and lead implementation in their schools. There were a total of 6 daylong meetings of the design teams during the 2013-14 school year.

PDSA was not an explicit expectation of the work in Spring 2013, but the district design team meetings did include several needs-analysis activities where the team discussed various types of data. PDSA was introduced in Fall 2013 as a tool to further develop the innovation prototype and align it to each of the school's unique contexts, and PDSA was the main vehicle for data use on the school design teams. External facilitators adopted practices conducive to successful evidence use, such as technical training (Marsh, 2012), cross-school collaborative opportunities (Chen, Heritage, & Lee, 2005; Murnane, Sharkey, & Boudett, 2005), and the establishment of norms for data use (Park & Datnow, 2009). When PDSA was first introduced, participants were given a general overview of this approach to continuous improvement that emphasized the role in testing practices and gathering feedback as part of the ongoing and iterative improvement of the innovation design. The enactment of PDSA was highly scaffolded with the facilitators and research partners coordinating the first rounds of PDSA testing that focused on school the teams piloting identical lessons on growth mindset and problem solving. During this initial phase of PDSA, survey instruments for students and teachers were developed by the research team, administered by school design team members, and compiled and analyzed by the research team.

In Spring 2014, facilitators and researchers continued to scaffold the PDSA process, although they now supported schools in planning and testing their own change ideas. They worked with the school design teams as they planned their work and helped them link outcome measures with their change idea. School design teams were encouraged to collect easy-to-gather data that was closely linked with their change idea. They were also urged to collect school

process data on program implementation or perceptions that could help teams understand how the innovation design should be modified, in addition to student outcome data.

In the 2014-15 school year, when schools began full implementation, PDSA continued to play a large role in the organization of their work. Meetings of the whole network occurred once a quarter, and the agenda focused around sharing what they learned from the most recently completed PDSA cycle and planning what they wanted to accomplish for the next PDSA cycle. At the end of each network meeting, the district design team (including central office administrators and principals) would attend to hear schools present what they accomplished in the past cycle and hoped to achieve next.

Data and Methods

Case Study Data Collection

The data used for this paper come from two sources. Table 1 contains detailed information about the frequency with which these data sources were collected. We refer to the primary data source as the "process data," and it includes observational and artifact data from district design team meetings, including observational fieldnotes and audio recordings of all meetings, artifacts distributed or produced at the meetings, and feedback forms completed after each meeting. Researchers collected these data sources by attending regular meetings with the school design teams (i.e., teacher-led teams) and program developers from Spring 2013 to Spring 2015. Audio data from each session were not transcribed, due to the length and complexity of each recording in which recorders captured discussions among an often dispersed groups of participants. Instead, graduate students listened to each recording in its entirety, and utilized reflection forms to partially transcribe and synthesize data according to the analytic framework described below. To supplement observations from the district meetings, the research team

conducted in-depth, semi-structured interviews with design team members. District design team members and a random sample of school design team members were interviewed in the Summer 2013, Summer 2014, and Fall 2015 to understand their experiences with the work.

In addition to attending design team meetings, we also conducted fieldwork visits in each of the innovation schools. Three visits took place between December 2013 and April 2015. For the first visit, teams of nine researchers visited each of the innovation schools for one day. For the second two research visits, teams of three researchers spent four days in each of the schools on each of these visits. Except in the case of turnover, we re-interviewed school design team members during both visits. In total during these visits, we conducted 66 semi-structured interviews with school design team members, all of which were transcribed verbatim. From these interviews, we focus on their descriptions of their experiences with data use and the continuous improvement process.

Data Analysis

A qualitative case study design (Lincoln & Guba, 1985) allows us to probe how school stakeholders, particularly the teacher-led teams, used evidence in the ongoing development of the innovation. The process and fieldwork data were analyzed separately using a framework that consisted of several *a priori* codes specific to the process and fieldwork data, in addition to codes that emerged inductively from the data. For both the process and fieldwork data, the coding process was iterative in nature with members of the research team comparing coding to ensure a consistent understanding and application of codes (Corbin & Strauss, 2008). The process framework included: attitudes and engagement in the design process; learning and understanding about design, implementation, and scale; the extent to which the design process was collaborative, needs-centered, aligned with existing system components, and included PDSA

cycles; evolving roles of design team members and integration of researchers; learning across schools; and evidence of sustainability and scale. After coding, detailed memos about each design team meeting were written around this framework.

Regarding the fieldwork data, the research team coded transcripts based on a coding schema related to the project's framework for quality implementation, which included codes related to how the school design team worked together, design team member capacity, external support to design team members, and how the school design team engaged in continuous improvement. The research team then wrote a summary memo for each field visit, probing similarities and differences across the three innovation schools. This analytic process used several strategies to address potential threats to the validity of our inferences, including cross-validation between researchers, triangulation among sources and perspectives (i.e., interviews and observations), and member checking (Miles & Huberman, 1994; Patton, 2002).

Focusing on organizational routines necessarily places attention on the actual practice of data use (Little, 2012; Spillane, 2012). The fieldwork interviews provide a chance for us to hear from design team members about their perspective on PDSA, while the process data provides opportunities to observe them working through data use activities and the PDSA process. This is particularly important as participants may not always have been good reporters of their behavior in interviews (Jerolmack & Khan, 2014). While the initial analytic framework for the process and fieldwork data differed, they overlapped in areas specific to this paper. These areas of overlap include: how design team members engaged in PDSA, their understanding of the continuous improvement process and capacity to engage in it, the integration of the research team in supporting PDSA, and the support design team members received to engage in PDSA.

Educator Engagement in PDSA

PDSA and Grappling with Data Use

In exploring how design team members responded to the introduction of PDSA as an organizational routine, we first report the findings on how they wrestled with the use of data as part of the PDSA process. Two main themes emerge, both of which foreground the overarching institutional culture of accountability under which design teams operated. First, design team members were familiar with pre-existing organizational routines to examine and act upon data, but resisted discussions that were focused around data due to their accountability-driven district. Second, they emphasized the need to frame data gathering efforts as improvement rather than accountability, yet still sought to use data for accountability. Across these findings, there is evidence that the institutional environment of the district—and its accountability-driven logic—shaped educators' responses to PDSA. We discuss these in turn.

Familiarity with, but resistance to, data use routines

It was clear that design team members were experienced in analyzing and discussing data, particularly student outcome data. In the first district design team meeting, members rated their capacity in seven areas. "Use data to identify needs" had the second-highest rating (behind "design innovative solutions"). When asked to reflect on the group's self-ratings, one member referred to data available due to the state and district accountability measures and said, they "have a lot of experience using data." Members' familiarity with data use routines was not only due to the accountability culture of the district, but also their university preparation and the more general process of reflection used by educators. For example, in the first data analysis activity, where district design team members examined existing district data on SOAR-related indicators, members demonstrated a familiarity with the data and were critical consumers. Several members asked questions about the response rate for survey data and reliability of specific measures, with

more than one referencing something they learned in a statistics class. Further, while only two members described prior experience with a specific continuous improvement model, many indicated that reflection based on data was a natural part of educators' work. One member said,

> I'm very familiar with just a change model. Basically that's what Plan, Do, Study, Act is, so whether it's this design, whether it's another design, you're pretty much looking at data, identifying a problem, coming up with an action plan as to how to address the problem, implementing it, gathering more data, looking at your results, evaluating it, and moving forward.

