

© Copyright 2015

Kathryn Elizabeth Torres

**Ready, Fire, Aim: Instructional Leaders' Evidence Use
and Assessment within a Latino and Emergent Bilingual-serving School**

Kathryn Elizabeth Torres

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Micheal S. Knapp, Chair

Susan B. Nolen

Frances E. Contreras

Program Authorized to Offer Degree:

College of Education

Abstract

Ready, Fire, Aim: Instructional Leaders' Evidence Use and Assessment within a Latino and Emergent Bilingual-serving School

Kathryn Elizabeth Torres

Chair of Supervisory Committee:

Professor Michael S. Knapp

College of Education

Accountability systems and instructional reform efforts have relied on various forms of data to improve student outcomes. Yet, it is unclear exactly how data is being used in practice, by leaders and educators within particular contexts for instructional improvement. This case study explored forms of evidence use of an instructional leadership team during collective data-focused inquiry, within an Latino and emergent bilingual-serving school. The goal of this study was to determine how available assessment tools and forms of evidence created opportunities to learn and respond to their Latino and emergent bilingual student population. This work posits that data and assessment use is a situated and collective process, embedded within state, district, school, and professional learning community contexts. Observations of instructional leadership team and professional learning community meetings, alongside interviews

with educators, revealed that assessment data was primarily used for targeting students for intervention and support services. Their accountability context, available assessment tools, and emphasis on testing created many tensions that decreased opportunities for educators to make connections to the instructional core. Findings indicate forms of data and assessment used within data-focused inquiry has the potential to shape practitioner discourse, as well as opportunities for educators to engage with each other to develop capacity to serve their student population. Implications for research, practice, and policy are discussed.

Table of Contents

LIST OF TABLES	iv
LIST OF FIGURES	v
Chapter 1:	1
Introduction: NCLB Accountability Within The “New” Latino Diaspora	1
ELL Designation Policies and Assessments	3
Data Use In The “New” Latino Diaspora	6
Chapter Overview	8
Chapter 2:	11
The Research Problem—Data Systems and Collective Leadership	
Assessment Practices in Culturally and Linguistically Diverse Schools	11
Responding to Pressure For Data Use Under Differing Accountability	
Contexts	12
Data Systems for External Accountability	13
Data Systems for Instructional Improvement	19
A Knowledge Gap: Collective Data Use Practices In Particular Accountability	
Contexts	22
Understanding the Microprocesses of Data Use Practices At the School	
Level	23
A Focus of Inquiry: Collective Data Use Practices In Latino	
and Emergent Bilingual-Serving Schools	25
Chapter 3:	32
Mechanisms and Meanings in Data Use Practices: Informing Literatures and	
Framing Ideas	32
Organizational Routines for Data Use	34
Identifying Routines for School Improvement	36
Data Use in Data-Driven Decision-Making Routines	37
Inside Organizational Routines for Data Use: Opportunities to Learn and the	
Construction of Meaning within Communities of Practice	40
Negotiation of Meaning in Team-based Data Use Interactions	42
Boundary Objects as Discursive Spaces for Collective Learning	45
Assessment-as-Practice Within Team Discourse and the	47
Potential for Situated Collective Learning	47
Types of Assessment	48
Expanded Notions of Evidence and Assessment in Team-based Data Use	49
Collective Data Use Outcomes and their Sources within Team Discourse	51

Factors that Shape Data Use Practice and Discourse in Latino and ELL-serving Schools	52
Documentary Assessment Factors	53
Discursive Assessment Factors in Team-Based Inquiry	53
Inherent Assessment Factors	55
Conceptual Framework	57
Chapter 4:	61
Study Design and Methods: A Situated Qualitative Case Approach	61
Site Selection and Setting	64
School and District Context for Instructional Improvement	66
Study Participants	72
Data Sources and Data Collection	74
Approach to Analysis	79
Phase 1: Observing, Processing and Describing Data Use Practices	80
Phase 2: Coding, Developing Matrices, and Identifying Key Themes	81
Phase 3: Tracking Opportunities to Learn from Data Use in Practice	83
Strengths and Limitations of the Study Design	86
Chapter 5:	90
Data-Focused Instructional Leadership and Perceptions of Assessment	
“Fit” Within Collective Data-Use Practices	90
The Accountability Context and Organizational Structures for Student Improvement	91
The Crossroads Continuum of Student Support Structures	92
Testing and Assessment Routines	96
Organizational Routines for Data Use	98
Local Meanings in Data-Use Practices	102
The “Mountain Valley Way”: Use of Multiple Measures for Student Tracking and Interventions	103
“Doing” Data: Challenges to Utilizing High-Stakes and	107
Interim Summary Assessments for Decision Making	107
Perceptions of Assessment “Fit” for Interpreting	113
ELL-Student Learning	113
Summary: Context for Data Use Practices in ELL-Serving Schools	115
Chapter 6:	118
Exploring Data Use in Practice and Sources of Evidence for Instructional Improvement within Collective Data-Focused Routines	118
Focusing on Outputs: Student Data and Practitioner Inquiry	125
Student Growth as Success	125

Interpreting Student Outcomes	132
Focusing on Inputs: Aligning Expectations to Teaching and Learning.....	138
Interrogating School Expectations for ELL Student Speech	138
Horizontal and Vertical Alignment.....	141
ELL Practitioner Knowledge and Analysis	144
Summary: Sources of Information for Collective Inquiry	150
Chapter 7:	151
Discussion, Conclusions, and Implications: Expanding the Forms	
of Evidence in Data Driven Decision Making	151
Summary of Findings.....	152
What the Study Contributes to Research on Data Use Practice.....	155
Implications for Practice and Policy	164
REFERENCES	168
APPENDICES	180

LIST OF TABLES

Table 1. Crossroads Elementary continuum of ELL student supports.....	94
Table 2. PLC and staff meeting rotation schedule	101
Table 3. Local framework for organizing student levels in Crossroads Elementary.....	104
Table 4. Examples of assessment practice in data use routines.....	122
Table 5. Connections to the instructional core within data-driven decision-making.....	124

LIST OF FIGURES

Figure 1. Data-Driven Decision Making Process Framework.....	36
Figure 2. Situated Collective Data Use Practices	55

ACKNOWLEDGEMENTS

My journey to the PhD has been touched by so many amazing individuals: First, to the instructional leaders and staff in my study, whom welcomed me as one of their own, let me scribble in my notebook as they engaged in the important work of improving student outcomes, and who shared their hopes and struggles of their everyday work.

I have been blessed with so many incredible mentors over the years. Dr. Frances Contreras who saw my passion for improving outcomes for Latino and emergent bilingual children, recruited me to the University of Washington to begin my doctorate, helped me to secure funding and research opportunities throughout my time here, and who taught me the ropes of professional development. Special thanks to my advisor, mentor, and teacher Dr. Michael Knapp for letting it be OK for me to fall in love with so many ideas, but who also thoughtfully and patiently cultivated and reined in my numerous ideas for the dissertation. Thank you for believing in the power of my ideas. To the CREST Faculty Mentors, Drs. Michael Knapp, Bill Zumeta, and Bob Abbott for all of the support and opportunities to build my skills as a P-20 Mixed Method Policy researcher, and to Angie Windus for keeping the lights on! To Drs. Ann Ishimaru and Joe Lott for their mentorship within the Equitable Parent-School Collaboration project. To Dr. Michelle Young and the UCEA Community, your work continues to inspire me everyday. To my committee members, Drs. Michael Knapp, Frances Contreras, Sue Nolen, and Joaquin Herranz, thank you for all the work that you do!

I would never have made it without the support of my University of Washington community. To my writing group Drs. Nitya Venkateswaran, Jessica Salvador, Sara Stull,

and Susanna Eng, you are all so amazing and inspiring! Thank you for listening to me rattle on and on about my dissertation and giving me much needed pep talks. To my CREST cohort, it has been an amazing experience to be counted among such driven and intelligent individuals. To many research team members I have worked with over the years, I have learned so much from working with you!

Finally to my family, I would have never made it without all of your love and support throughout the many cross-country moves for my various degrees. To my husband, partner, and in-house copy editor Jack, your daily love and support kept me going through all of the ups and downs of this process, and into the final stretch of my dissertation!

DEDICATION

This dissertation is dedicated to all of the students in
Langford Elementary's 2005-2006 4th grade ELL classes, who continue to inspire me to
work to find ways to improve opportunities to learn for all students.

Chapter 1:

Introduction: NCLB Accountability Within The “New”

Latino Diaspora

During the last decade of accountability reform, the federal No Child Left Behind (NCLB) policy has expanded the role and complicated the practice of instructional leaders. It has created a focus on statewide standardized assessment and the use of data to improve systems of practice and build internal capacity, while drawing criticism for constraining teacher and student learning opportunities, particularly for Latino and *emergent bilingual* students¹—that is, those students who are still in the process of gaining academic language proficiency in English. As the population with the highest percentage of designated English Language Learners (ELLs) (Shin & Kominski, 2010), Latino emergent bilingual students have been especially vulnerable to the outcomes of state and local responses to accountability. The Latino achievement gap is especially salient due to the projected 36 percent increase in K-12 Latino students by 2019 (Hussar & Bailey, 2011) alongside the simultaneous rise in the number of ELL students.

Scholars have traced the impact of accountability generally and specifically for Latino emergent bilingual students (Contreras, 2010, 2011; Darling-Hammond, 2010; Valli, Croninger, Chambliss, Graeber, & Buese, 2008) through examination of assessment documents and classroom practices (Diamond & Spillane, 2004; Pandya,

¹ It is important to point out that a student who is no longer designated as “ELL” by federal or state language

2011; Valli, et al., 2008), and through implementation of organizational assessment practices (Diamond & Spillane, 2004; Heilig & Darling-Hammond, 2008; Vasquez Heilig, 2011). Meanwhile, research on improving student outcomes has elicited promising leadership practices and data use practices to improve decision making through cycles of inquiry around assessment and instruction (Knapp, Copland, & Swinerton, 2007; Leithwood & Riehl, 2005; Portin et al., 2009) and district partnerships and organizational learning (Honig, Copland, Rainey, Lorton, & Newton, 2010).

Parallel to this work, researchers and practitioners have endeavored to find new and innovative ways to build capacity to improve outcomes for Latino emergent bilingual students. Studies of effective leadership within Latino-serving schools have extended the school and district dynamic to include strong partnerships within Latino families and communities (Contreras & Stritikus, 2008; Gonzalez, 2010; Gonzalez, Huerta-Macias, & Tinajero, 1998; Riehl, 2005) and, in particular, highlighted best practices for teaching and assessing ELLs and Latino English learners (Abedi & Dietel, 2004; Abedi & Gándara, 2007; Aguirre-Muñoz & Solano-Flores, 2010; Collier & Thomas, 2004; Tinajero, Munter, & Araujo, 2010) and improving professional development for teachers of Latino ELL students (Minaya-Rowe & Ortiz, 2010).

Yet while developing capacity around data-driven assessment and inquiry practices is a hurdle for any organization, research on implementing NCLB

accountability within Latino and emergent bilingual serving schools² highlights additional challenges and considerations in leading for instructional improvement within these contexts. Prominent among these challenges is the focus of this dissertation: how school and district leaders in these settings interpret and use the now ubiquitous assessment data to improve education for the students they serve. Their efforts take place in a distinctive context, described in the remainder of this Introduction, that sets the stage for this research.

ELL Designation Policies and Assessments

English language learners (ELLs) are students who are generally in the process of developing proficiency in English. Federal legislation refers to limited English proficient (LEP) students as being school age; enrolled in primary or secondary school; either non-native born or living in an environment where English is not dominant; and with difficulties speaking, reading, and writing in English so that they cannot be proficient in state standardized tests and participate fully in classrooms (Hakuta, 2009). Since state standardized tests measure achievement and language, the persistent underperformance of this group may be due to the wide ranges of English proficiency in the ELL student subgroup (Abedi & Dietel, 2004).

Therefore, leadership practice can be complicated by particular state policies that affect language policies and assessment practices required by the district (e.g., how

² For the purposes of this study, Latino and emergent bilingual serving schools will be defined as schools with a critical mass of 25 percent or above of Latino and ELL students

learners of English are identified, and included in assessment), which can affect the validity of those assessments in gauging the achievement of students with varying levels of English proficiency (Duran, 2008). Part of the challenge of building capacity nationwide for ELL students lies in the fact that many states and districts have been given discretion in how to define and place their ELL students. In a 2010 study, only eight states and the District of Columbia had established aligned statewide criteria for identifying ELL students. The remaining states provided districts with the discretion for identification (U.S. Department of Education, 2012). Generally, a home language survey is used to determine potential for English Language proficiency. Once identified by the state or district criteria, the student would then remain in ELL programming until they met exit criteria.

Thus, state and district-specific ELL policies and assessments dictate entry and exit points for ELL students. These policies and assessments characterize English acquisition as a dichotomy (proficient or not) rather than as a continuum of language acquisition. For example, legislation in California and Massachusetts makes assumptions about the length of time it takes for English learners to gain proficiency—one and three years respectively (August, Goldenberg, & Rueda, 2010). Work by Hakuta, Goto-Butler, and Witt (2000) indicates that it can take up to 3 to 5 years to acquire oral language proficiency and 4 to 7 years to acquire academic language. Given that language proficiency is not the same thing as academic proficiency, long-term English language learners, in particular, demand particular scrutiny among school leaders and researchers (Hakuta, 2011). In particular, school leaders' support for their

emergent bilingual population can often be dependent on how a state (or district) defines, identifies, and transitions students in and out of ELL programming. These definitions have distinct implications for service supports and English as a Second Language (ESL) curriculum policies in response to meeting accountability sub-group goals, such as Title 1 funding for additional tutors, programming, and other student support services. Regardless of a student's placement inside or outside ELL designation and support services, emergent bilingual students—still in the process of gaining academic language proficiency in English— must have language support throughout their curricular experiences even after moving out of ELL designation.

Finally, the amount of time required to transition students from needing language support services could be mitigated by the type of programmatic approaches used to provide instruction, such as an instructional environment that reflects the benefits, rather than the deficits, of dual language acquisition, and includes parental engagement and support of the program (Collier & Thomas, 2004). Yet many districts do not yet have the professional capacity to implement dual language programs (Gándara & Hopkins, 2010; Garcia, et al., 2009). Teachers—and those who help them improve their instruction— need to learn how to offer rigorous grade-appropriate content, while at the same time ensuring that students with limited English proficiency can access that content through research-based effective practices for language acquisition. These challenges are particularly salient where increasing numbers of Latino emergent bilingual students have redefined the demographic makeup in school and district contexts that have traditionally not been home to Latinos.

Data Use In The “New” Latino Diaspora

Scholars Murillo and Villenas (Wortham, Hamann, & Murillo Jr., 2001, 2015) have referred to the recent growth of Latinos in the South and other non-traditional areas of Latino settlement as the “New Latino Diaspora.” This term denotes the many Latino immigrant populations settling temporarily and permanently in areas outside of the “traditional” Latino settlements.³ These new areas include North Carolina, Maine, Georgia, Indiana, Arkansas, and other rural areas. Hamann and Harklau (2010, 2015) argue that these newcomer Latinos are often faced with improvisational educational responses—particularly in response to language learning needs—whereby Latino students are subject to “pull out” programs such as Structured English Immersion (SEI), and are more likely to be taught by less credentialed teachers who are unfamiliar and untrained in working with this new student population. This study sits within this “improvisational” school context. It seeks to better understand how practitioners and leaders new to serving in culturally and linguistically diverse schools navigate and negotiate potential conflicting messages between accountability and ELL policies.

This research has the potential to inform how aspects of NCLB accountability policies impact practitioner learning that can lead to stagnant accountability gains. Specifically, it is important to critically examine school improvement work within Latino and emergent bilingual serving schools. Many studies lump together Latino students with other students of color, making any analysis of the impact of high-stakes

³ i.e., Arizona, California, Colorado, Florida, Illinois, New Jersey, New Mexico, New York, and Texas

accountability on ELL students an afterthought. Furthermore, practitioners within the field need better tools and understanding of how to apply effective leadership practices for Latino students, along with accountability reform policies that stress the use of data for decision making. Although these data-driven approaches hold promise for improving student outcomes, we are only beginning to understand how these ongoing interactions around accountability practices impact school leaders' motivation and capacity to empower practitioners to serve their students and make complex decisions in a situated context (Moss, 2012). In particular, we need to know more about how and in what ways state and district policies, staff knowledge and access to professional development, and available tools for data-driven decision-making impact collective instructional leadership in Latino and emergent bilingual serving schools.

Thus, in this study, I contribute to our understanding of collective instructional leadership practices by examining the extent to which school leaders consider and navigate their embedded context when collectively utilizing student data to improve student outcomes within a predominantly Latino and emergent bilingual serving school. Through this research, I aimed to understand how accountability, alongside ELL curriculum, designation, and assessment policies, influences how school leaders and staff interact and make sense of available data for decision making. For districts within states that have seen swift growth among their Latino student body, tools to implement accountability policy while building organizational capacity for culturally and linguistically diverse schools cannot come too soon. For that reason, this study of assessment practices of instructional leadership teams within Latino and emergent

bilingual serving elementary schools is an important first step in building integrated theory around organizational learning for cultural and linguistic capacity building.

Chapter Overview

In Chapter 2, I identify the research problem in more detail by examining how NCLB created data systems for accountability and school improvement purposes, and considering how the use of these data systems is challenged within Latino and emergent bilingual serving schools. I review the extant research on how instructional leaders have used data systems for instructional improvement. I also review the sources of data and factors that influence engagement with data for decision making. I then draw from emergent research on situated and interpretative data use practices to consider the extent to which collective data use practices can consider and be influenced by their surrounding contexts. The chapter concludes with a specification of the focus of inquiry and a set of research questions that my study will answer.

Chapter 3 offers a way to conceptualize in theoretical terms what I am investigating. Here, I delve deeper into the mechanisms and interactions within data use practices by highlighting the ostensive and performative aspects of data use routines. I then draw from organizational and sociocultural theories of data-driven decision making and practitioner learning to integrate opportunities for practitioner learning and sensemaking within data use micro-processes, as well as considering the influences of various sources of data and evidence in data use. I then review a variety of factors that impact the forms of assessment present in data use practice and present my integrated

conceptual framework of situated social learning and decision making within data use processes that outlines the influences of multiple contextual influences during collective data use discourse.

In Chapter 4, I present a rationale for a situated qualitative case study and methodological approach, focusing on my case selection, their situated context for educational reform and my approach to collection and data analysis. The chapter goes on to discuss in detail the design choices I made, and their suitability for pursuing the research questions identified in Chapter 2.

In Chapter 5, I offer the results of a first round of data analysis. The chapter describes the context of accountability, data use routines, and instructional reform within my case, and considers how practitioners interpreted and made sense of their reform context during the use of data for decision making. Particularly, I highlight identified problems of practice and methods of tracking and labeling students for interventions, while outlining key tensions identified by staff in their use of available tools and improving outcomes for struggling students—particularly Latino emergent bilinguals.

Chapter 6 offers a deeper dive into the data, illuminating the role of data evidence use in practitioner knowledge generation and decision making. I do so by examining patterns of engagement in episodes of “data talk” within a variety of team and school meeting settings. I examine the extent to which different patterns and configurations of individuals yield different outcomes for decision making and shared learning, and consider the role of available tools in the mediation of collective discourse surrounding problems of practice.

Chapter 7 provides a summary of findings and considers the implications of this study for research on data use and for practice within Latino and emergent bilingual serving schools. This chapter concludes by outlining the promise and limitations of this work for research and practice.

Chapter 2:

The Research Problem—Data Systems and Collective Leadership Assessment Practices in Culturally and Linguistically Diverse Schools

No Child Left Behind (NCLB) created opportunities for increased equity through policies that stressed high standards for all students and increased transparency for assessments, practice, and student subgroup outcomes. Central to this policy shift was a focus on building data systems to inform local decision-making to improve instruction. The data systems centrally feature the results of annual state assessments of performance in high-priority subjects (reading/literacy, mathematics), though other data are also sometimes included. The systems operate district-wide, providing administrators and teachers at the school level with access to data warehouses (some linked with other data systems, others not) that report on the progress of every student subgroup in meeting academic proficiency, while providing additional student-focused information (i.e. student demographics, attendance, student grades, course enrollment, etc.) (Means, Padilla, & Gallagher, 2010).

The system of accountability in public education puts pressure on local educators to respond in ways that meet external expectations for the outcomes of education, while also motivating and supporting the improvement of practice within schools. Understanding how educators manage this response under different accountability expectations reveals some important and unexplored aspects of educational reform, especially as it pertains to particular historically marginalized student populations, and

especially for emergent bilingual students in schools that serve high concentrations of such students. Within this area lies the knowledge gap my study will address, inside of the *collective data use practices* in which the leaders of schools serving emergent bilingual students engage.

Responding to Pressure For Data Use Under Differing Accountability Contexts

Local, state, and national educational agencies implemented these data systems with the assumption that strict sanctions and accountability for administrators and practitioners, and public reporting of student test scores will motivate practitioners to improve student outcomes (Elmore, 2000; Heilig & Darling-Hammond, 2008). It is also assumed that local school leaders would have the capacity to use these data systems to facilitate and support the organizational change needed to achieve accountability targets (Finnigan, 2010). Weiss (2012) argues that—while both uses of data systems (external accountability and instructional improvement) emphasize data as a core component for educational change—the theories of action behind these data systems are distinct. Examining the differences between data systems designed to hold educators accountable to external stakeholders *versus* data systems designed for educators to make informed decisions to improve instruction has implications for the ways in which educators respond to the call to “use data” for instructional improvement.

Data Systems for External Accountability

Systems that utilize data for external accountability monitor student progress by standardizing student achievement in order to measure and compare across differing educational contexts (Weiss, 2012). This data is then used to provide the basis for holding practitioners and schools accountable for persistent achievement gaps. State standardized student data can be a powerful tool for equity as stakeholders (such as parents) can use this data as ammunition to advocate for educational change (Ishimaru, 2014). Yet, the availability of data cannot guarantee its use by practitioners, nor can the use of data for accountability purposes presume that unfortunate unintended consequences will be minimal (Fuhrman, 2003). Furthermore, such data concentrate on selected outcomes of instruction, not the nature of instruction itself. Therefore, providing educators with information about the outcomes of their interventions without support to identify aspects of practice that would influence these outcomes can be problematic.

Most forms of standardized testing tend to not be very helpful for educators working to improve student outcomes (Supovitz, 2012). Standardized student achievement data only reveals the levels of performance, and hence the existence and extent of the achievement gap for student subgroups, but often fails to offer guidance as to its causes, or remedies (Linn, 2005). Educational leaders are often tasked with meeting a myriad of student sub-group achievement goals (e.g. Special Education, English Language Learners, African American, and so forth) wherein many students are cross-classified (i.e., “counted” in more than one category), which can oversimplify the learning needs of culturally, linguistically, and economically diverse students. Big

questions remain unanswered by these data systems concerning the different ways that these various sub-groups might best be served, and what improvements in instruction would be most likely to produce the desired outcomes across all.

Issues of Validity for ELL Students. Assumptions behind data used for accountability purposes often neglects to consider whether *all* students' performance is adequately and accurately measured by state-standardized assessments. This situation is especially problematic for certain sub-groups in the student population. This calls into question the validity of using standardized assessments in English for students currently or previously designated as English language learners (ELLs) (Gándara & Baca, 2008; Menken, 2008; Solorzano, 2008), particularly for schools that have students with varying levels of English proficiency (Duran, 2008).

Standardized assessments with ELL students for external accountability purposes suffer threats to external validity. Examination of district data by Abedi, Leon, and Mirocha (2000/2005) found that language proficiency is associated with performance on content-based assessments, a performance gap exists between ELL students and their Non-ELL peers, and that language background may be a source of measurement error. Additionally, Thompson et al. (2002) has suggested that even when scores do improve for ELL students, the achievement gap is not actually diminished. In short—while ELL student performance on English is mediated by their English proficiency—there can be issues in how data is reported; such as the misuse or misinterpretation of percentile rank scores, the averaging of data across years and grades, and the aggregation of all levels of

ELL proficiency for external reporting and comparison within and across schools and districts.

Yet, these discrepancies in student performance and reporting are difficult to compare across states, districts, and schools. As mentioned previously, while state student learning objectives have been standardized across districts, the identification and classification of ELL students across states and districts is not (U.S. Department of Education, 2012). This lack of standardization of district defined entry and exit points for language proficiency status creates a false dichotomy of oral language proficiency (either proficient or not proficient), which complicates school leaders' understandings of the necessary supports for ELL students with varying levels of English proficiency (Hakuta, Goto-Butler, & Witt, 2000). Additionally, the implementation of language programs can also vary across states, districts, schools, and classrooms (August & Hakuta, 1997). This research suggest a potential lack of "fit" between standardized assessments and the needs of learners in schools serving culturally and linguistically diverse students. Despite these issues of reliability and lack of ELL identification standardization, educators and their emergent bilingual students remain accountable to their performance in the terms that the externally focused accountability systems set up.

Therefore schools that serve large proportions of ELL students are likely struggling to demonstrate adequate yearly progress (AYP) for one or more student subgroup. For schools that consistently fail to meet AYP, the consequences under NCLB vary in severity by the number of years the school has been labeled as "in need of improvement." When schools progress through the federal school improvement

timeline, every year “in need of improvement” potentially involves increased restrictions, sanctions, and eventually school restructuring from the local education authority. These sanction policies within high stakes accountability attempt to influence individual classroom practice by increasing the degree of oversight within struggling schools. However, the degree and type of oversight within a struggling school may incentivize school leaders and practitioners to focus narrowly on short-term or surface goals (such as a test score) instead of finding ways to alter school culture and instructional practices that would have longer term benefits to students and practitioners (Diamond & Spillane, 2004; Weiss, 2012). Data systems that reward finding ways to improve test-scores—rather than building overall school capacity—has led to unintended outcomes for curriculum and instruction, such as “teaching to the test” within culturally (Diamond & Spillane, 2004) and linguistically diverse schools (Menken, 2010; Pandya, 2011).

Challenges to data system use within Latino & ELL-serving schools. When systems emphasize data for external accountability purposes and create high-stakes pressure to make adequate yearly progress for student subgroups, educators serving Latino and ELL students may easily lose sight of what it means to serve these students’ learning needs effectively. For example, in one study the result of failed interventions within Latino ELL classrooms lead to an overall “testing-as-learning” culture in order to avoid strict sanctions and school closure (Pandya, 2011), or—in another case—created a “new definition of ESL” to more closely resemble English Language Arts coursework, in order to prepare ELL students to pass a high-stakes exam (Menken, 2006). Latino

students, in general, have markedly lower achievement on standardized tests, participation in college preparatory coursework, high school graduation, and college enrollment and completion than their White and non-White peers (Gándara & Contreras, 2009). Despite our focus on achievement for minority groups, reading and math achievement gap has remained relatively unchanged (Lee & Wong, 2004). In addition, the increased transparency of educational disparities has not motivated states to address issues of educational inequity within their districts (Kantor & Lowe, 2006). The persistent achievement gap and lack of responsiveness to serve this population may be due the many issues leaders face in understanding how to serve the unique learning needs of ELL students coupled with educational barriers already experienced by the Latino population.

It is often difficult for educators to understand the complex sources of the performance gap between ELL and non-ELL students—caused by factors such as parental education, poverty, learning a new language, and inequitable learning opportunities—that can create a significantly lower baseline of performance for ELL students (Abedi & Gándara, 2007). Moreover, within schools without a coherent language policy within the school to guide implementation of data-focused interventions for ELL populations, many school leaders may find themselves creating *de facto* English-only language policies trying to prepare ELL students for high-stakes testing (Menken, 2006). Without clear direction, incentives and support within accountability policies to improve teacher and leadership preparation for this population, leaders within Latino and emergent bilingual serving schools face will continue to face distinct

contextual challenges in providing equitable learning environments within high-stakes testing environments. Therefore the combination of school setting, Latino and ELL student outcomes, and subsequent accountability pressure to improve may create a myopic view of immediate and short-term goal setting and solutions that ignore the capacity building needed by staff to effectively serve these students (Hargreaves & Shirley, 2009).