Another member concurred, "as a teacher myself…you teach a lesson, you have to come back and look at the data and everything, and then revise it if you need to and then act upon it and then do it again."

Despite this familiarity with routines around examining data, there was evidence of resistance to including data use routines as part of this improvement work. Some of this resistance came from questions about how specific indicators were to be measured and how they might be used for accountability. In one of the first discussions around PDSA, the research team led a discussion around possible indicators that could be used to determine if SOAR was changing practices or outcomes. As the researchers shared initial drafts of indicators and asked for design team members' help in clarifying them, several members expressed concern over how the measures would be collected. One member said, "This is going to be too much for the teachers; if we're asking them to do lessons, and collect data on all of these markers of success, it's going to be too much!" When one member pointed out that "we're already collecting a lot of this, tardy students, on time assignments, so it's not that hard," another member countered, "but we need measures of how we got there, not just the end result." In this exchange, design team members struggled with wanting to move beyond existing administrative data, but recognized the burden associated with developing new measures to collect data from their busy colleagues.

A member who was also a central office administrator was the most explicit about connecting this resistance to the district's accountability-driven logic and existing data-use routines. This member said (s)he recognized that the intent was for the design team to use the indicators to improve the innovation, but that,

> This could become entirely a tool for data, and that it might be the only focus of the district and might harm teachers…I think [the draft measures] are really good. I think that we need to have walkthrough "look-fors" to get the measures. But, ultimately, I don't think this data will matter because the bottom line is going to be standardized test scores only.

Notably, this member relied on an existing organizational routine widely used in the district (walkthroughs) as a form of data collection. (S)He was also highly aware of how any data collected, even for improvement purposes, was likely to be used in the institutional context of the district that prioritizes accountability and impact on high-stakes assessments. (S)He expanded on this concern in an interview, where (s)he linked the accountability culture in the district to her/his concerns that the district would remove the innovation from its school-based context and distort the underlying goals of the innovation in the process. (S)He summarized:

> When we take something that should be going on in class that is very instructional and about relationship building, and conversations, when we boil it down to this, leaders will just look at the data report and not go into the class and care about the conversations and the instruction that is happening. We're boiling down a human-centered objective into a data point that will be hung around necks as an albatross at some point to collect data, to compile data and to say it's not being effective.

Other members expressed similar concern that they would lose something by quantifying complex aspects of teacher practice and student beliefs. For example, during a small group discussion about PDSA, a member said,

> Growth mindset is such a personalized whole student. All this will turn into, who knew it, yes or no. It's a data report, check the boxes. It's not about the kid. That's the only thing I'm worried about with all this innovation is how is it not going to become a data numbers check box. And how is [it] going to stay focused on making sure the kids are there?

As evident in this quote, this member was worried that a quality such as growth mindset will be turned into a number, particularly that this qualitative change they were trying make would be turned into a "check box." Similarly, later in the year, another member said they tried to measure student growth mindsets, but it "was a struggle at first to determine how to quantify and measure growth mindset when it seems like a more qualitative issue." In another activity, another member expressed a similar concern about how he wanted to use qualitative information to assess their progress, but felt pressure to get specific about measures:

> If we get some things going and if we're trying to be so specific about [measuring] this is the specific thing we're trying to improve. I mean, I'm going to know if our implementation went well if I can go and ask a kid, "What did you do this year?" and the kid says, "Well this year was different because we did blah, blah, blah…." Even if it didn't produce the other outcomes, I've changed the behavior.

For these teachers, the pre-existing organizational routines around data in a context that privileges quantitative data meant they felt pressure to produce quantitative measures, even though they expressed skepticism that a quantitative indicator could provide useful information about the qualitative changes they were trying to create. In this way, the institutionalized approach to data use influenced how teachers engaged in a continuous improvement process.

One way that members tried to resolve the challenge of burdensome data collection in the PDSA process was to use data that existed in the district's existing data system. For example, the district prepared comprehensive data reports every six weeks for each school, which included data on attendance, discipline, passing rates, and benchmark assessments. Members wanted to incorporate these data into a PDSA cycle, but struggled because the data were not directly related to what they wanted to know. For example, in Spring 2015, one member said,

> What are we comparing -- if we're looking at change, we're comparing students' change from the beginning of the six weeks to the end of the six weeks or -- Because it seems

> like comparing this six weeks versus last year's same six weeks is just completely different data, right? Like totally discontinuous. So that's my problem with the PDSA cycle.

This member conflated the PDSA process in which she was asked to engage with the specific data on student outcomes that was readily available in the district. In this meeting, members at one school spent a large period of time debating what the appropriate comparison should be when using the six week reports, such as last year's 9th graders to this year's 9th graders (who are different students) or last year's 9th graders to this year's 10th graders (who are approximately the same students but had matured). Here, we see what appears to be a conflict between the district's existing routines around data and the needs of the newly introduced routine of PDSA.

Other design team members expressed concerns on the potential misuse of any data they collected. While discussing an article the design team was given that explained the PDSA process, two members[1] at one school described how administrators in the district used low-stakes data to misrepresent teacher practice. This was evidenced in their following excerpt from their conversation:

> Deanna: I think that part of the problem that we're having is that we haven't figured out ways to monitor, to assess, that what is working is working. And I think [that is] part of the issue with lots of the district initiatives we've had in the past. Let's use learning walks. How do we know that learning walks are working?...
> Rebecca: You can you have wicker posters, and your pictograms, and your word wall but, cool, you checked it off, but really how are you practicing that? That's really the important question.
> Deanna: … there are eight out of ten teachers who have wicker posters up or are doing word walls. That to me is an arbitrary number that has in no way to do with what I'm doing in my classroom.

The evidence that most design team members had negative experiences with data also came

---

[1] All names are pseudonyms.

through in a workshop focused on deepening understanding about PDSA. As an introduction to the workshop, the facilitator asked for a couple members to share their experiences with data, and the members who answered shared negative experiences of data use. When the facilitator then asked if anyone would share a positive experience using data, no one spoke, and a few members laughed in the awkward silence. While their connection of PDSA to a cycle of inquiry showed that many members did have positive experiences of using data for personal reflection, in this context, most members thought of data in negative terms. The concerns about the use of data within the district's accountability-focused environment appeared to shape their overarching skepticism about how data could be used and misused to hold schools and teachers accountable.

Continued ambivalence about data for improvement and accountability

Despite this concern about potential misuse of data by the district, design team members grew to value data use for their own purpose of continuous improvement. The more they engaged in PDSA and understood its purpose, the more they were able to see data gathering as a tool for improvement rather than accountability. Evidence of this shift in attitudes was apparent at the beginning of the 2014-2015 school year, when design team members indicated that the data they gathered as part of the PDSA process helped them to refine the SOAR innovation. At one school, one member said,

> I like the review part, because a lot of the time other programs and initiatives I've tried out, they just come in and they say, 'hey, try this', and then it's vaguely a did it work or did it not work. It's not a, let's go back and review it. Let's see what we could change. Let's try it again… I like the forced aspect of, you have to go back and review, and tweak, and try it again.