Just as for all students, context matters; yet instructional leaders working within culturally and linguistically diverse schools must also contend with the potential deleterious impact of accountability mandates on school language policies and curriculum that arise within implementation. While standardized test scores will continue to be a driving force of instructional reform, how school leaders use test-score data to gauge performance and influence instructional practices is an important aspect of how accountability and ELL policies are understood by the staff and are leveraged into reform (Spillane et al., 2002). Yet—despite system-wide data systems that incentivize performance over improvement, and standardized student data that emphasizes outcomes rather than the antecedents of achievement gaps—not all outcomes of external accountability have led to unintended consequences. Many school leaders have begun to internalize these higher standards and have used data as an opportunity to challenge expectations for student performance and support local interventions (Elmore, 2000). As a result, some educational leaders have begun to collect their own data to build local data systems (often specific to a given school) focused on supporting—rather than hindering—instructional improvement.

Data Systems for Instructional Improvement

Given the complexities of standardized student outcome data, a central problem of practice confronting educators within Latino and emergent bilingual serving schools is the challenge of making informed decisions about instructional improvement based on available assessment data and whatever other data, interpretive lenses, and expertise in serving linguistic diversity they can bring to bear on the problem. In addition, their decision making is best served by assessments that “fit” the needs of their students and it is possible that not all forms of data are useful or appropriate. To make these decisions, educators are often participating in data systems that blend external accountability information with other sources in an attempt to address the specific teaching and learning situation they face. Examining systems in such instances that use *data for improvement* provides insight into the assessment processes and practices used by educators and school leaders for improvement.

Other than providing a “snapshot” or a lens through which to view students and teacher work, educators must be able to use data as a tool to diagnose issues and causes of student achievement, monitor goals and process, guide instruction and organizational change, and legitimize or justify decisions made in the interest of instructional improvement (Jennings, 2012). In essence, it is not only important to have data, but also to be able to use data effectively for instructional improvement. Studies of effective uses of data explore how principals and other formal school leaders work to build team-orientated cultures focused on the utilization of data for instructional improvement

priorities (Portin et al, 2009, others). They create organizational routines and practices for data-informed inquiry and decision making (Knapp, Copland, & Swinnerton, 2007; Leithwood & Riehl, 2005; Portin et al., 2009; Wayman & Stringfield, 2006), and create spaces for educators to engage with particular tools (e.g., teacher evaluation tools, Halverson & Clifford, 2006). These opportunities to engage with and learn from data are often on-going, smaller cycles of inquiry and informal evaluation within the larger evaluation cycle (Portin et al., 2009), which suggests that student test scores and other data relevant to student learning can be used for a variety of purposes by different educational actors at different times.

Sources of information for instructional improvement. Along with standardized data, instructional leaders marshal additional material and conceptual resources as they take information about student performance to action. Often administrators, instructional coaches, and teacher leaders of several kinds, these individuals exercise leadership in the school's efforts to substantially improve the teaching and learning of educators. Regularly working as a leadership team, these individuals are positioned to interpret the externally provided assessment data, bring other kinds of information and interpretative frames to the task, and attempt to fashion improvement strategies that keep the needs of the school's student population in focus, all while also responding appropriate to external accountability demands.

These individuals help to identify the on-going needs of the school and series of decisions needed for school improvement. Moss (2013) highlights an important dimension of school data use— time scale and grain size—and suggests that the types of

data instructional leaders employ are contingent on the type of large or small-scale decisions that are needed for on-going school improvement, such as changing instructional strategies mid-year versus deciding how to allocate school funds. Therefore, she argues that depending on the type of decision needed, data used by instructional leaders and their teams can take a variety of forms:

[Sources of evidence can be] formal and informal, quantitative and qualitative—to which educators have access or that they can develop themselves. These range from the information available in ongoing classroom interactions and samples of students' work; teacher accounts of classroom practice, instructional artifacts, and discussions of standardized test results; data from videotapes, interviews, and surveys; various indicators of resources and social structures; and published research reports. (p.93)

However, creating systems and practices for the exploration of data must also include adequate supports for educators to develop actionable knowledge for improvement.

Factors that influence engagement with data for improvement purposes.

Although armed with multiple sources of data and opportunities to collectively engage with data-focused inquiry, educators still struggle to effectively work with data for improvement purposes. In their investigation of educational leaders, Anderson, Leithwood, and Straus (2012) found that overall principals acted more as enablers of data use than users themselves, and that only in a minority of settings were principals and teachers going beyond using data to identify issues towards active problem solving. This is reminiscent of Temperley's (2008) identification of the "activity trap" in data

use, wherein educators in engaging with data became “stuck in activity traps in which examining data and having conversations was seen as a good thing to do with only a vaguely defined purpose for doing so” (p. 69). Therefore, without effective supports, educators may interact with data, but may not utilize its intended uses and interpretations.

In order to combat ineffective uses of data, schools and districts have provided a variety of supports to help build the capacity of educators to work with data. In her exploration of efforts to support data use, Marsh’s (2012) exploration of Data Driven Decision Making (DDDM) processes found that effective data-support intervention programs, activities, and materials included “human support” (i.e., data coach, training and professional development), time for social interactions with peers surrounding interpretation of data, and by creating “safe spaces” for teachers to engage in criticism of instructional practices rather than individual teachers. In addition to these factors, Ikemoto and Marsh (2007) also suggest that aspects of the data itself—such as timeliness, accessibility, and perceived validity of the data—also impact educators’ use of data.

A Knowledge Gap: Collective Data Use Practices

In Particular Accountability Contexts

While research on data-support interventions suggests important aspects supporting school leaders and educators in data use, it raises questions about how the various configurations of data user capacity, time, tools, support for data use and interpersonal dynamics interact within collective data use practices, particularly as they

engage with contradictory, confusing, and/or misleading forms of standardized data. Therefore, in order to understand how well intentioned data use can lead to unintended consequences (and also how to avoid that possibility), we must explore how and in what ways educators collectively assess, interpret, and respond to available data at hand. This becomes especially important to do in settings such as schools serving emergent bilingual students, where leaders face contradictory pressures and where the assessment date may not be well aligned with students' unique learning needs. Doing so means taking a closer look at data use practices in these settings than scholarship has yet attempted.

Understanding the Microprocesses of Data Use Practices

At the School Level

Deeper examination of data use practices in action may provide insight into the variety of processes, conditions, and contexts that impact data use in schools. Emerging research on data use practices conceptualize data use as a largely situated, interpretive process (Marsh & Farrel, 2014; Spillane, 2012; Coburn & Turner, 2011). Therein, these practices are impacted by a variety of internal (i.e., environmental, organizational, and group) contexts and broader institutional systems of meaning (i.e., classifications, categories, designations) (Coburn & Turner, 2011). Little (2012) situates the practice of data use as a series of microprocesses:

[Microprocesses] "...attend systematically to the ways in which interaction is meaningfully situated, shaped by and constitutive of organizational structures,

norms, and resources (the context of particular schools or districts, for example) as well as broader institutional and societal structures, processes, and logics (common arrangements for and ideas about education). (p.145)

These series of tasks or microprocesses in data use practices may provide clues to unpacking the variety of situated influences to data use in schools. For example, Halverson & Clifford (2006) investigated how assessment tools—in this case teacher evaluation tools—were used in practice by tracking how leadership tasks flowed through a complex educational system. This process allowed them to analyze how the enactment of key tasks revealed the relevant structural supports, or artifacts, that supported leadership practice. Put another way, one could unpack a macro process of instructional leadership such as “leading through data” by examining a series of micro tasks in assessment practice, such as “examining 4th grade science assessments by race, free/reduced lunch, level of English proficiency, and migrant status.” Yet, we are only beginning to understand how these micro-processes impact how educators make complex decisions in a situated context (Moss, 2012).

Thus far, there is little research focused on examining the aspects of discourse and activity in data-focused decision making practices. Little’s (2012) review of microprocess studies in data use provide examples of the influences of instructional leaders’ facilitation or moves in the engagement with data and the co-construction of evidence for student performance, as well as highlighting the importance of the clarity of purpose in data use. Similarly, Lasky et al. (2008) found that nascent data users tended to primarily focus on the processes and procedures of using data rather than what the

data means in context. Rainey's (2014) investigation of data use in decision making found that conditions of uncertainty caused by lack of available data provided educators the opportunity to draw from personal knowledge to add as additional evidence for instructional decision making.

In summary, this research provides important insight into the ways in which organizational, group, and individual contexts can influence interpretation of data in situated collective practices. Moreover, they provide examples of a variety of interaction analysis that sit at the intersection of institutional, organizational, and interpretive traditions (Little, 2012). Yet, much of the research on data use continues to make the assumption that the forms of data that instructional leaders employ is suited to the diagnostic and assessment needs of all of their students. This assumption neglects to consider the impact of "fit" of available assessment tools, and its impact on the outcomes of data use, particularly in emergent bilingual-serving schools. Furthermore, many investigations of data-focused reform has been primarily located in urban settings, which may have more developed systems of data and access to innovative professional learning than their rural counterparts.

A Focus of Inquiry: Collective Data Use Practices In Latino and Emerging Bilingual Serving Schools

Despite the apparent promise of local, improvement focused data-use practices, and the potentially helpful role that leadership teams may play in promoting and sustaining such practices in a school, there is no guarantee that their efforts will help the

school staff stay focused on what matters most for the linguistically diverse population they are serving, nor that they will avoid the downsides of data systems aimed at external accountability. As mentioned previously, data designed for accountability purposes creates additional challenges for school leaders stemming from misaligned assessment and ELL policies. As a result of this lack of “fit” and the enhanced accountability pressures due to low performing ELL students, focusing predominantly on student sub-group assessment output scores runs the risk of discursively focusing on these students as “numerical liabilities” (Koyama & Menken, 2013), rather than on interventions for school improvement. Examining collective data use practices by leaders within Latino and ELL serving schools⁴ provides opportunities to explore mechanisms that lead them to the design of instructional improvement interventions, as well as the meanings assessments bring to their understanding of student learning for emergent bilingual students.

In particular, exploring the mechanisms by which educators make meaning during collective data use practices—as well as the actual meanings they do make—in this particular context is important to better understand how they will try to serve an emergent bilingual student population. In particular, this examination may reveal how the external and organizational environments of Latino and ELL-serving schools impact the social constructions of “evidence,” “data,” “achievement,” and even “ELL students” in ways that may lead to unintended outcomes for this student subgroup. It is also

⁴ I use “ELL-serving school” and “emergent bilingual-serving school” interchangeably throughout this dissertation.

important to examine what types of interpretive or explanatory worldviews or frameworks educators use to make sense of low or stagnant student achievement, especially when student achievement data does not support or reflect their understandings of effective practice and effort. This study of educator data use discourse in emergent bilingual serving schools builds on the work of Koyama and Menken (2013) to examine how “emergent bilinguals” are quantified as ELLs into school and district data, that are then implicated into “what works” (or doesn’t) within their school context (p.96). Essentially, emergent bilingual students can be labeled as ELL and seen as a monolithic group of student assessment performance, rather than students with individual learning needs.

This quantification of emergent bilingual students as test scores and subsequent categorization of these students is similar to what Goodwin (1994) describes as “coding schemes” or the informal categorization of students in relation to teachers’ perceptions of their work. Horn (2007) argues that these coding schemes or categorization of students employed by teachers “model problems of practice simultaneously” while “communicating beliefs about students, subject matter, and teacher” (p.42), which have implications for the types of instructional decisions made for students, and their subsequent academic outcomes. If Latino and ELL students are seen by educators as “numerical liabilities” (Koyama & Menken, 2013), the potential for reinforcing negative belief systems by focusing on student or community “deficiencies” rather than on instructional and pedagogical improvements may be high. In essence, the act of labeling

and blaming categories of students as “the problem” may shift the locus of control away from educators’ agency.

Thus, my study focuses on the collective data and assessment practices of an instructional leadership team within data use practices. I explore how the team members and other educators collectively make sense of available data within organizational routines to inform instructional improvement, within a Latino and emergent-bilingual serving school. I explore these issues through an analysis of the leaders’ discourse surrounding data, within organizational routines devoted to considering data for improvement purposes. In this context, I pay particular attention to the tools the participants use, and additional information they bring that shapes their decisions about how to improve instruction for their students. The following questions guide my study:

1. In what ways, and how well, do current assessment tools in high-accountability contexts help to inform educators in an instructional leadership team about the learning needs of an emergent bilingual school population?
 - a. To what extent, if at all, do educators’ perceptions of existing assessment tools map unto their current understanding of student needs (i.e., “fit” for their student population)?
 - b. To what extent does this “fit” or lack of fit lead to contradictory or confusing evidence about emergent bilingual students during ILT assessment practices?

2. How, then, do the team members and educators collectively interpret and use this evidence to address or design their efforts in instructional improvement?
3. In what ways, if at all, are the assessment tools themselves shaping the collective thinking and practice for these educators?
4. In what ways does the external accountability system and language policy context constrain or shape their team members' interpretations or use of data for improvement purposes?

Synopsis of and Rationale for the Research

To answer these questions, I turn to research within data driven decision making (DDDM) that explores the mechanisms of how school leaders identify and utilize forms of data to make informed decisions for school improvement. These data use practices are often found within organizational routines that utilize team-based inquiry for school improvement. Therein, the types of interactions educators have surrounding forms of data, evidence, and inquiry for school improvement can be impacted by a variety of external and internal organizational factors. I utilize the theoretical framing of ostensive and performative aspects of organizational routines as a way to differentiate between the idealized and abstract understandings of data use practices to explore what data use actually looks like “in practice in particular places at particular times” (Spillane, 2012, p.115). Then I identify sociocultural research on professional learning which investigates the meanings educators—and the tools they carry—bring into collective data use practices. I will particularly explore how research within professional learning

communities considers how participation in the “joint work” of data-informed decision making provides opportunities to learn for school improvement. Together, both strands of literature work to light on the social construction of evidence for decision making, and consider the degree to which assessment tools have the potential to become powerful actors within assessment practice. These analytical frames provide an opportunity to delve more deeply into the complex interplay between policy, practice, and assessment actors within Latino and emergent bilingual serving schools.

For this study, I employ theoretical concepts from assessment-as-practice, communities of practice and DDDM theories to suggest a different paradigm for collective learning and inquiry of data use in practice. I unpack traditional conceptualizations of organizational routines for data-use practice in four interpretive and inter-related dimensions: 1) the use of inherent, discursive, and documentary *assessments*; 2) the establishment of routine discursive *spaces* for collective inquiry; 3) the *processes and practices* that create conceptual change through collective discourse; and 4) the *affordances and constraints* created by access to information, participation, and positioning of individuals that impacts leadership team inquiry. This framework is used as a lens to explore one aspect of collective instructional leadership practices—accountability discourse and collaborative decision-making strategies— in order to understand what school leaders and educators within Latino and ELL-serving schools learned about the effectiveness of their instructional improvement strategies through collective data use practices.

I use case study data of an instructional leadership team to provide an analysis of collective discourse and problem framing utilized in data use practices within a predominately Latino and emergent bilingual serving rural elementary school in the Pacific Northwest. I explore how this framework can be used to describe how collective discourses surrounding accountability and assessment for Latino and ELL elementary students often problematize and, yet sometimes overlook the learning needs of their linguistically and cultural diverse student body in the face of external organizational accountability pressures. Finally, I explore how this framework may be used by leaders to consider how their use of available tools and access to resources (human, training, etc.) could provide affordances and constraints to responding to student learning needs in schools and districts facing rapid demographic change.

I envision that my approach to the investigation of microprocesses in data use will help provide analytical clarity to the various organizational, group, and individual contexts that influence data use practices and social constructions of evidence, achievement and possibly even students themselves. Finally, this work also has implications for decision-making when available forms of data are not sufficient or appropriate for external and internal accountability purposes within culturally and linguistically diverse schools.

Chapter 3:

Mechanisms and Meanings in Data Use Practices:

Informing Literatures and Framing Ideas

Educators working within changing cultural and linguistic demographic schools and communities need more opportunities to learn and adapt their practices to serve student learning needs. This requires that educators have access to a myriad of effective tools and training in order to properly assess student learning and progress, and provide responsive practice. In response to these changing contexts, routine collective data use practices to foster school improvement have intensified under NCLB (Heilig & Darling-Hammond, 2008), particularly for low or under-performing schools. These collective data practices require that educators interpret and use assessment tools and other forms of student data in order to craft coherence between external demands of accountability policies and mandates and their school's internal goals and needs (Honig & Hatch, 2004). As a result, schools and districts have created organizational routines surrounding "data use" or "data-informed decision-making" to help inform and guide their own curricular interventions that are responsive to external demands of accountability and local school and community needs.

Yet, policies that support the use of data-informed decision making for school improvement often ignore the interaction of key aspects of data use that can create changes to practice. This is due, in part, to the underlying assumptions that educators "use" data from data warehouses to inform and guide their own curricular interventions,

with little support or guidance surrounding how and in what ways data ought to be used. These assumptions border on “technological determinism” wherein the “effects” (i.e. improved student outcomes) are assumed to originate from the availability of the data itself rather than what is ultimately a “people problem” (Wayman & Cho, 2013, pp.4). That is, it is not just the availability of multiple forms of assessment and data that drive school change, but the ways in which educators engage in inquiry surrounding persistent problems of practice and the subsequent outcomes of shared meaning and decision making for school improvement. However, as mentioned previously, changes to classroom practice may be dependent on the extent to which educators feel that they have the agency and flexibility within their school’s accountability context to explore additional definitions of academic success beyond externally defined standards such as a “4th grade Math proficiency.” Thus the impact of increased availability of multiple sources of data for decision making depends on the mechanisms and meanings that educators craft throughout their uses of data-focused decision making for instructional improvement.

In this chapter, I highlight interactive dynamics at play within data use practices by utilizing theories of learning and change that arise through collective engagement with problem solving through various forms of data and evidence. Building on prior conceptual work that combines strands of sociocultural and organizational learning theories (Honig, 2008), I argue that a more complete conceptualization of data use in practice in leadership teams integrates our understandings of organizational mechanisms of change with interactive and situated perspectives of practitioner learning. Employing

idealized and performative aspects of data use allows us to analytically locate and track meanings derived from the joint work of data-informed decision-making within culturally and linguistically diverse schools.

Drawing on these literatures, the chapter presents a framework that integrates understandings of data-driven decision making (*DDDM processes*) (Ikemoto & Marsh, 2007), aspects of *communities of practice* that shape sense making (Wenger, 1998; Star & Griesmer, 1989; Nolen, Horn, Ward, & Childers, 2011), and the affordances and constraints to data use practices within the conceptual frame of *assessment-as-practice* theory (Jordan & Putz, 2004). This unified framework provides tools to explore how the interaction of individual interpretations and understandings, collective social discourse around organizational functioning, and the conceptualization of learning espoused in documentary assessments *shape* the daily practice and decisions of school leaders and teachers. This is followed by a short summary of team-, school-, and district-level factors that have been found to impact data use practices. Finally, I elaborate how this framework informs the research questions.

Organizational Routines for Data Use

Conceptualizations of data use in complex educational systems—such as in culturally and linguistically diverse schools—must integrate our understandings of how data use ought to be and what data use actually looks like in particular places and situations. However, conceptual consensus of what “data use” or “data informed decision making” looks like in practice remains varied and draws from a variety of

organizational and socio-cultural conceptual frameworks and tools. To draw together these different theoretical strands, I utilize Spillane's (2012) framing of data use in practice to show how these theories help to clarify how micro-interactions of data use lead to particular decision and shared learning outcomes.

Spillane draws on both organizational and socio-cultural theoretical traditions to attend to the *ostensive* (idealized and abstracted) and *performative* (in actual practice, in situated contexts) aspects of organizational routines. He argues that considering the extent to which formal organizational structures enable and constrain interactions surrounding data use allows us to examine the relationships between school-level reform efforts and their larger context, and how these relationships shape micro-interactions within these routines. Organizational routines that utilize data may differ depending on the configurations of individuals (i.e., leadership team, grade-level teams, coaches with classroom teachers), goals and purposes (i.e., evaluating after-school programs, checking student progress, teacher evaluations), and educators' notions or understanding of the uses of data and how it informs aspects of their practice (Cho & Wayman, 2014). For example, tracking instances of "data talk" within weekly leadership team meetings geared for the evaluation of programs versus examining efforts to make sense of data within mandated grade-level PLC-time (Professional Learning Communities) provides two very different contexts and potentially different outcomes of data use. In essence, while the forms of evidence remain relatively stable (i.e., a test score remains the same across contexts), the ways in which that evidence is used and under what circumstances provides different opportunities for data focused interaction and ultimately

organizational change. Therefore, differentiating the ostensive and performative conceptual tools in data use practices allows us to focus on the mechanisms of data use practices and the opportunities for meaning making within situated data use in practice.

Identifying Routines for School Improvement

Organizational routines for data use have evolved alongside our understanding of collective instructional leadership practices for school improvement. In response to the complex nature of the task at hand, conceptions of instructional leadership have moved away from focusing on best practices of singular leaders and have come to the conclusion that one leader alone cannot accomplish the task of systemic instructional reform (Hallinger, 2005; Honig et al., 2010; Honig, 2008; Honig & Louis, 2007). Instead, research on instructional leadership has evolved into focusing on the "shared work and commitments that provide direction for instructional improvement, and that engage the efforts and energy of teachers and others in pursuit of powerful, equitable interactions among teachers, learners, and content, in response to environmental demands" (Knapp, 2012, p.8). As additional forms of student data and information became readily available through data warehouses and technology, organizational routines utilizing data became central to this "shared work" for instructional improvement.

Organizational routines—or "repetitive, recognizable pattern[s] of interdependent actions, involving multiple actors" (Feldman & Pentland, 2003 p.311)—for school improvement and instructional planning often utilize or directly address forms

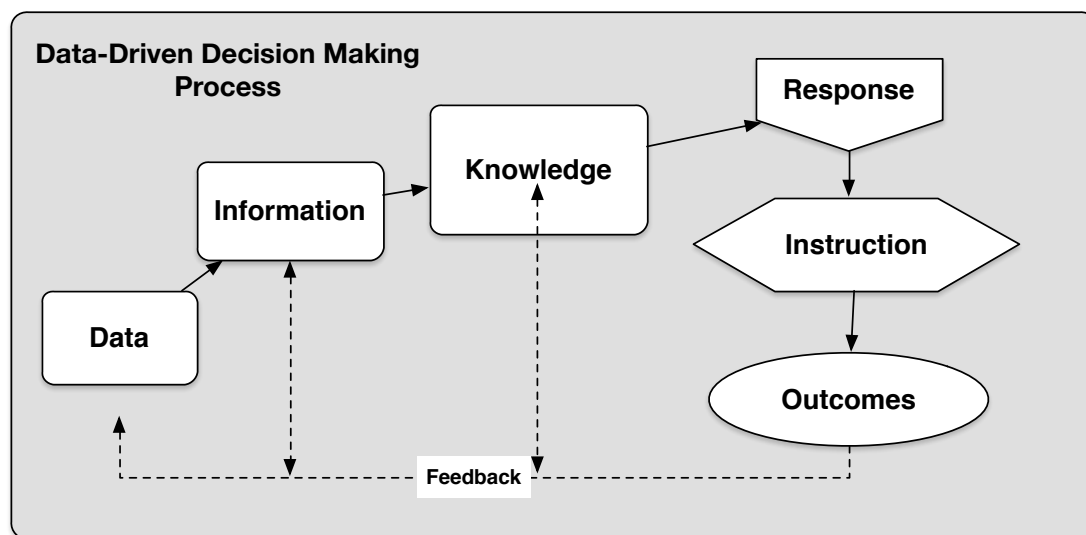
of data regularly throughout the year. Elmore (2012) argues that, once institutionalized, organizational routines have the ability to store and legitimize organizational learning, and have the possibility to either disrupt and transform the status quo or maintain it through surface-level performance (e.g., going through the motions to maintain legitimacy). As part of these organizational routines, data—in its many forms—potentially plays a powerful role in organizational learning and change. Therefore it is important to better understand how data is used within these routines for instructional improvement.

Data Use in Data-Driven Decision-Making Routines

Data-focused or data-driven approaches to decision making are widely used ways to conceptualize data use in schools and allow us to unpack the “ostensive” or idealized aspects of team-based uses of assessment data. Routine school-wide and team-based instructional planning can take many forms at the school level. Often it engages the principal and other staff (who exercise leadership formally or informally) in discourse about the nature of their joint work, the tools (such as data, assessments, etc.) for undertaking that work, and their cumulative knowledge on approaches to practice for the purpose of improving practice. Studies of instructional leadership teams (ILTs) have generally placed principals as the “leader” of one or more instructional leadership teams (Portin, et al., 2009). These teams are generally comprised of teacher leaders (Little, 2003; Smylie & Denny, 1990) and instructional specialists (Dexter, Seashore Louis, & Anderson, 2009; Matsumura, et al., 2009). These ILTs have often been the focus of

research on data for improvement and decision making (Knapp, Copland, & Swinnerton, 2007; Leithwood & Riehl, 2005; Portin et al., 2009; Wayman & Stringfield, 2006).

These formal and informal teams focused on instructional improvement often meet regularly throughout the year to collect data and assess their practice and student outcomes. Inquiry into organizational routines and practices has highlighted the cyclical nature of inquiry and informal evaluation within the larger yearly evaluation cycle of state assessments (Portin et al., 2009); while theory-building within data-driven decision making (DDDM) processes has emphasized the ways in which educators collect, organize, and process various forms of data and evidence to make actionable decisions for school improvement (Marsh & Farrell, 2014; Ikemoto & Marsh, 2007; March, 1994).



Adapted from Ikemoto & Marsh (2007) and Marsh & Farrell (2014)

Figure 1. Data-Driven Decision Making Process Framework

In particular, Ikemoto and Marsh's (2007) framework of the DDDM process (Figure 1) provides an understanding of the sequences, flow, and connections between

data use activities and the co-construction of actionable knowledge and decision making. Their framework sketches out the larger opportunities for interpretation within DDDM through four main steps: first (1) various forms of *data* (input, process, outcome, and opinion data) are collected and presented to the group; then (2) these forms of data are combined with individual understandings of the situation (i.e., personal insights and experiences) to yield various forms of *information* or evidence that speaks to the issue at hand; then (3) together these “data users” convert sources of information into actionable *knowledge* by negotiating, prioritizing, hypothesizing, evaluating, and assessing varied problems of practice; and (4) ultimately coming up with actionable *responses*, that is, decisions to be implemented. Outcomes of implemented solutions then become new sources of data and feedback for the DDDM process. This framework is useful in that it provides us an abstracted understanding of opportunities for interpretation by “data users” and allows us to differentiate between different dimensions of “data talk” within the DDDM process.

However, as the authors note, this framework “fails to capture the nuances and variation that occur when educators go about making decisions in real-world settings with competing demands on their time and attention” (p.110). They point to the fact that—in practice— data users might skip the sources of data and rely directly on personal knowledge and intuition, as Rainey (2014) found when school leaders were faced with gaps in their data; or users may triangulate multiple forms of data and evidence to form a more comprehensive body of knowledge to inform their decision. Additionally, this framework also differentiates between simple (i.e., less

comprehensive, or one-dimensional) and complex (i.e., more integrated and multidimensional data) forms of data, the types of data (input, process, outcome, and opinion), and their sources (i.e., state tests, individual opinion, personal story or account), which—when grouped together—interact to yield deep or shallow inquiry for instructional improvement. Yet, within this model all forms of data and their sources are assumed to be equal in terms of being seen as “valid” sources of evidence for decision-making, and neglects to consider how situational contexts inform which forms of data and evidence are more likely to be used than others.

Finally, this model does not provide us an understanding of how educators interact with each other and with data to create deep or shallow forms of inquiry and meaning. Therefore a more complete model of data use in practice must draw on additional conceptual tools that represent the ways in which educators, along with their forms of data and evidence, create shared meaning in practice, and the sources that enhance or inhibit that sense making.