At another school, a member emphasized how their team elicited feedback from a variety of school stakeholders as part of their improvement process:

> I think that just the open communication and also how we continue to refine [SOAR] and make it our own. I feel like we're doing a really good job as a team.

> Like what's not working?…being like, okay, it's not working, why, and getting other people's input and really trying to solicit other people to make…whatever we're doing better.

Still, given the accountability context of the district, design team members were sensitive to the ways in which their colleagues would regard their data collection efforts and were worried that any data teachers were asked to collect would be seen as part of the accountability-driven district culture. One approach used by members was to intentionally shift their language to emphasize the ways in which the data they collected would be used non-punitively, hoping the staff at their school would be more sympathetic. A design team member from one school, stated,

> It seems like that's the language that would make this easier. What we want to see is how it's being done and how we need to change it. Not, you're being held accountable [for] whether or not you're doing it. We want to know how's it going.

Other members of the team recognized the need to convey to their staff that they would be collecting and reviewing data related to the SOAR innovation, but that it would not be used for accountability purposes. They emphasized how feedback would be vital for the process of improving the innovation, so that SOAR was usable for teachers and could lead to positive student outcomes. An administrator on the district design team who worked a school that was not one of the targeted innovation schools, summarized this idea as, "Evaluation of you as a teacher versus the evaluation of the program." Another administrator at another school added,

> We're monitoring to see what works well, what doesn't work well… If somebody knows I'm doing this to see if this is successful, I want to see where the gaps are, I want to see how to make this better. That's a different motivation [than the teacher evaluation system].

Ironically, despite wanting their use of data to focus on improvement and not be misused as a form of accountability, design team members also began expressing support for using data for accountability. First, they contended that some accountability was necessary, or teachers would likely discount the innovation as unimportant when held up against all the other demands

placed upon them for which they were held accountable. Further, it was a signal of legitimacy given the district culture. Initially, the emphasis on having some accountability structures was expressed primarily by the administrators on the design team. During a whole-group conversation, one administrator on the district design team said,

> There has to be an accountability piece, otherwise I'm not going to do it because I don't want to. There has to be some level of, you know, we want you to do this. If you're struggling, we need to know how to help you, but this is a requirement, you must do this.

This administrator is indicating that the accountability culture is so pervasive that if people are not held accountable for particular initiatives, those initiatives are perceived as not important. Later in the school year, as they began to prepare for full implementation, teachers on the design team were more open to the need for outcome measures that would satisfy district demands for accountability. Shifting from her prior attitude that data collected to monitor implementation were not good measures of what was actually happening, Deana recognized the need for more systematic data collection to monitor implementation due to its perceived importance in the institutional environment. Deana said,

> I don't think we can in good conscious say to [the district], this is something that is working for us if we have not figured out a systematic method of checking that the things that we are actually doing, that we are planning right now worked. And if we don't come up with a method to check did [SOAR] work, we can't tell them this worked for us so we should do this too.

Another member, who also initially resisted that qualitative ideas like growth mindset could be quantified, later emphasized how they could monitor proximal student outcomes:

> If a student is taking ownership of their learning, that is a mindset, that is a thing that is not easily measured. However, we should know that they are going to be getting their homework done, they are going to be doing these things that we can easily report on with a checkbox even though that's not really getting at the core. So when we're looking at the data we need to collect, we are going to intentionally look at these very basic [measures such as] homework completion, zeros, tardies.

Despite wanting to clearly differentiate data collection for improvement from accountability, as members looked ahead to full implementation, they began to realize the difficulties of such a clear distinction and the role of data in convincing the district their work was valuable. Improving implementation required having some mechanism for knowing the extent to which teachers were implementing the innovation and they were achieving positive outcomes. As members gained experience with PDSA, there was a shift in their attitudes as they began to embrace the use of data for improvement. Yet the institutional logic of the district continued to shape their response to PDSA as they expressed continued ambivalence about how to use data for both improvement and accountability.

Using evidence for improvement within an accountability-driven setting

The data allow us to explore not only members' responses to the use of PDSA as part of this improvement approach, but how they engaged in PDSA. This section focuses on the ways design team members used evidence and how the specific context of evidence use shaped what forms of data they privileged. The definition of internal perceptual evidence draws on Dormann and colleagues (2016) definition as evidence that is based on stakeholder perceptions, but is collected by individuals in the school, such as teacher and student feedback surveys or focus groups. Internal perceptual evidence is distinguished from evidence substitutes in the extent to which the data is systematically collected and analyzed. Evidence substitutes include non-systematically collected information that may include their personal experience or intuition that informs their professional judgement (Dormann et al., 2016; Ingram et al., 2004). Overall, design team members relied most heavily on internal perceptual evidence (especially teacher feedback) and evidence substitutes to guide design and implementation. Yet when school design team members presented their work to district administrators, they privileged student outcome data.

Evidence use amongst equals: Primacy of internal perceptual evidence

Table 2 presents information on the focus of each school's PDSA cycle and the types of evidence they drew upon over two years. When teams began using PDSA at the start of the 2013-2014 school year, the research team created student surveys that assessed student understanding of growth mindset and problem solving for the first two PDSA cycles. By spring 2014, teams were given more leeway in conducting their own PDSA cycles. As evident in this table, school design teams placed a high value on internal evidence from both teachers and students through surveys or focus groups to get feedback about specific innovation practices.

Examples from the three schools illustrate how teacher feedback, in particular, was seen as most beneficial for making improvements to the innovation on their campuses. For instance, for Cycle 3 in Spring 2014, design team members at the three schools piloted three separate activities. The Forest Glen design team tested the revised growth mindset lessons with a larger group of teachers. Valley tested the response to a professional development around growth mindset-oriented praise language. The Desert Grove design team piloted a behavioral reflection activity designed to help students take responsibility for their behavior, rather than routinely sending them to the front office. Valley did not specify in advance what evidence they would collect in their PDSA cycle, but both Forest Glen and Desert Grove planned to collect evidence about teacher and student feedback on how these practices were working. At the next meeting, the teams shared what they had learned. Valley described what the professional development looked like, with only a brief reference to "overall, teachers were very receptive." Forest Glen also came back with no evidence, but described how they "spent a lot of time creating the questions and the scale" before encountering problems in administering the survey. Desert Grove collected teacher and student feedback and reported on this to the whole group. What seemed to

most excite the Desert Grove team was the feedback they received from teachers. When reporting on their experience testing this practice, one member described that they met with volunteer teachers to explain the reflection form and then said,

> We had another meeting a few weeks later where teachers gave us feedback on it. The disciplinary sheet, the feedback we got was that it works really, really well on freshman and sophomores… one of the great things was we got results that we were looking for. I mean, we were hoping that it would be a certain thing. We were hoping teachers would like it and feel like it was useful, and they did.

The team continued to emphasize the value of this teacher feedback, even though they also had data from students, which they described after a question from another school: "We learned that students don't want their classmates to be immediately sent to the AP [assistant principal]. Students with disciplinary infractions also expressed that they preferred the timeout forms rather than being sent to the AP." Both Forest Glen and Valley were excited about what Desert Grove had learned and made plans to test the practice in their own schools.