Inside Organizational Routines for Data Use: Opportunities to Learn and the Construction of Meaning within Communities of Practice

A second set of ideas takes us beyond ostensive aspects of data use in school-based teams to the more “performative” or data use in situ. Using various forms of data to construct actionable knowledge for collective decision making is largely an interpretive, iterative, and discursive activity. Work within communities of practice theory (Wenger, 1998) provides a deeper exploration of the largely situated and

interpretive DDDM micro-processes that are at work in school-based teams working with assessment data. Communities of practice research takes a situated and social perspective on learning and refers to a deeper level of engagement through “active participation] in the practices of social communities and constructing identities in relation to these communities” (Wenger, 1998, p. 4). The active participation of instructional leadership teams or professional learning communities with cycles of inquiry provides insight into how data-focused organizational routines develop and shape “shared practices, historical and social resources, and common perspectives”(Coburn & Stein, 2006, p.28) within local data use practices.

Research exploring the impact of organizational routines and the establishment of communities of practice has examined a variety of instructional improvement endeavors. Coldren (2007) utilized Wenger’s (1998) communities of practice theory to explore concepts of boundary practices and boundary objects to examine how organizational routines and tools established and maintained connections between school leadership activities and teaching and learning. Halverson, Grigg, & Thomas (2007) explored how organizational routines were used by leaders to shape professional communities, incorporate research into faculty discussions, and provide feedback on effectiveness of instructional programs in light of accountability policies and pressures to shape school improvement plans that could lead to improved practice. This research suggests that organizational routines and tools for decision making paired with individuals who share common understandings, worldviews, and professional language

help to provide educators useful connections and shared meaning during school reform initiatives.

Negotiation of Meaning in Team-based Data Use Interactions

The logic of change within this work assumes that instructional improvement is not entirely derived from a series of data-informed decisions, but also from the formation of shared meaning[s] that educators derive from participation in a community or communities of practice inside and outside of their organization. Utilizing aspects of communities of practice theory provides useful framing ideas in which to explore how school leaders within DDDM processes engage with each other and negotiate meaning from forms of assessment data and evidence. For the purposes of this inquiry, I focus primarily on the negotiation of meaning, as laid out in communities of practice theory, and consider how participating in data- or assessment-focused discourse interacts with and is shaped by reified—or abstract—notions of “data,” “evidence,” and “student learning” found in artifacts of data or evidence used for inquiry. Wenger (1998) imagines the concept of reification to “cover a wide range of processes that include making, designing, representing, naming, encoding, and describing, as well as perceiving, interpreting, using, reusing, decoding, and recasting” (p.59). In essence, reification can refer to both a process and its product, such as the process of designing a data display wherein the categories of students become localized heuristics for educators to talk about grouped student performance instead of specific students.

This process of reification within communities of practice theory draws some similarities to the way that Ikemoto & Marsh (2007) envision how school leaders engaged in collective DDDM integrate and process “raw” assessment data with their own understanding of the situation to provide forms of information and evidence for possible solutions. In particular, their conceptualization of DDDM process indicates key opportunities for the negotiation of meaning, such as organizing “raw” data for display, utterances of data users’ interpretations of the situation in relation to this information, or the collective process of created shared knowledge to inform actionable decisions for problems of practice. Integrating the concept of reification within their conceptualization DDDM process, educators can take various forms of evidence (*Data*) of student learning such as a state-standardized assessment which already represents a fixed notion of “4th grade math proficiency” and compare their students’ performance against this standard to their own understandings of what “4th grade math proficiency” looks like, which may or may not align (*Information*). Together, educators can then take these scores and rank or categorize students to provide a useful display or heuristic (*Knowledge*) for educators to talk about how students are doing (such as talking about “red,” “yellow,” or “green” students) and discuss and plan decisions for academic improvement (*Decision Outcomes*). In this example, the notion of “math standards” found in a state-standardized assessment was used to create a new form of “data” or evidence in the form of a color-coded data display; that is, an established—or reified—“math standard” found in an assessment informed and shaped the discourse of student achievement to create new local understandings of student success.

In this way, data use in practice ideally provides data users opportunities to learn from and form shared meaning with each other. Yet—similar to DDDM processes—communities of practice theory and its promise for collective learning suffers from aspects of the organization and context for learning. For example, Gallucci (2003) found that communities of practice among teachers can vary by relative strength or weakness of the community, and the relative openness of the community to actively engage with new ideas. Therefore inquiry into the micro-processes of data use must also consider aspects of the organization or configuration of educators that can potentially impact participation and learning within data use.

Additionally, it is important to note that while forms of assessment data may seem “raw” or unbiased to educators, assessments themselves already contain previously negotiated forms of meaning that have been reified into what is now a “4th grade math standard.” The extent to which educators feel that they have agency to negotiate what they believe a “4th grade math standard” should be in contrast with what high-stakes accountability measures define as standards has implications for the types of meanings that get privileged or negotiated (Gallucci, 2003), and the extent to which subsequent decisions reflect short-term or long-term solutions. In high-pressure school contexts, it is important to consider what meanings of student success educators take out of their interactions with static notions of student learning (see Horn’s (2007) “Fast Kids, Slow Kids, Lazy Kids;” or Koyama and Menken’s (2014) exploration of ELL students as “numerical liabilities”). Therefore, focusing in on patterns of interaction in DDDM—such as aspects of discourse and participation— in conjunction with various

forms of assessments may provide opportunities to unpack the complexity of DDDM processes, by allowing the research to spotlight the negotiation of meanings that are part of that interaction.

Boundary Objects as Discursive Spaces for Collective Learning

Within data use practices, student assessments often serve as focal points or the main catalyst for focused inquiry. As stable representations of student learning, school leaders can use assessments to facilitate coordinated activity across the boundaries of different learning communities inside and outside the school, while maintaining a unified understanding of student achievement goals. The various forms of assessment utilized within participation of instructional leadership teams and ongoing inquiry to improve instruction can be viewed as creating the boundaries or “discursive spaces” for joint activity of collective instructional leadership. One further idea taken from communities of practice theory (Wenger, 1998), is useful at this juncture: the concept of *boundary objects* (Star and Griesemer, 1989) allows us to track how assessments facilitate coordinated activity across communities (Nolen, Horn, Ward, & Childers, 2011). Due to their shared language and evaluative statements, assessments can both facilitate coordinated activity across social boundaries inside and outside of schools (Nolen, Horn, Ward, & Childers, 2011) and can be helpful in tracking and comparing how forms of meaning within assessments get taken up (or not) by educators. As boundary objects, assessments serve as an impetus for collaborative leadership activity and as a “problem space” for discourse between different practitioner communities.

Within these discursive spaces, instructional leaders make sense of state and district expectations, learn from each other, and begin to negotiate the implementation of accountability mandates.

Wenger (1998) and Star and Griesemer's (1989) work on boundary objects illustrates how assessment tools (i.e., standardized tests, in-class student assessments, language proficiency tests) can shape the discourse around assessment and school improvement practices. As boundary objects, state standardized assessments are "...both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites (Star & Griesemer, 1989, p.393). Star and Griesemer (1989) also point out that "...the creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds" (p.393). Therefore, state standardized assessments can be seen as an attempt by policy makers to craft coherence within curriculum and instruction, and to set a bar for student achievement across school sites.

Thus far, I have explored conceptually how collective data-focused organizational routines create opportunities for shared meaning- and decision-making. I have also unpacked how, within DDDM mechanisms, educators interpret forms of data and evidence and collectively craft actionable knowledge, decisions, and additional forms of reified "data." Finally, I have suggested that this situated and interpretive "joint work" takes place within various communities of practice that are bound together by stable representations of student learning found in student data and assessments that constitute "boundary objects" within team discourse. I argue that the interaction between

these three aspects of team-based data use practices—individual assessment, collective discourse, and forms of data as boundary objects are important areas for focused inquiry into the situated nature of data use micro processes. In the next section, I provide a final set of ideas that characterizes the individual, collective, and student data as part of a larger assessment system operating within educational organizations; outline several factors that contribute to or detract from the interaction of these features of school assessment; and provide an integrated conceptual framework outlining these ideas in action.

Assessment-as-Practice Within Team Discourse and the Potential for Situated Collective Learning

A third set of ideas gives us more useful and specific ways to consider the actual kinds of data at play in the school-base instructional team’s interactions, and in particular, the discourse it engages in within this space. Integrating the more social and situated aspects of meaning-making in communities of practice theory with DDDM understandings of assessment data, I turn to *assessment-as-practice* theory (Jordan & Putz, 2004) to characterize and expand our understandings of the dynamic assessment activities of communities of practice within a school. *Assessment-as-practice* theory (Jordan & Putz, 2004) situates assessment into a broader social context and examines how the everyday often informal assessments (“inherent assessments”), collective social discourse, and the formal standardized assessments (“documentary assessments”) shape the daily practice of school leaders and teachers. For example, Jordan & Putz (2004)

argue that an over-reliance on documentary assessments—such as in a high-stakes accountability school environment—could lead to dysfunctional effects on practice, decision-making, and the structure and culture of the organization. Therefore it is important to consider the ways in which aspects of assessment within schools could lead to inequitable learning practices, particularly within Latino and emergent bilingual serving schools.

Types of Assessment

Assessment-as-practice theory differentiates the organic and informal kinds of assessments produced by individuals and groups as a social process—*inherent and discursive assessments*—from the formal, standardized assessments used in organizations, or *documentary assessment*. Inherent and discursive assessments happen in everyday practice through either nonverbal monitoring or group conversation where the knowledge produced is either tacit, implicit, or made explicit through discussion. For example, when a principal observes a classroom and makes an inherent assessment about the quality of a teacher's work, the assessment can become discursive if s/he chooses to make aspects of that assessment salient to the teacher and/or an instructional leadership team.

Conversely, documentary assessments involve "...externally mandated, stable symbolic representations of evaluations in the forms of tests, surveys, check lists, plans, targets and similar instruments" (Jordan & Putz, 2004, p.351). Similar to ideas surrounding reification, assessment tools, such as standardized tests, can be viewed as

manifestations of a “management rationale” of student learning that provides indicators of student learning that are “linked to educational, political, and economic concepts” within policy-makers’ organizations (Jordan & Putz, 2004, p.351). Systematic investigation of organizational responses to accountability (e.g., discourse and types of evidence used as “data” for decision making) can potentially build greater internal responsibility of school leaders for the unintended negative outcomes of student learning (Jordan & Putz, 2004).

Expanded Notions of Evidence and Assessment

in Team-based Data Use

Therefore, assessment-as-practice theory (Jordan & Putz, 2004) asks us to re-examine our conceptualizations of “assessment” and consider the shared and reflective talk of instructional leaders engaged in assessment practice as an avenue for examining school leaders’ learning and shared understanding of their role and responsibility in instructional leadership. This understanding is based on their consensus of the available capabilities, resources, and other problems of practice impacting their student test scores. Examining how and in what ways assessment discourse shapes the joint activity of instructional leaders is an important first step in uncovering to what degree and in what ways high-stakes accountability assessment practices enhance or detract from organizational opportunities to learn within culturally and linguistically diverse schools. Exploring this discourse within organizational routines for data use provides one avenue for investigating where conceptual change happens within schools.

Analyzing inherent and discursive assessment practices within an organization is imperative for leadership work aimed at school improvement and for identifying potentially deficit practices within a school's culture. Since the performance assessment movement in the 1990s, "assessment" in an accountability context has predominantly focused on students' "mastery of small, isolated bits of knowledge or skill, attested by darkened circles on a multiple-choice answer sheet" (Pullin & Haertel, 2008, p.31) found in summative assessments. Conversely, "assessment" could also refer to the range of formative individual and collective assessment practices of school leaders aimed at teachers to improve instruction (Fink & Resnick, 2001; Hallinger, 2005; Honig, Copland, Rainey, Lorton, & Newton, 2010; Knapp, Mkhwanazi, & Portin, 2011).

Reframing assessment as informal, formal, and documented assessment *practices* allows us to situate assessment in a broader social context (Jordan & Putz, 2004). They highlight that internal-group assessment—besides creating learning and shared understanding of individual roles and responsibilities—often creates "public verbal representation of the capabilities, resources, and issues for a group that enable them to consider implications of the current state, as they understand it, for behaving more effectively" (p.351). Additionally, the types of assessments introduced in instructional leadership team meetings can privilege a particular theory, or rationale, of learning over others. The authors argue that—when left largely unexamined—documentary assessments have negative consequences such as: 1) manipulating the numbers; 2) changing work practices; and 3) modifying organizational structure, climate, and culture. Therefore a first step in investigating the degree of equity consideration within

assessment practices is identifying *how* and in what ways shared meaning around assessment is formed within collective instructional leadership practices within a particular school context. By framing assessment in this manner, we can examine school practices that have implications for equity and accountability.

Collective Data Use Outcomes and their Sources

within Team Discourse

DDDM and communities of practice theories suggest that two key outcomes of collective data use practices are actionable decisions and shared meanings within and across communities of practice. Examining discursive assessments, in particular, externalizes aspects of group learning by providing insight into how aspects of inherent and documentary assessments interact with one another as educators verbalize and frame problems of practice, offer explanations, debate potential avenues for action, and form solutions. How data users interact with one another, the forms of evidence and data presented to them, and their reactions to particular forms of interpretations and solutions may provide insight into not only how decisions are made in data-informed decision making, but also the extent to which the collectivity of data users form shared meaning surrounding problems of practice, students, curriculum, etc.

Framing assessment as (1) an internal activity of individuals; (2) shared and reflective talk generated by a group of individuals engaged in joint activity; and (3) stable representation of assessment and theory of learning, serves as a common thread between situated/sociocultural theories of learning and organizational learning theories.

Additionally, research has suggested that group reflective talk in discursive assessment is an important avenue for learning and innovation (Jordan and Putz, 2004; Greeno and van de Sande, 2007). Examining the discourse of instructional team leaders in meetings surrounding aspects of internal assessment and accountability data may allow us to uncover potential mechanisms through which seemingly uncontested examination of “data” can, at best, maintain existing deficit assumptions and practices, and, at worst, actually reinforce or reify inequity for Latino and ELL students. Therefore, it is important to identify aspects of internal, discursive and documentary assessment data that may have implications for affordances and constraints to individual and group learning during data use practices.