As full implementation of SOAR began in Fall 2014, design teams experienced less difficulty executing PDSA, and the process had begun to become routinized. Each of the meetings in 2014-15 had a similar agenda: time to share informally across school design teams what they had accomplished in the past cycle, school-based time to consolidate their learning and plan for the upcoming cycle, and time to share with the entire district design team what they had accomplished and what their next steps were. This format provided an avenue to explore how teams used evidence for improvement and how their use of evidence varied by audience.

Forest Glen and Desert Grove continued to rely heavily on perceptual evidence, particularly teacher feedback. For example, in Cycle 1, Desert Grove reported on teacher buy-in on growth mindset before and after its introduction to students at the beginning of the year, showing that 11% more teachers strongly agreed that the growth mindset lessons for students

would be worthwhile for the school. Forest Glen reported that 90% of teachers agreed or strongly agreed that SOAR is moving the school in the right direction. In interviews, design team members from both schools explained that they valued this perceptual evidence as a way to ensure they were meeting the needs of their school and build teacher buy-in. A Desert Grove member shared, "…going through the [PDSA] cycle we've …refined the lessons to fit what our campus is going to be using for professional development, which then means we're changing it to fit our campus." By eliciting feedback from their colleagues, design team members customized the innovation to meet the needs of their teachers. Another teacher on Desert Grove's team, described how this feedback was treated as a tool to build buy-in so teachers "feel like they have a say or like they're being heard." A member at Forest Glen similarly said, "We were always collecting and looking at data and figuring out what our colleagues wanted, needed, and trying to base our next actions based on that." During early implementation in these schools, the design team members focused on teacher feedback to improve SOAR in ways that better met the needs of their teachers, treating PDSA as a tool to build teacher buy-in.

Evidence substitutes and decision-making

This high use of perceptual evidence in PDSA cycles masked how the design teams actually made decisions to improve the innovation on each of their campuses. In the broader context of improving the SOAR innovation, design teams often used evidence substitutes, such as their own experience, intuition, and anecdotal information, in addition to perceptual evidence that was systematically gathered. One example of how design team discussions around PDSA did not necessarily inform decisions about improving the innovation comes from Forest Glen's second PDSA cycle in Winter 2015. As demonstrated above, members of the design team at Forest Glen describe using teacher feedback as a tool for monitoring implementation and better

understanding teachers' perceptions of the innovation. For a PDSA cycle around their disciplinary reflection form—the "Think It Out" form—they surveyed ninth grade teachers to understand how it was working in this grade. They asked teachers the following questions:

1. Per week, how many Think It Out sheets do you sign?
2. Think It Out sheets prevent what percent of your students from needing another intervention?
3. Do you think it would be beneficial to move to a campus-wide form—do you think other kids, other grade levels would benefit from it?

As they discussed the results from this PDSA cycle, they noted that eight out of nine freshman teachers thought this tool was an effective behavioral intervention and that it was keeping behavioral problems from escalating. They completed the "act" portion of this PDSA cycle by saying that the form should be expanded by introducing it to teachers in the tenth grade. They also outlined some minor changes to how the form was implemented within the context of the school's larger behavior management system. However, discussions of this form were absent from later discussions that day about their next steps for implementation, and they did not spread this practice to tenth grade. Having completed a cycle focused on this disciplinary reflection form, their attention had turned to how teachers mentor students in the advisory period.

Similarly, during an end of year presentation on what they learned, two members of the Desert Grove team did not share results from the teacher survey they conducted, but described ways they drew from their own experience implementing a grade monitoring activity to identify areas for improvement for the next school year. These included:

> Member 1: To have proactive rather than retroactive grade recording. So maybe doing it at the end of six weeks, maybe doing it in the second and fifth week of the six weeks so students can see what their grades are before their final instead of after its been finalized and having to react to what's happened.
> Member 2: Looking at the rewards and incentives, doing a better job, a more consistent job of rewarding and incentivizing students in their successes, honor roll, things like that.

We see that in the case of the Desert Grove design team, their own experience informed

improvements to the design alongside perceptual evidence that was systematically collected from the broader school community. During the first meeting of the 2014-15 school year, a member from Valley used her own classroom as an example of the success the school experienced:

> My class, personally, of my students 33 percent made A/B honor role. And they just took the poster and took control of the class and I sat back and they said, 'we're going to go for 100 percent [on A/B honor roll],' and I was like 'wow, okay.' And they're like 'we are.' And I said, 'how are you going to get there?' And they wrote their five steps and even came up with an accountability buddy, and they exchanged phone numbers, and they said they were all going to make sure each was passing, and they all signed off on it… So they took ownership of it.

Anecdotes that members told about their own experiences tended to be positive. When anecdotes were shared about other teachers in their schools, they sometimes focused on peers who were antagonistic to SOAR. While sometimes these comments were dismissed, they were often used as opportunities to identify ways that could improve the innovation, or how staff was trained in its delivery. For example, the following exchange occurred in Spring 2015 when the Valley team discussed the "XYZ chart," which was the form the developed for students to monitor their grades:

> Member 1: It's interesting there has been pushback, not from not many teachers, but just one teacher who has been very vocal about feeling like the XYZ chart shows kids how little they need to do.
> Member 2: Well there is always going to be a negative side to everything.
> Member 1: I just want to acknowledge that there was this presence of it.
> Member 3: But that's when teachers [should] take that opportunity to emphasize where they want to go and not just getting by. Yeah that exists, so there's your opportunity to instill growth mindset.
> Member 1: That might be an interesting conversation for us, to look for opportunities to have as a faculty. To say, 'we recognize that a kid may look and see that they only need to get a 30 to pass. How do you as the teacher - what are ways that we can respond? What are teachers' ways to respond to kids when they set their standards that low?'

In this instance, an anecdote about a teacher unsupportive of the innovation is dismissed by one member of the team, but, for others is treated as an opportunity to improve their delivery of the innovation practices to the faculty.

Another form of evidence substitute is intuition, which design team members routinely drew upon. In an interview in fall of 2014, one member illustrated this point:

> We found that some of the lessons were too long, so we changed them. We found that in order to implement properly, that [SOAR] needed to be done in a completely separate grouping outside of the classroom for multiple reasons, one being that it does take away teachers' time from their content, directly anyway, and then two, having it in a separate setting, we could control the class sizes better so that you wouldn't be presenting it to 30 kids. So now with this advisory period that we've implemented, you're just presenting it to 15 kids…

While this member used the phrase "we found," (s)he was not referring to data from piloting the lessons, but work their team did thinking through the logistics of the time and space for how SOAR would be implemented. A Desert Grove member also said that their reflection involved intuition informed by their professional experience:

> I think that we operate in many ways very conversationally around the idea of reflecting on what went well and what did not. We could definitely do a better job of trying to track that, and have something that we could use as an ongoing history of how the work has changed, and how maybe we've learned some lessons that we could pass on.

In this quote, this member was recognizing that they made decisions based on a "conversational" type of reflection that was not documented or accessible to those not on their team.