Factors that Shape Data Use Practice and Discourse in Latino and ELL-serving Schools

In this section, I review research on the various affordances and constraints within inherent, discursive, and documentary assessments that may impact educators’ decision making and opportunities to form shared meaning within Latino and ELL-serving schools. These factors have implications for the ways in which these forms of assessments interact with each other and contribute to the forms of negotiation, prioritization, evaluation, and meaning making of available forms of evidence for decision-making.

Documentary Assessment Factors

As mentioned previously, the presence of assessment data does not provide clarity as to how to improve student outcomes. In fact, it may provide confusing, misleading information or be missing altogether for timely decision making. For example, forms of state-standardized assessment data for ELL students may not provide accurate information on their student learning progress due to English proficiency issues (Gándara & Baca, 2008; Menken, 2008; Solorzano, 2008). Additionally, aspects of the data itself—timeliness, grain scale, accessibility, perception of validity or applicability— may also contribute to the extent to which educators effectively use data within inquiry surrounding instructional improvement (Ikemoto and Marsh, 2007; Moss, 2013). Finally, the school’s accountability context (i.e., if they are “under review” or “probation”) has implications for the weight that state-standardized test carry when educators focus on short- versus long-term problems of practice (Hargreaves & Shirley, 2009).

Discursive Assessment Factors in Team-Based Inquiry

Team dynamics, collaboration, facilitation for data-focused decision making, and environment for practitioner-led innovation all contribute to the forms of negotiation, prioritizing, and evaluation that take place in discursive assessment. Research on intersecting communities of practice has shown us the road to conceptual understanding can be bumpy (Coburn, 2001; Gallucci, 2003). It may take time for members of instructional leadership teams to develop “shared practices, historical and social

resources, and common perspectives” that characterize a community of practice (Coburn & Stein, 2006, p. 28). Additionally, the strength and openness to learning of a community of practice (Gallucci, 2003) can impact how educators push each other’s thinking around data and instruction. In particular, instructional leaders often try to support discursive practices that maintain effective and consistent collaboration within instructional leadership teams and communities of practice. Fenwick (2007) suggests that negotiation between organizational communities involves unique discursive practices that involve the capacity within an individual “...to be critically attuned to the shifting discursive patterns that emerge in negotiations among different constituents: overlapping discursive communities, troubling discursive intersections and resistant discourses” (p.140). While principals within instructional leadership teams must share instructional leadership with teacher leaders, they must also be sensitive to the dynamics of the individuals within the room, understand their particular data needs, and help them stay on track and form consensus around a plan of action (Leithwood & Riehl, 2005).

Finally, environmental constraints to practitioner-led innovation may also impact the types of “solutions” to problems of practice. For example, a study by Kerr et al. (2006) found that when assessment data identified areas that needed re-teaching, teachers felt that they lacked discretion to veer from the pacing of district-mandated curriculum guides. As a result, they followed the guides instead of addressing problem areas identified by the data. A study by Black (2006) examining Texas accountability and ELL transitional assessment practices found that teachers interpreted the tension between conflicting priorities within the school—to develop the native language

sufficiently or quickly transition students to be eligible for state tests—as interference to their practice, which led some teachers to disregard research on language acquisition and cast ELL students as needing sympathy more than high expectations.

Inherent Assessment Factors

Lastly, aspects of leadership capacity to lead for instructional improvement, data literacy and notions of data use, and factors that shape individual interpretation of data all interact with documentary and discursive assessment practices. Assumptions surrounding the role of principals in instructional leadership (Kowalski, 2010) may create tensions between leading or creating opportunities to engage with data. For example, the extent to which principals involve themselves in curriculum and instruction can be influenced by their knowledge of their role in instructional leadership (Mangin, 2007) and the immediacy of short-term goal-setting caused by accountability pressures (Hargreaves & Shirley, 2009). Principal effectiveness in supporting instructional learning (Matsumura, Sartoris, Bickel, & Garnier, 2009) and their perceptions of the appropriate theory of action in instructional leadership can vary (Reitzug, West, & Angel, 2008). Given these discrepancies between principal knowledge, capacity, and implementation of instructional leadership, it is unclear how school leaders should participate in and contribute to data use practices within schools. Furthermore, even when school leaders provide opportunities for educators to interact with data to investigate problems of practice, they are held up by different level of data literacy possessed by the team. Being able to read graphs, understand statistical findings, and

draw conclusions to relevant changes in practice are all required in order for educators to move from data, to information, to actionable knowledge within DDDM practices (Marsh & Farrell, 2014). Additionally, educators may also choose to under- or over utilize forms of data they feel is applicable to their practice based on personal definitions of forms of data and their uses (Cho & Wayman, 2014).

Educators serving Latino and emergent bilingual students may also not have the capacity to create responsive changes to instruction regardless of an abundance of data. For example, the majority of mainstream teachers have had little to no professional development to teach ELLs; few have taken courses specific to ELL needs; and many do not feel adequately prepared to teach ELLs (Lucas, Villegas, & Freedson-Gonzalez, 2008). This lack of “fit” between educator training and student learning needs may create a vicious cycle of attempts to improve student outcomes with little to no success. Conversely, educators’ beliefs about student ability, race, and language can be constraints on an individual’s understanding of the implications of data on their practice. For example, Diamond and colleagues (2004) found that within schools with a majority of African-American and low-income students, teachers held more deficit beliefs about student abilities (e.g., lack of motivation, negative family influences, and limited skills), which they felt undermined their ability to effectively teach. Datnow et al.'s (2003) work in examining comprehensive school reforms in culturally and linguistically diverse schools found that the "social construction of language minority, immigrant students as low ability and lacking in basic skills served as a constraint in both educators' initial receptiveness to reform designs as well as in their subsequent implementation of them"

(p.158). As a result of these barriers to responsive decision making, practitioners may find themselves with eroded self-efficacy in their ability to improve instruction in response to specific accountability mandates and pressures. Additionally, general stress and anxiety due to governmental policies thought to be unclear or meaningless can further malign implementation of organizational routines for student improvement (Leithwood, Steinbach, & Jantzi, 2002).

Conceptual Framework

In summary, I drew from an emerging body of research in education that takes a situated perspective on the micro-processes that develop shared meaning and decision making within data use practices. I considered the ways in which various forms of assessment practices—inherent, discursive, and documentary—shape interactions within DDDM mechanisms and provide insight into factors that contribute to or inhibit these interactions. This study assumed that organizational routines within schools create opportunities for educators to collectively engage with data in varying levels of collaboration, and that, within schools with established instructional leadership teams and communities of practice, these educators come together for joint discourse and action surrounding various student assessments throughout the school year. Joint discourse is made possible by assembling a meeting surrounding a particular “boundary object” such as a documentary assessment (e.g., state test scores).

Located within a particular school context—those serving a concentration of emergent bilingual students—these dynamics will play out in ways that have important

implications for how the students are attended to (or not), how the students are viewed and understood, and how problem of instructional practice related to their learning needs are addressed (or not). Because such schools sit within a high-stakes accountability environment, and because the student population served is likely to score poorly (on average) on the assessments, the discursive work within school's data-use routines will take on a particular urgency.

This integrated conceptual framework (Figure 2) suggests that embedded within conceptualizations of DDDM mechanisms outlining how data users collect, interpret and process data into actionable decisions, there are opportunities for data users within communities of practice to form shared meaning and understandings as they engage with forms of assessments as “problem spaces” for inquiry. In particular, assessment tools are not only the “glue” or the boundary in which discourse about assessment and problems of practice takes place, but also they contribute to the prioritization of evidence, information and merit of potential solutions during the translation of “raw” data to shared actionable knowledge.

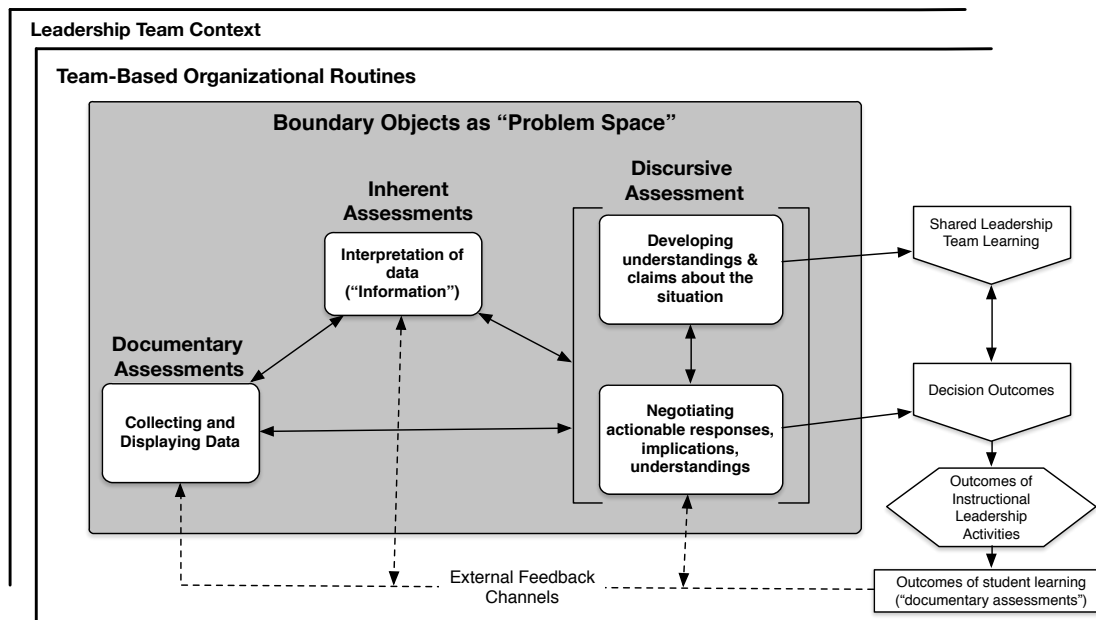


Figure 2. Situated Collective Data Use Practices

The outcomes of instructional leadership team discourse (problem framing, solutions generated and conceptual change) are influenced by aspects of inherent, discursive and documentary assessments. Interactions between these mechanisms then impact the forms of negotiation, inquiry, evaluation, and prioritizing that occur within situated and collective data use practices. As mentioned previously, the success of data use practices depends on affordances and constraints to information and concepts (such as data literacy and preparation, belief and knowledge systems of individuals); aspects of the tools themselves for sense making (e.g., data representation, preparation, number and range of assessments); influences on team dynamics and participation structure (such as ability for facilitation, norms of participation); and finally, overall environment for joint action for instructional reform.

The fact that data use is situated within an embedded school context, means that high-stakes accountability reform and state- and district-specific policies concerning identification and assessment of English Language Learners also contribute to the interaction of assessment practices within the school. The influence of this embedded context has implications for the types of dialogue and decisions about instructional practices and evaluation instructional leaders make within their school.

This framework's focus on situated social learning processes highlights how discourse is used to form conceptual knowledge through joint action of community practices that utilize various planning, coordinating, and evaluation practices in the service of accountability. This framework also suggests that there may be several avenues where the complex nature of ELL student learning and accountability policy contexts that may complicate leadership practice. Therefore, its focus on examining micro-processes of instructional leadership team discourse on assessment is an important first step in understanding how, and in what ways, school leaders and educators challenge or perpetuate student inequality and inequitable opportunities to learn for Latino and ELL students.

Chapter 4:

Study Design and Methods: A Situated Qualitative

Case Approach

In order to examine situated collective instructional leadership team decision-making policies, procedures and responses within a Latino and emergent bilingual serving school, this study employed a situated case study (Merriam, 2009) of the collective assessment practices of an instructional leadership team within a rural elementary school in the Pacific Northwest serving largely Latino, ELL, and Title 1 students. This study centered on how instructional leaders interpret forms of data and evidence—such as documentary assessments—to form decisions and construct shared understandings of instructional improvement. I utilized a situated sociocultural approach to examine the data use practices of an instructional leadership team within a high-stakes accountability environment and Latino and emergent bilingual serving school. This work situates itself in previous critical sociocultural studies of education policy as a practice of power (Levinson, Sutton, & Winstead, 2009) by ethnographically examining how and in what ways an instructional leadership team appropriates, or not, the theory of learning proposed by high-stakes accountability measures for their culturally and linguistically diverse students within their data use practices.

Within sociocultural learning traditions, “learning is viewed as an essentially social process, situated within cultural, institutional, and historical contexts” (Knapp, 2008, p.527). This perspective situates learning and knowing as a process of practice and

participation that considers the interconnected and mutually defining elements of meaning, community, identity, and practice within learning (Lave & Wenger, 1991; Wenger, 1998). Thus, this examination of situated data use practices considers the ways in which the social structures of schooling (i.e. ELL policies, accountability contexts) and local practitioner dynamics interact with their enactment of collected evidence and data use practices in the service of instructional improvement.

To zero in on these local and situated dynamics of data use practices, I conducted a case study of instructional leaders' discourse and interactions with a variety of assessments and sources of evidence within their collective data use practices. Case study allows for an in-depth analysis and "thick description" of a bounded system (Merriam, 2009), which allowed me to unpack how situated data and evidence use within organizational routines impact decision-making and opportunities to create shared meaning. The central task of this qualitative study was to "...discover the specific ways in which local and nonlocal forms of social organization and culture relate to the activities of specific persons in making choices and conducting social action together" (Erickson, 1986, p. 129). A qualitative design allows the researcher to perform "inductive" investigations of how instructional leaders navigate and negotiate multiple demands of staff, students, and resources in responding to problems of practice articulated by NCLB-mandated targets and knowledge of "best practice" for Latino and Latino ELL students (Erikson, 1986, p.121).