Finally, looking across the practices tested in the various PDSA cycles in Table 2, few teams saw these cycles as iterative, a chance to test a refined practice again or in a new context. In almost all instances, teams approached the focus of their PDSA cycles as singular events, something that they implemented, collected and reviewed evidence on how it went, and then moved on to a different practice altogether. One member from Desert Grove expressed frustration with their team's approach to PDSA as a "missed opportunity" because "we'd kind of get back into, well, just kind of planning your next one." Forest Glen, in particular, often did not complete the PDSA documentation, deciding to use time allotted for school design teams to reflect together on the implications of what they learned to instead immediately begin planning

for the next activity. In a later session, one member wrote on a feedback form that (s)he was realizing that PDSA cycles did not have to be "discontinuous cycles that are not building off each other."

Evidence use with administrators: Using "data" where it counts

As noted above, the meeting agendas in 2014-15 included a presentation to other school design teams in the morning and a presentation to the broader district design team in the afternoon, which included administrators from each of the schools and the central office. Thus the meetings offered a chance to see how the design team members discussed evidence with various audiences. During these meetings, it became apparent that the district focus on outcomes and accountability created challenges for the design teams who were trying to use data for improvement. For example, in the fall meeting, Desert Grove described the results from the teacher survey they conducted, the lessons they took from these data, and how they were refining their approach to move forward. When they opened for questions, the only question was from a district administrator, who asked "do you have any metrics to measure if it is working with kids?" The school design team member responded, "we are looking to track attendance…We are also looking into more ways to measure these effects. A lot of the data is qualitative and intangible evidence from students." While that answer appeared to satisfy the administrator, it was not an accurate reflection of the data they discussed earlier that day. Indeed, they did not plan on collecting attendance data, but did plan to collect student artifacts in order to examine student engagement in the grade reporting activity, as well as student grades.

The interaction around Forest Glen's presentation was similar. Like Desert Grove, they shared positive results from a teacher survey, as well as positive results from a student survey. After the presentation, one of the first questions from a district administrator was again about

outcomes, “have you noticed any changes in students?” In response, the principal interjected and prompted the design team member by saying, “[Forest Glen] students are doing well so far this year, could you talk more about that?” The teacher responded, “We have not got to that data yet, but we can see it every day. We can see SOAR in students and [we are] seeing the results.” Through these reactions to their PDSA presentations, design team members saw district administrators focus on student outcome data. Notably, there were fewer questions for the Valley design team, whose presentation was given by an assistant principal that was a new addition to their school team. Instead of sharing perceptual evidence, this administrator shared that the use of the disciplinary reflection sheet was associated with a decrease in disciplinary infractions, which was one of the indicators the district routinely evaluated in their six-week data cycles. The importance of student outcome data to district administrators was again emphasized when a senior district administrator closed the meeting by noting that the district needs to make a commitment to this work, and said, “the data will be important to show folks. We need to think about engineering opportunities for people to see changes in behavior.” In this meeting, administrators were espousing the district’s norm of using outcome data as evidence of whether the innovation is a valuable investment of time, while teachers were trying to use data for continuous improvement. As they struggled with differentiating data for improvement and data for accountability, district administrators made clear which one they valued.

At the second set of presentations to the district design team that year, the school design team members again described the perceptual evidence they had collected. Desert Grove and Valley also described student work they examined and what they learned from it. After each school’s presentation, which were done by teachers, administrators from each school added their own thoughts. All of them focused on indicators from the district’s data system. At Desert

Grove, the administrator added,

> I will share one small bit of data. The first semester our passing rates for our ninth grade students who are traditionally our most difficult students, especially when they are coming in from middle school and they're trying to get the high school way of things, they've increased a minimum of 12 percent across the board.

An administrator from Valley explained how at their school, "Behavior of our kids has improved, and looking at the data, disciplinary referrals have been down." Another administrator at Valley reiterated the benefits from the use of the behavioral infraction form, saying, "The discipline rate has decreased significantly, and the freshman path to graduate is up to 95 percent." An administrator at Forest Glen qualified why they have not seen positive increases in these measures at their school:

> So our failure rate is not to say, you know, we corrected it by 15 percent or anything like that. We already had a good passing rate. And we had a good attendance rate. And we had a good discipline rate. It's not that it's making that kind of impact, but performance related to core courses, well all of the courses, we are seeing that difference.

This distinction between data school teams found useful for improvement, and what they felt pressure to share in district design team meetings, was most apparent in the third meeting of the year. As the end of the school year was approaching, design team members felt pressure to report on improved student outcomes. At this meeting, the Desert Grove team focused student outcome data around passing rates, one of the indicators used in the district's six-week data reports. The team spent their time collectively analyzing a set of tables that compared passing rates from year to year. Forest Glen and Valley, however, did not have similar quantitative data to discuss. They did have various forms of perceptual evidence from teachers, and in the case of Forest Glen, from students. Specifically, the Forest Glen design team had conducted a teacher and student survey and had shared the results with the entire faculty the previous day in a series of focus groups. Their conversation focused around the discussions they had in that faculty

meeting about the survey results. Despite this rich amount of data, a member from Forest Glen began the presentation to the district design team by apologizing for "still not having data," by which he meant data on student outcomes. He then went on to describe their survey data. The team identified data as inclusive of only those things to which they were held accountable. In an interview, one of the team leaders referenced the disciplinary reflection form and how their principal shared data on disciplinary infractions: "As far as the Think It Out stuff, that was the first time we had actually had, like, real data." Again, despite having collected several types of teacher and student feedback, this member distinguished that perceptual evidence from "real data." This point is indicative of tension between data for accountability versus data for improvement. Teams came to learn to value and use perceptual evidence for improving the innovation, but still felt constrained by the district's institutional context to share data from the accountability system with those outside their team.

Valley organized their third PDSA cycle around developing a schoolwide approach to problem-solving. Their aim was to get feedback from teachers and students after it was introduced. They did not execute a plan to get feedback from students, but facilitated discussions about their experience implementing this problem-solving approach during a series of department-based meetings. They also encouraged teachers to tell them how they adapted the approach in their own classroom. While planning for their presentation to the district, several members voiced concern that they did not have any data. For example, one member expressed frustration that they "haven't decided how [they] are going to measure anything." They began brainstorming a possible teacher survey, and another member said they needed to "measure student understanding and change in behavior." The facilitator reminded them of the structured discussions they had with teachers, and the team began discussing specific points of pushback

they heard from teachers and how they would revise their approach moving forward. The facilitator said, “this is what the continuous improvement cycle is supposed to be about…take some credit for [the teacher feedback you collected] and this is a great opportunity to share with faculty…a 2.0 version.” Another facilitator recognized the team felt “pressure to look at outcomes like grades or attendance, but there other, smaller data points that can show whether we are engaged in improvement.” In her presentation, the design team member who presented led with a recognition of the importance of quantitative data, by saying,

> Everything we’ve done this year is increasing As and Bs at our school and decreasing behavior problems, but we set that to the side, and within this, we changed our aim to create more of a schoolwide vocabulary and process for exploring and solving problems so that we all have a common language…we’ll concentrate more on As and Bs for the future, but for now we want that common language.

Across these two schools, we see the influence of the district institutional context as they felt pressure to focus on student outcome data, even as they used other evidence for purposes of improvement. In fact, the hesitancy among design team members to present their teacher surveys as data was noted by a researcher. This research team member had missed a few meetings, but reflected on the progress the school design teams had made, saying,

> I sensed some reluctance and anxiety last summer about the continuous improvement and that today it feels like it’s being incorporated into implementation… on the measurement side…you are actually getting a lot of data that you are not framing as data right now. Think about them as intermediate measures, which I define as those short-term things that need to happen for the ultimate outcome to occur.

The third school, Desert Grove, also demonstrated how design team members prioritized data they thought the district would value, even as they used perceptual evidence in their decision making. For example, when preparing their presentation for the district, a researcher asked, “Can you say a little bit more about the data you’ve been reviewing, how it’s influenced some of the changes you’ve made, and, if you could start again, what additional data would be

helpful to you in making decisions?" A Desert Grove member responded by reviewing the detail by which they analyzed passing rate data.

> We're totally looking at passing rates right now. We have passing rates by grade level and subject area. And we have it all the way down to ethnicity as well. So one of the things we've been looking at particularly today is comparing last year's data to this year's data, and how did ninth graders do in the first six weeks of 2013 and 2014 and passing rates, how many kids passed English in the first six weeks of this year. We're also looking at ninth grade last year to tenth grade last year, see how they improved. And across the board, we're seeing improvements for the most part. Fourth six weeks is where we have not seen as much change, as much improvement. *I don't know if there is any other data right now we feel like we've collected.*

Another member from that team reminded her of the teacher and student surveys they collected, but the first member discounted that by saying they did "use teacher surveys and teacher feedback as well…that data is much more qualitative than quantitative." Two other members described how the teacher feedback data was influential in revising what they did. One member offered:

> We've used that to adjust things in our grade reporting… we incorporated some more [professional development] about adjusting and differentiating the sheet for their particular students. … And we also decided to completely adjust our grade reporting schedule because a lot of teachers found that they didn't have enough time to complete everything, and so because of that, we were able to take a couple more minutes from all seven class periods for that grade reporting. So we changed grade reporting from about 23 minutes to about 35 minutes.

As noted above, the Desert Grove design team was clearly using perceptual evidence from teacher feedback to inform how they revised SOAR in their school, yet when thinking about a presentation to the district design team, that perceptual evidence was discounted in favor of describing details about student outcomes. Across the three schools, the findings suggest that perceptual evidence was considered acceptable for their own work focused on improvement, but their idea of what constituted "data" shifted when they discussed their work with district administrators to better align with the district's accountability-driven logic.

Discussion and Conclusion

A central tension in the literature on educational data use relates to the ways in which data should be used for improvement versus accountability purposes (Ingram et al., 2004; J. A. Weiss, 2012). Improvement science, with its use of PDSA as a tool for organizational change, applies data use routines to build an understanding of effective practices and the conditions that enable their success (Bryk et al., 2015; Cohen-Vogel et al., 2015). Evidence from this study suggests that when PDSA is adopted as a new organizational routine within a district with a deep-seated history of high-stakes accountability, this distinction between data for improvement and accountability becomes blurred. In one instance, design team members supported the use of data for accountability, contending that teachers only viewed the SOAR innovation as legitimate if school-wide participation was monitored. In other instances, design team members resisted any efforts to quantify their change efforts, as they felt the logic of high-stakes accountability would distort the underlying goals of the innovation in the process.

To understand how design team members adopted PDSA as part of their school-wide reform efforts, we argue that their approaches to the use of data are best viewed through the lens of the new intuitionalism in education. While there was some resistance to PDSA, over time, teachers learned to enact this routine and adapt it to their specific needs. Moreover, we found that the district culture of accountability and data-driven decision-making shaped their engagement in PDSA. These findings reflect prior research on how the social and institutional context shapes educator data use (Coburn & Talbert, 2006; Farrell, 2015; Grissom et al., 2017; Honig & Venkateswaran, 2012). Three aspects of the district context were particularly important in shaping teacher engagement in PDSA. First, their initial resistance to developing measures that could be used in PDSA was associated with concerns that the resulting data would be misused by

the district. Their past experiences with data-use reflect prior research where the data do not provide enough guidance to future action while also causing demoralization, disruption, and loss of local control (Mintrop & Sunderman, 2009; Supovitz, 2009).

A second mechanism by which the institutional culture shaped their engagement in PDSA is that even as they resisted the use of data for accountability, they also embraced accountability indicators when they thought about their work as leaders of the implementation of SOAR. As they implemented their innovations in their schools, design team members articulated the belief that without some form of accountability, the reform would not be taken seriously. Accountability within their schools was in fact a signifier of importance. In other words, design team members recognized that the data collection required by PDSA had its own accountability logic as teachers would interpret SOAR as important if they were asked to provide some data to the team. This reflects the institutional logic of their district, driven in turn by state and federal policy, that prioritizes accountability as a central goal for schools (Rowan, 2006). As shown elsewhere, teachers may express resistance to test-driven accountability, while still enacting the institutional logic envisioned by such accountability (Booher-Jennings, 2005; Palmer & Snodgrass Rangel, 2010). Even as the design teams adopted new routines around data use, the overarching institutional environment was salient and repeatedly seeped its way into how they went about doing their work.

A third aspect of the district culture around data and accountability that shaped their engagement with PDSA was in how they defined data in specific ways. Specially, our analysis demonstrates that design team members used evidence differently depending on both purpose and audience. While working as teams, with the researchers and facilitators, participants relied mostly on perceptual evidence and evidence substitutes. They valued teacher feedback, depended

on their own professional intuition, and, to some extent, used student feedback to shape their efforts. This is consistent with other research that suggests educators rely on feedback from stakeholders and evidence substitutes to ground their decision-making (Honig & Coburn, 2008; Ingram et al., 2004). However, when teams were asked to present to district-level administrators they shifted their stance, downplayed this evidence, and instead privileged quantitative data on student outcomes, under the assumption that this is what district administrators wanted to see. As teams navigated the institutional power structure, and interacted with figures to whom they were typically accountable, they shifted their approach to align more closely with their dominant institutional arrangement. This different definition of what counts as evidence reflects prior research on patterns of data use across schools and districts, with stakeholders' role and social location influencing their definition of evidence (Coburn & Talbert, 2006). It also reflects that educators came to realize that different forms of data need to be used in different ways (Farrell & Marsh, 2016). In some instances, data was used instrumentally to make decisions. In others, it was used conceptually to shape how their thinking or symbolically to mobilize support (Coburn et al., 2009; Penuel et al., 2016; C. H. Weiss, 1977). Ingrained institutional norms influenced design teams' efforts to legitimate SOAR to powerful people in the district.

There are several implications of these findings for supporting educators in enacting continuous improvement approaches such as PDSA. Improvement science emphasizes that measurement for improvement and measurement for accountability require different questions and different measures, and should operate in different social contexts (Bryk et al., 2015). Yet this study highlights challenges with attempts to focus data-use around improvement rather than accountability. Specifically, educators cannot always control the institutional context in which they work. The use of evidence for improvement was more apparent when the school teams were

working with each other, but the context of sharing their accomplishments with school and district administrators created pressures to measure outcomes that, while important, provided less guidance for action. While the design team members valued the focus on improvement and recognized that the improvement questions they were asking would not be answered by student outcome data, they were challenged to adopt an organizational routine with a logic quite different from the way in which data is typically used within educational organizations. Even when teachers are able to adopt new organizational routines (e.g. those around data use), the institutional environment in which they continue to operate remains the same, suggesting the need for a larger district infrastructure of continuous improvement to provide guidance on how to reconcile both improvement and accountability purposes of data use (Peurach, 2016).