Examining data use and assessment practices of instructional leadership teams provides a window into how these processes within data use practices shape the

problem framing, interpretation of data and evidence, and the possible solutions aimed at instructional improvement. Within this study, affordances of and constraints—internal and external to the school— on instructional leaders’ ability to form shared meaning from assessment were examined through analysis of team meetings, interviews with leaders and practitioners, and through documentary assessment analysis. Utilizing interviews with school leaders allowed for insight into the culture within the school surrounding high-stakes testing and bilingual education, as well as an understanding of their role and responsibility in leading instructional change. Also, utilizing “rich descriptions” of instructional leaders’ assessment-focused problem-solving processes (Merriam, 2009, p.28) and a description of the outcomes of instructional leader team meetings—afforded by the triangulation with staff interviews—created a more robust description of school leadership assessment practices. By moving beyond problem identification and decision-making in data use practices, I was able to critically examine the “social arenas where the interests and languages comprising a normative policy discourse get negotiated into some politically and culturally viable form” (Levinson, Sutton, & Winstead, 2009, p.778).

Additionally, *how* individuals interact with data and with each other also has implications for shaping ideas, norms, and values of a school community (Coburn, 2011). Therefore, the focus of this study is not specifically on accountability or language policies within a school, but on the discourse of accountability practices and the outcomes of this discourse within their embedded context. I assume that an instructional leadership team can be viewed as a ‘community of practice,’ which can be defined as

“an aggregate of people who come together around mutual engagement in an endeavor” (Ekert & McConnell-Ginet, 1992, p.464). The discourse practices within a community of practice display both observable behaviors, as well as underlying sociocultural belief systems. Therefore I drew on analytic techniques developed by Wells (1996) to explore episodes of data talk and sequences of DDDM processes to examine how micro-level instructional leadership interactions form conceptual knowledge through joint actions of a community of practice—specifically, the actions that utilize various planning, coordinating, and evaluation practices within a larger, social, political, and historical context.

Site Selection and Setting

In order to understand how school leadership makes sense of standards-based accountability messages for Latino and Latino emergent bilingual students, one must first examine the distinct “...structural and historical conditions framing practice” for school leader implementation of NCLB policy (Merriam, 2009, p. 35). The rural context of Washington state presents a unique opportunity to capture local capacity building activities to serve a rapidly growing Latino and emergent bilingual student body. Different from states in the southwest that have a longer history of learning to provide equitable practice for Latino and emergent bilingual students, and unlike southern states characterized as the new “Latino Diaspora” (Wortham, Hammond, & Murillo Jr., 2001) that are only recently experiencing concentrations of Latino students, Washington state falls between these two extremes. In 2013-2014 English Language Learners represented 102, 339 students (out of a statewide total student population of 1,005,517), an increase

of .7 percent from the previous year with the majority of the enrollment concentrated in the urban areas along the Interstate 5 corridor and the rural areas of the Yakima Valley (OSPI, 2014, 2015).

Also, unlike states which have explicit English-only policies of language acquisition, Washington state's Transitional Bilingual Instruction Act provides opportunities for districts to implement models of bilingual instruction that fit their needs and organizational capacity. However, as OSPI reports, the state still struggles with the shortage of properly trained ELL instructors to implement more ambitious bilingual programs (Malagon, McCold, & Hernandez, 2011). The Latino Achievement Gap Report (Contreras & Stritikus, 2008) found that this issue was especially dire within rural districts, which generally serve homogenous language minority students in contrast with the diverse range of languages found in the western Washington context. Furthermore, the period of language acquisition is especially sensitive for young learners, and secondary schools often employ pullout methods of language acquisition rather than transitional or dual language programs.

Rural Washington provides a unique example of a state context (and regional context within it) that offers opportunities for districts to create bilingual language policies to serve their predominantly homogenous language population, but that struggles with the recruitment of highly qualified staff trained in Second-Language Acquisition and potentially with pressure from high-stakes accountability environments. This context, similar to many new Latino settlements, describes a community in transition and was chosen to highlight some of the challenges instructional leaders face

in building capacity within their schools. This bounded organizational system was selected in order to "critique and challenge" accountability cultures within high-stakes testing and "to transform and empower" organizational change toward increased Latino student achievement (Merriam, 2009, p.34).

School and District Context for Instructional Improvement

Within a rural part of Washington state, I selected a particular school district and within it an elementary school that provided a good setting for pursuing my examination of assessment practices in leadership team discourse. In particular, Mountain Valley School District and Crossroads Elementary School offered a particularly appropriate kind of context for understanding leaders' assessment practices and the implications for the way they served an emergent bilingual student population.

Mountain Valley School District⁵ is a small, rural school district in Washington state serving over 6,000 students enrolled in its 10 schools. It has experienced rapid demographic change in its student population over the past two decades. In 1998, 28.9 percent of students in the district were Latino. By 2012, 54.4 percent of students identified as Latino. Of the 24 percent English Language Learners within the district, over 90 percent of ELL students identify as Latino. In addition, the Mountain Valley School District also has 67.3 percent of its student population qualifying for federal free and reduced lunch (OSPI, 2013). From 1998 to 2013, the district's migrant population

⁵ Pseudonyms were used during this study.

slowly diminished from 12 to 8 percent. This suggests a growing number of permanent settlements of migrant farm workers working within the surrounding fertile valley.

Crossroads Elementary is a small elementary school serving 630 students with a higher proportion of Latino and ELL students within the district, with 74 percent of students identifying as Latino and 48 percent of students listed as Transitional Bilingual. While the remaining percentage of students are predominantly White, 86 percent of the students in Crossroads Elementary qualify for free and reduced lunch. The 35 teachers at Crossroads possessed an average of 14 years of teaching experience; 75 percent have Masters degrees; 100 percent of teachers have met ESEA standards for highly qualified certificated staff; and the majority of staff identify as White.

During the 2012-2013 school year, Crossroads was designated a “Priority School,” which places it “...among the lowest five percent of Title 1 schools in the state, based on achievement on the statewide assessments, with a demonstrated lack of progress on those assessments over three years.”⁶ The goal for these schools is to improve school performance, close opportunity gaps, and substantially improve student learning and outcomes.

Nested Organizational Contexts for Understanding

Leaders’ Assessment Practices

Leadership within this district and school site is exercised within two nested organizational contexts: that of the population they serve, as a Latino and ELL majority

⁶ Retrieved from the Mountain Valley School District website.

school, and that of a high-stakes accountability environment. Educational outcomes of educational policy implementation must be understood as a function of how multiple, embedded contexts influence (and potentially constrain) school leaders' sense-making and interpretation of data use within a high-stakes environment. Using a situated sociocultural discourse analysis of school leadership accountability practices can potentially show how "...inequality emerges from structuring activities to become external and constraining on social actors" (Mehan, 1992). For example, schools serving a critical mass of Latino and ELL students often struggle with meeting AYP for their student sub-groups. Schools that are seen by the district as "academically unacceptable" or "under probation" are perhaps under a more strict and prescribed schedule to address their low-performing scores (Diamond & Spillane, 2004). Therefore, even for schools not "in need of improvement," the fear of losing autonomy can be a motivator to change instructional practices, at times to the detriment of student learning. The following ELL designation policies, available assessments, and ESL curriculum for emergent bilingual students within this district provide important context for how instructional leaders were able to label, track, and be held accountable for their ELL student subgroup outcomes.

ELL designation policies and assessments. Mountain Valley School District currently uses the Washington English Language Proficiency Assessment (WELPA) in order to annually assess the growth of emergent bilingual students in grades K-12. The WELPA consists of two tests: the Placement test to determine initial eligibility for the English language development (ELD) services, given to all students whose families answer "a language other than English" to question 2 or 3 on the Home Language

Survey; and the Annual Test, given to all students who qualified for ELD services with the Placement test, to measure students' growth in English language knowledge skills and determine if students are still eligible for ELD services (OSPI, 2015).

Students are tested in their speaking, listening, reading, writing, and comprehension skills, and their overall score places them in a Proficiency Level of 1 (Beginning/Advanced Beginner), 2 (Intermediate), 3 (Advanced), or 4 (Transitional). Students who score at Level 1, 2, or 3 continue to receive ELD services. Score reports available to staff include individual Student Proficiency Reports and Group List Reports of a roster of all students in a group, such as a class. In addition, scores from speaking, listening, reading, writing, and comprehension are combined to provide both parents and teachers with information surrounding the students' English comprehension, productive language (such as expression), oral skills for collaboration and interaction in social and academic tasks, and literacy in both reading and writing (Perrone, 2014).

ESL curriculum. Mountain Valley School District supports Washington state's Transitional Bilingual Program wherein students may be assigned to an English as a Second Language (ESL) classroom for a portion of instruction in English or receive tutorial assistance. Crossroads Elementary utilizes a program called "Content-Based English as a Second Language" that is designed to teach English to non-English speakers through the use of English instruction of grade-level material (Reed & Railsback, 2003). This program is designed to provide ELL students with background knowledge and vocabulary (as opposed to more simplified ESL texts) through the use of a variety of strong language learning components (Brown, 2004).

Due to the large number of emergent bilingual students within Crossroads Elementary, all teachers are assumed to provide Content-based ESL within their mainstreamed classes, with bilingual instructional assistants providing tutoring and support during class, as well as the ELL Specialist providing focused support for a select group of students outside of the classroom. Additionally, it was reported by the instructional leaders and staff that they have received some training in nationally recognized professional development programs to support teachers in working with second language learners, such as the Guided Language Acquisition Design (GLAD) or with the Sheltered Instruction Observation Protocol (SIOP[®]).⁷

High-stakes and interim summary assessments. Prior to Washington state's transition to the Common Core State Standards and the Smarter Balanced Assessment Consortium (SAC) in the 2014-2015 school year, Mountain Valley School District and Crossroads Elementary used a variety of summative assessments for both accountability purposes and interim assessments to support decision-making for instructional improvement throughout the 2012-2013 school year.

Measurement of Student Progress (MSP). The MSP was the state's exam for Math (grades 3-8), Reading (grades 3-8), Writing (grades 4 and 7), and Science (grades 5 and 8) as required by state and federal law.⁸ The state assessment system is required to:

⁷ GLAD is a model of professional development designed to promote English language acquisition, academic achievement, and cross-cultural skills. GLAD is a program to train teachers to develop instructional strategies effective for the learning of ELL students, along with the theory and research behind the model. SIOP, another national model of ESL professional development, focuses on teaching English language development through the content areas.

⁸ www.k12.wa.us/assessments/StateTesting/FAQ.aspx

- Test all public school students across the state, including students with disabilities and students with limited English proficiency.
- Be administered annually in selected grades.
- Measure performance based on the state’s learning standards.
- Report on the performance of individual students, schools, and districts.
- Serve as one basis of accountability for students, schools, and districts.

Within Crossroads Elementary, student test scores served as a “high stakes” test due to the school being placed on probation as a “Priority School” because of a lack of adequate yearly progress (AYP) within student subgroups.

Measures of Academic Progress® (MAP®). The MAP is a computer-based assessment program created by the Northwest Evaluation Association (NWEA) that is aligned to Washington state standards. It was used as an interim trimester assessment to monitor student growth throughout the year and to support teachers in making data driven decisions about instruction. Results are reported as *RIT* scores to show a student’s current achievement and growth over time, *Percentiles* to compare students by grade level and age, and *Lexile Scores* to monitor a student’s reading level and growth.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The DIBELS was also a trimester measurement of early literacy skills in *phonemic awareness* (hearing and using sounds in spoken words), alphabetic principle and *phonics* (knowing the sounds of the letters and sounding out written words), accurate and *fluent reading* (reading stories and other materials quickly and easily with few errors), *vocabulary* (understanding and correctly using a variety of words), and *comprehension* (understanding what has been

spoken or read).⁹ While designed as a formative assessment, the developers of the DIBELS have argued that using the DIBELS for high-stakes decisions such as student labeling, tracking, or grading should be avoided due to issues surrounding validity for such purposes and a lack of program evaluation tools (Kaminski & Cummings, 2007). The developers feel such uses have the potential to compromise instruction and they encourage the use of multiple measures for high stakes decisions.

Study Participants

Study participants were selected based on the likelihood of utilizing data for informed decision making. Therefore, initial participants were selected based on their membership within Crossroads Elementary core instructional leadership team. Teacher leaders were identified and selected based on “snowball” sampling (Bogdan & Biklen, 1998), based on recommendations from the leadership team and by their participation in a variety of teacher leader positions and activities, such as being their grade level representative within whole-school leadership team (site council) or participating in the school’s ESEA reauthorization team. These teacher leaders served as additional informants outside of the core leadership team and as members of the teaching staff.

Instructional Leadership Team:

1. *Patty, Instructional Support Specialist.* Patty taught at a different local elementary school in grades 1, 2, and 3 for thirty years where her team taught multi-age classrooms. Originally the position was created to

⁹ <https://dibels.org/bros/DIBELSBrochure.pdf> “What are DIBELS?”

support the Success For All whole-school improvement model at Crossroads Elementary. Her position then changed into providing a variety of literacy support for students and professional development and instructional support for teachers.

2. *Rebecca, ELL Support Specialist.* Rebecca previous taught in a mainstream 4th grade classroom in Chicago prior to arriving at Crossroads Elementary as an ESL teacher, with primarily small group pull-out instruction for the last ten years. Her job has since evolved to include administration, coaching, and providing professional development to teachers on ESL classroom and support strategies.
3. *Edward, Crossroads Elementary Principal.* A local and Mountain Valley alumnus, Edward began his career at Crossroads Elementary in 1996 as a student teacher, and then taught 5th grade for 11 years. During that time he was tapped by the then principal to consider the principalship. After completing his Masters and internship, he took a position as a middle school principal before returning to Crossroads Elementary as their current principal. Along with his role as principal, he also leads mathematics-focused PLC meetings.

Teacher Leaders:

1. *Simone, 6th Grade Teacher.* Simone grew up in Eastern Washington and has been working at Crossroads Elementary for the past 15 years. She started out teaching 3rd grade, but currently teaches 6th grade. She has served in a variety of teacher leadership positions and is currently the lead teacher for her grade level and site council representative. She is also part of the ESEA reauthorization planning team, and previously served as the digital literacy coach and as a model teacher in accelerated math for the district.
2. *Robin, 3rd Grade Teacher.* Robin spent years teaching in multicultural and multilingual schools in California and Hawaii before settling in Mountain Valley in 2003. She took her position as 3rd grade teacher after receiving her Masters. She is also a site council representative and participated in the ESEA reauthorization planning team.