Further, educators still work in educational environments where accountability reigns supreme, and the pressures of this reality continue to influence how they use, present, and conceptualize data. Emerging research on other improvement-focused measurement systems reflect similar challenges in grappling with both improvement and accountability purposes (Marsh, Bush-Mecenas, & Hough, 2017). Our findings demonstrate that continuous improvement and its organizational routines and logic can function in environments of high accountability, but these findings should also encourage proponents of improvement science to pay close attention to the macro-level pressures that structure the everyday experiences in which teachers and other educators work. More generally, these findings highlight the ongoing importance of institutional environments, and how they shape reform efforts that attempt to change them. Educational leaders who wish to engage in more continuous improvement in their system need to attend to this institutional context of accountability and how it shapes behavior. Indeed one of the greatest obstacles to the ever ringing call to reform education is the institution

itself, with its deeply embedded customs, norms, routines and structures of power.

References

Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading Data Use in Schools: Organizational Conditions and Practices at the School and District Levels. *Leadership and Policy in Schools*, *9*(3), 292–327. https://doi.org/10.1080/15700761003731492

Au, W. (2011). Teaching under the new Taylorism: high-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45.

Bertrand, M., & Marsh, J. A. (2015). Teachers' Sensemaking of Data and Implications for Equity. *American Educational Research Journal*, *52*(5), 861–893.

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231–268.

Bryk, A. S., Gomez, L., & Grunow, A. (2011). Getting Ideas into Action: Building Networked Improvement Communities in Education. In M. T. Hallinan (Ed.), *Frontiers in Sociology of Education* (pp. 127–162). United States: Springer Publishing.

Bryk, A. S., & Gomez, L. M. (2008). Reinventing a Research and Development Capacity for Educational Improvement. In F. M. Hess (Ed.), *The Future of Educational Entrepreneurship: Possibilities for School Reform* (pp. 181–206). Cambridge, MA: Harvard Education Press.

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.

Cannata, M., Rubin, M., Goldring, E., Grissom, J. A., Neumerski, C. M., Drake, T., & Schuermann, P. (2017). Using teacher effectiveness data for information rich hiring. *Educational Administration Quarterly*.

Cannata, M., Smith, T. M., & Taylor Haynes, K. (2017). Integrating Academic Press and Support by Increasing Student Ownership and Responsibility. *AERA Open*, *3*(3), 1–13.

Chen, E., Heritage, M., & Lee, J. (2005). Identifying and Monitoring Students' Learning Needs With Technology. *Journal of Education for Students Placed at Risk (JESPAR)*, *10*(3), 309–332. https://doi.org/10.1207/s15327671espr1003_6

Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What's the evidence on districts' use of evidence? In J. D. Bransford, D. J. Stipek, N. J. Vye, L. M. Gomez, & D. Lam (Eds.), *The role of research in educational improvement* (pp. 67–87). Cambridge, MA: Harvard Education Press.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the Terrain. *American Journal of Education*, *112*(4), 469–495.

Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A. R., & Wang, Q. (2015). Implementing Educational Innovations at Scale Transforming Researchers Into Continuous Improvement Scientists. *Educational Policy*, 0895904814560886.

Corbin, J. M., & Strauss, A. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* (3rd ed.). Los Angeles, CA: Sage Publications, Inc.

Dormann, C., Binnewies, C., Koch, A. R., van Ackeren, I., Clausen, M., Preisendörfer, P., … Zlatkin-Troitschanskaia, O. (2016). Transferring best evidence into practice: Assessment of evidence-based school management/Transfer wissenschaftlicher Evidenz in die Praxis: Messung evidenzbasierten Schulmanagements. *Journal for Educational Research Online; Münster*, *8*(3), 14–38.

Farrell, C. C. (2015). Designing School Systems to Encourage Data Use and Instructional Improvement A Comparison of School Districts and Charter Management Organizations. *Educational Administration Quarterly*, *51*(3), 438–471.

Farrell, C. C., & Marsh, J. A. (2016). 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

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly*, *48*(1), 94–118.

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. In B. J. Fishman, W. R. Penuel, A.-R. Allen, & B. H. Cheng (Eds.), *Design-based implementation research: theories, methods, and exemplars* (pp. 136–156). New York: Teachers College, Columbia University.

Goldring, E., Grissom, J. A., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make Room Value Added Principals' Human Capital Decisions and the Emergence of Teacher Observation Data. *Educational Researcher*, *44*(2), 96–104. https://doi.org/10.3102/0013189X15575031

Grissom, J. A., Rubin, M., Goldring, E., Cannata, M., Drake, T., Neumerski, C. M., & Schuermann, P. (2017). Central Office Supports for Data-Driven Talent Management Decisions: Evidence from the Implementation of New Systems for Measuring Teacher Effectiveness. *Educational Researcher*, *46*(1), 21–32.

Hallett, T. (2010). The Myth Incarnate Recoupling Processes, Turmoil, and Inhabited Institutions in an Urban Elementary School. *American Sociological Review*, *75*(1), 52–74.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The New Instructional Leadership: Creating Data-Driven Instructional Systems in School. *Journal of School Leadership*, *17*(2), 159–194.

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using Improvement Science to Better Support Beginning Teachers: The Case of the Building a Teaching Effectiveness Network. *Journal of Teacher Education*, *66*(5), 494–508.

Hatch, T., Hill, K., & Roegman, R. (2016). Investigating the Role of Instructional Rounds in the Development of Social Networks and District-Wide Improvement. *American Educational Research Journal*, *53*(4), 1022–1053. https://doi.org/10.3102/0002831216653205

Honig, M. I., & Coburn, C. (2008). Evidence-Based Decision Making in School District Central Offices Toward a Policy and Research Agenda. *Educational Policy*, *22*(4), 578–608.

Honig, M. I., & Venkateswaran, N. (2012). School–Central Office Relationships in Evidence Use: Understanding Evidence Use as a Systems Problem. *American Journal of Education*, *118*(2), 199–222.

Horn, I. S. (2016). Accountability as a Design for Teacher Learning: Sensemaking About Mathematics and Equity in the NCLB Era. *Urban Education*, 0042085916646625.

Ingram, D., Seashore Louis, K., & Schroeder, R. (2004). Accountability policies and teacher decision maing: Barriers to the use of data to improve practice. *Teachers College Record*, *106*(6), 1258–1287.

Jerolmack, C., & Khan, S. (2014). Talk Is Cheap: Ethnography and the Attitudinal Fallacy. *Sociological Methods & Research*, *43*(2), 178–209.

Kerr, K., Marsh, J., Ikemoto, G., Darilek, H., & Barney, H. (2006). Strategies to Promote Data Use for Instructional Improvement: Actions, Outcomes, and Lessons from Three Urban Districts. *American Journal of Education*, *112*(4), 496–520.

Lachat, M. A., & Smith, S. (2005). Practices That Support Data Use in Urban High Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *10*(3), 333–349.

Langley, G. J. (2009). *The improvement guide: a practical approach to enhancing organizational performance*. San Francisco: Jossey-Bass.

LeMahieu, P. G., Grunow, A., Baker, L., Nordstrum, L. E., & Gomez, L. M. (2017). Networked Improvement Communities: the discipline of improvement science meets the power of networks. *Quality Assurance in Education*. https://doi.org/10.1108/QAE-12-2016-0084

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

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. New York: Sage.

Little, J. W. (2012). Understanding Data Use Practice among Teachers: The Contribution of Micro-Process Studies. *American Journal of Education*, *118*(2), 143–166.

Louis, K. S., Febey, K., & Schroeder, R. (2005). State-Mandated Accountability in High Schools: Teachers' Interpretations of a New Era. *Educational Evaluation and Policy Analysis*, *27*(2), 177–204. https://doi.org/10.3102/01623737027002177

Lynch, D., Smith, R., Provost, S., & Madden, J. (2016). Improving teaching capacity to increase student achievement: The key role of data interpretation by school leaders. *Journal of Educational Administration*, *54*(5), 575–592. https://doi.org/10.1108/JEA-10-2015-0092

Marsh, J. A. (2012). Interventions Promoting Educators' Use of Data: Research Insights and Gaps. *Teachers College Record*, *114*(11), 110309.

Marsh, J. A., Bush-Mecenas, S., & Hough, H. (2017). Learning From Early Adopters in the New Accountability Era: Insights From California's CORE Waiver Districts. *Educational Administration Quarterly*, *53*(3), 327–364. https://doi.org/10.1177/0013161X16688064

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from Recent RAND Research*. Santa Monica, CA: RAND Corporation.

Meyer, H.-D., & Rowan, B. (2006). Institutional analysis and the study of education. In H.-D. Meyer & B. Rowan (Eds.), *The New Institutionalism in Education* (pp. 1–14). Albany, NY: SUNY Press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *The American Journal of Sociology*, *83*(2), 340–363.

Mintrop, H., & Sunderman, G. L. (2009). Predictable Failure of Federal Sanctions-Driven Accountability for School Improvement—And Why We May Retain It Anyway. *Educational Researcher*, *38*(5), 353–364. https://doi.org/10.3102/0013189X09339055

Moss, P. A. (2012). Exploring the Macro-Micro Dynamic in Data Use Practice. *American Journal of Education*, *118*(2), 223–232. https://doi.org/10.1086/663274

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using Student-Assessment Results to Improve Instruction: Lessons From a Workshop. *Journal of Education for Students Placed at Risk (JESPAR)*, *10*(3), 269–280. https://doi.org/10.1207/s15327671espr1003_3

Palmer, D., & Snodgrass Rangel, V. (2010). High Stakes Accountability and Policy Implementation: Teacher Decision Making in Bilingual Classrooms in Texas. *Educational Policy*, *25*(4), 614–647. https://doi.org/10.1177/0895904810374848

Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: District and school connections in data-driven decision-making. *School Leadership and Management*, *29*(5), 477–494.

Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., … Allen, A.-R. (2016). *Findings from a National Study on Research Use Among School and District Leaders* (No. Technical Report No. 1). Boulder, CO: National Center for Research in Policy and Practice. Retrieved from http://ncrpp.org/assets/documents/NCRPP_Technical-Report-1.pdf

Peurach, D. J. (2016). Innovating at the Nexus of Impact and Improvement: Leading Educational Improvement Networks. *Educational Researcher*, *45*(7), 421–429.

Rowan, B. (2006). The new institutionalism and the study of educational organizations: Changing ideas for changing times. In H.-D. Meyer & B. Rowan (Eds.), *The New Institutionalism in Education* (pp. 15–32). Albany, NY: SUNY Press.

Sherer, J. Z., & Spillane, J. P. (2011). Constancy and Change in Work Practice in Schools: The Role of Organizational Routines. *Teachers College Record*, *113*(3), 611–657.

Spillane, J. P. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education*, *118*(2), 113–141.

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational Routines as Coupling Mechanisms Policy, School Administration, and the Technical Core. *American Educational Research Journal*, *48*(3), 586–619.

Stecher, B. M., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A. E., McCombs, J. S., … Naftel, S. (2008). *Pain and gain: Implementing No Child Left Behind in three states, 2004-2006*. Santa Monica, CA: RAND.

Supovitz, J. A. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, *10*(2–3), 211–227. https://doi.org/10.1007/s10833-009-9105-2

Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education.

Tichnor-Wagner, A., Wachen, J., Cannata, M., & Cohen-Vogel, L. (2017). Continuous improvement in the public school context: Understanding how educators respond to Plan-Do-Study-Act cycles. *Journal of Educational Change*.

Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, *21*(1), 1–19.

Weiss, C. H. (1977). Research for Policy's Sake: The Enlightenment Function of Social Research. *Policy Analysis*, 531–545.

Weiss, J. A. (2012). Data for Improvement, Data for Accountability. *Teachers College Record*, *114*(11), 110309.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, *19*(3), 239–259.

Yeager, D. S., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (n.d.). *Practical measurement*. Carnegie Foundation for the Advancement of Teaching.

Table 1. Data Sources

Data Sources	Design Phase January 2013 – June 2014	Development Phase August 2013 – June 2014	Implementation Phase August 2014 – June 2015	Total 34 months
Process Data				
Audio Files	62 h 22 m	66 h 21 m	61 h 50 m	186 h 33 m
Field Note Logs	28	24	19	71
Artifacts	147	236	74	457
Research Reflection Forms	6	5	5	16
Participant Feedback Forms	97	141	82	320
Interviews with district design team members	24	23	0	47
Fieldwork Data		December 2013	October 2014 – April 2015	
Interviews with school design team members		21	24 21	66

Table 2. PDSA Data Collection

	2013-14				2014-15		
	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 1	Cycle 2	Cycle 3
Desert Grove	Focus: Growth mindset lessons Data: Student mindset survey	Focus: Problem-solving lessons Data: Student problem-solving survey	Focus: Disciplinary reflection Data: Teacher feedback and student survey	Focus: Recruit early adopters Data: Teacher feedback	Focus: Growth mindset lessons and beginning implementation Data: Teacher survey	Focus: Grade reporting process Data: Student artifacts; Teacher survey	Focus: Grade reporting process Data: Passing rates; Teacher survey
Forest Glen	Same as above	Same as above	Focus: Growth mindset lessons Data: Planned to collect teacher and student surveys; Data not collected	Focus: Disciplinary reflection Data: Planned to collect teacher reflection and student survey; Data not collected	Focus: Growth mindset lessons and beginning implementation Data: Teacher and student surveys	Focus: Disciplinary reflection Data: Teacher and student surveys	Focus: Teacher mentoring of students Data: Teacher focus groups
Valley	Same as above	Same as above	Focus: Praise language Data: Unclear; Data not collected	Focus: Disciplinary reflection Data: Unclear; Data not collected	Focus: Disciplinary reflection Data: Administrative data on referrals	Focus: Goal setting process Data: Student artifacts, classroom observations	Focus: Problem-solving process Data: Teacher feedback

Note: All name are pseudonyms.