Data Sources and Data Collection

My data sources for this study drew primarily on 41 hours of real-time observations of instructional leadership team meetings and interviews with leadership team members and staff, and later involved observations of other formal and informal collaborations for instructional improvement, such as PLCs, Site Council, and staff meetings. Documents from these observations (such as agendas, handouts, data displays, and pictures of collaboration activities) were also collected and used for analysis. These observations during the 2012-2013 school year served as primary data of instructional leaders' communication about sources of evidence, problems of practice, and instructional solutions. These recorded observations also offered insight into the processes and practices of team leadership collaboration and inquiry, shed light on the issues of learning in assessment practice, documented leadership team decision-making, and provided an historical account of the constantly shifting problems of practice.

Preliminary interviews with members of the school leadership team (principal, literacy specialist, and ELL specialist) allowed for insight into the culture of schools around high-stakes testing and Content-Based ESL education, as well as an understanding of their role and responsibility in leading instructional change. Follow-up interviews with the leadership team provided "rich description" of instructional leaders' assessment-focused problem-solving processes (Merriam, 2009, p.28) and a description of the outcomes of instructional leader team meetings.

Finally, document analysis from data use practices and interviews with key informants within the school served as supplementary data. Triangulation of multiple

data sources (Merriam, 2009) such as interviews with key staff informants, observations of professional learning community meetings, monthly staff meetings, and written documents of leadership activity around assessment and improvement, created a more robust description of school leadership assessment practices and outcomes.

Observations

Since engagement with student assessments and data is an ongoing collaborative process throughout the school year (Portin et al., 2009), I focused my observations on the leadership team meetings at Crossroads Elementary as they occurred during their cycle of inquiry. They met weekly during the 2012-2013 school year to discuss student achievement and instructional improvement, and the team consisted of the principal, a literacy specialist, and an ELL specialist.¹⁰ I attended these meetings on a weekly to biweekly basis and observed and recorded nine of these weekly meetings from January to June. I collected information about their assessment practices —such as their uses of evidence (e.g., documentary assessments such as particular test scores) and their discursive assessments (e.g., identified problems or issues within the school that are causes for concern or barriers to achievement)— and tracked the details of their decision-making activities. These meetings were convened to address a variety of issues such as collectively interpreting high-stakes and interim summative assessments, as well as planning, assessing, and implementing ongoing instructional practices, interventions, and testing.

¹⁰ Pseudonyms were used for all participants

During the course of this work, I added observations of staff meetings, professional learning community (PLC) meetings, and site-council meetings (a school-wide leadership team) with representatives of each grade-level; plus, interviews with school staff participating in site-council leadership in order to provide descriptions of instructional leaders' assessment-focused problem-solving processes and school leadership assessment practices and outcomes. During these meetings I took an "observer as participant" stance that allowed me to "observe and interact closely enough with members to establish an insider's identity without participating in those activities constituting the core of group membership" (Adler & Adler, 1998; as cited by Merriam, 2009, p.124).

I audio recorded and transcribed their meetings in order to effectively capture the specific problem-framing and discourse surrounding instructional improvement and their student population. I took fieldnotes and collected meeting documents to be able to track and document the ways in which these instructional leaders interacted with each other and with sources of data in non-verbal ways, such as facial expressions, using data displays or handouts as reference points in their group discussion, and other non-verbal actions in their practice. These fieldnotes and documents were then used to provide a more descriptive narrative to the audio transcriptions.

My observations of their discourse surrounding assessment practices within an embedded context provided real-time examples of problem-framing and problem-solving surrounding a particular problem of practice. My analysis of a particular meeting's topic and discourse allowed me to employ a "nondirective" interview strategy,

and structure the follow-up interviews around a problem of practice they are currently engaging with to answer my questions surrounding motivation, value systems, norms, and concepts without directly introducing the topic (Wolcott, 1995). Therefore, the collection of data surrounding inherent, discursive and documentary assessment activity occurred in tandem with observations and follow-up interviews.

Interviews

I conducted two 45-60 minute semi-structured interviews—at the beginning and end of my data collection— with the instructional leadership team members. This allowed me to examine their understanding of instructional leadership, and their role in assessment, perceptions and planning activities specific to Latino and ELL students. My first baseline interview protocol covered such subjects as a) their leadership and background characteristics; b) viewpoints and perceptions of state-specific accountability and ELL policies; c) personal beliefs of how high-stakes accountability shapes their own and others' practice; d) specific tactics that they utilized to respond to accountability requirements; and e) contextual factors within their campus that they felt affected the success or failure to reach their accountability requirements.

The second interview protocol administered at the end of the school year asked questions aimed at identifying shifts in the leaders' thinking surrounding problem-formation, the solutions generated, and changes in their conceptions of their instructional leadership, assessment practices, and opportunities for school improvement. These interviews also allowed me to examine my own interpretations and

assumptions surrounding data use for instructional improvement and probe further about specific decisions and dynamics observed during the course of data collection, serving in part as a “member check” of the observed practice (Merriam, 2009, p.217).

Additionally, I identified two teacher leaders during the course of this work who participated in PLC meetings, served as grade-level representatives within Site Council, and who were involved in the School Improvement Planning for the upcoming year. I conducted one 60-minute interview with each teacher at the end of the school year to help elaborate on key issues that were identified during the course of the school year and to provide an external perspective of the instructional leadership team practice. Similar to the end-of-year interviews with the instructional leaders, these semi-structured interviews also examined their perceptions of the school’s context, motivation, and capacity to impact effective practice issues that were or were not being addressed by the school leadership.

Finally, I conducted interviews with school leaders at the end of the PLC meetings they facilitated to interpret and clarify the assessment activity within the meeting, allow the instructional leader an opportunity to reflect on their practice, and track their understanding of the affordances and constraints to their practice over the course of the observation. These interviews served as a member check for my interpretation of the PLC meeting. All interviews were recorded and transcribed anonymously with approval of the subject.

Documents

Within this work, I collected current and historical documents that referenced the ongoing assessment, student tracking, collaboration, and planning materials (such as agendas, data displays, etc.) used by the instructional leadership team. I additionally collected both the 2012 and 2013 school improvement proposals and took pictures of school improvement planning displays that were co-created by school leaders and staff, such as brainstorming and prioritization posters. Beyond providing a more descriptive narrative to my observations, these documents—particularly the data displays and handouts—were crucial in my ability to track micro-interactions within data use practice. For example, many times participants would reference student or grade level passing percentages by number only or would refer to a student or students by their highlighted color as an in-action shorthand as they interacted with the data display or handout. As such, these documents served as important reference points in which the discourse and dynamics of decision making were enacted by participants.

Approach to Analysis

Interview, observation, and documentary assessment analyses were used to understand both the embedded context of instructional leadership practice and how documentary assessments can shape organizational opportunities to learn (Miles & Huberman, 1994). The analysis of this study was conducted in three distinct phases: 1) utilizing jottings and fieldnotes to guide post-meeting reflection on problems of practice and contextual influences, and later processing jottings, fieldnotes, and documents with transcribed meetings to create a detailed account of data use practices; 2) followed by

coding and creating analytic memos with interview and meeting data for overarching themes that characterized the tensions between their accountability context, ELL policies and curriculum, and data use practices; and 3) finally, identifying patterns and generating claims using matrices of micro-interactions within data use in practice.

Phase 1: Observing, Processing and Describing

Data Use Practices

In order to delve into how leaders and staff understood their context and the uses of data to inform their practice, I utilized aspects of ethnographic participation within my weekly or biweekly visitations to Crossroads Elementary. While there was not full immersion within the school, by regularly visiting leadership meetings, and then expanding my observational reach by invitation, I was able to follow the instructional leaders "...to see how they respond to events as they happen and experiencing for oneself these events and the circumstances that [gave] rise to them" (Emerson, Fretz & Shaw, 2011, p.3).

As mentioned previously, the first phase of analysis happened concurrently with observation data collection to provide context and questions for post-meeting interviews. During observations I took jottings of scenes, events, and interactions of the meetings and used these jottings to develop additional post-meeting questions surrounding problem-framing and negotiation of accountability within their particular context. For example, I noted how problems of practice were identified and framed in order to later clarify my understanding of the changes stemming from their team discourse during data

use practices. This practice helped me to avoid imposing exogenous meanings and appreciate how local meanings and concerns were felt within their context (Emerson, Fretz, & Shaw 2011). Later, these jottings were processed alongside meeting transcriptions and collected meeting documents to create a richer and more detailed account of the various moves and micro-interactions surrounding data use in practice. Jottings within meetings where audio recording was not prudent (such as all-staff meetings, or an all-day ESEA reauthorization working meeting) were turned into descriptive fieldnotes.

Phase 2: Coding, Developing Matrices, and Identifying Key Themes

The next stage involved conducting a close reading of a selected sample of the corpus to identify descriptive patterns and overarching themes to create inductive codes from the participants' words and concepts (Maxwell, 2005). This data was then coded (utilizing Atlas.ti qualitative analysis software) using a "two-level scheme" of broad categories and deductive codes based on my conceptual framework (Miles & Huberman, 1994).

First round coding. To create the initial codebook, I focused on transforming the interview and observation text from descriptive "open codes" into larger codes derived from my conceptual frameworks and identifying alternate coding trends (Emerson, Fretz, & Shaw, 1995, p.152). This was in order to "integrate the data into an explanatory framework" using the "ladder of abstraction" model outlined by Carney (1990) (as cited by Miles & Huberman, 1994, p. 91). Utilizing Atlas.ti's Network Views

tool, I examined and categorized these generative codes along with codes derived from my conceptual framework in order to create descriptive and thematic connections with grounded codes and analytical codes from my framework (Glaser & Strauss, 1967). This first round of descriptive coding helped to identify and describe the context of data use (such as forms of assessment, evaluation, and accountability context; and organizational routines) alongside aspects of inherent assessment (such as perceptions of capacity for data use; understandings, perceptions, and utility of available data sources; and capacity, beliefs and motivation for instructional improvement). Throughout this process, I wrote analytic memos and used data displays to collate evidence, reflect on emerging themes, and further develop key theoretical propositions (Miles & Huberman, 1994).

Analytic memos. A series of initial analytic memos was used to capture preliminary analysis and integrated to “clarify and link analytic themes and categories” (Emerson, Fretz, & Shaw, 1995, p.143). The analysis of analytic memos draws from “grounded theory” approaches to develop, rather than confirm, deeper understanding of the policy/practice praxis within school leadership. Major themes that emerged from assessment practices of instructional leadership teams served as illustrative cases of common problems within Latino and ELL-serving schools and were organized into matrices based on my analytic questions to facilitate the interpretations and make comparisons (Miles & Huberman, 1994) to problems of practice, perceived affordances and constraints to practice, and other shared understandings of the impact of accountability on Latino and ELL serving schools. This early analysis helped to develop

and track connections during the coding and analytic memo-ing process that initially focused on the following main topics: 1) the impact of state and district accountability mandates on leadership and teacher practice and data use; 2) the interaction of ELL support policies and programs and data use; and 3) the extent to which data use practices were responsive to community-based needs. This first level of analysis was also important in identifying common problems of practice identified by school leaders, so that the second level of discourse analysis could draw conclusions about the joint activity mechanisms and constraints and affordances that led to particular decisions (or not) about instructional reforms or solutions, and/or potential changes in ways of thinking about their school's problems of practice.

Once coding and memo-ing was completed, I triangulated my findings through multiple data sources collected throughout the year (i.e., preliminary and post-meeting interviews with leaders, interviews with key teacher informants, and documentary assessment collection). I also incorporated a form of member check by reviewing some key themes from the observations of the meeting with the school leaders in their post-meeting interviews and through end-of-the-year interviews with the instructional leadership team.

Phase 3: Tracking Opportunities to Learn from Data Use in Practice

The next stage of analysis sought to explore some of the tensions and issues surrounding data use practices within this context by discursively exploring instances of

data-focused inquiry within organizational routines for data use. Given the nature of the accountability context of this school, data was often used as a basis for framing persistent problems of practice, for tracking outcomes of intervention strategies, and as general feedback on student progress. During such instances, educators are required to engage in some sort of inquiry and analysis of the presented data and collectively interpret and problem solve identified issues. This stage of analysis helped me to consider the ways in which educators interpreted and used data and other forms of evidence to interpret and address identified problems of practice, and form decisions for instructional improvement.

To examine instances of micro-interactions surrounding data use in practice and their potential relationship with decision outcomes and shared meaning, I drew on analytic techniques employed in Wells' (1996) organization of spoken discourse, and Horn's (2007) identification episodes of pedagogical reasoning to identify relevant episodes of *assessment talk*. Drawing from activity theory and systemic linguistics, Wells' (1996) analysis of talk proposes that the "exchange of meanings between participants in order to perform some function(s) in a specific situation" mediates activity goals and outcomes (p.77). In the case of data-informed decision making, it is not just the data that mediates meaningful action but also the aspects of collective discourse. Horn's (2007) conception of episodes of pedagogical reasoning allowed me to extend my conceptualization of "data talk" to also include "...moments in teachers' interaction in which they are describe issues in or raise questions about teaching practice that are accompanied by some elaboration of reasons, explanations, or justifications"

(p.46). To identify episodes of what I call assessment talk, I searched within instructional leadership team and PLC meetings to find instances where problems of practice and documentary assessments—such as test score data displays—triggered extended discourse and sense-making surrounding evaluation and planning for instructional improvement efforts. My decision rules for the beginning of these episodes started with conversations sparked by the “boundary object” of assessment data, standards, problems of practice surrounding evaluation and testing, with the completion of the episode ending with the educators moving onto a different “boundary object” or non-assessment focused discourse, such as budgeting or scheduling.

Within my analysis of evidence and assessment use in practice, I identified episodes of assessment talk as a starting point for examining micro-interactions of data use in practice. I then organized these episodes of assessment talk into an effects matrix (Miles, Huberman, & Saldaña, 2014) to explore the identified problems of practice and tracked: a) participants; b) types of evidence and data (forms of documentary and inference assessments offered in discussion) utilized during the data use practice sense making; c) a summary educator responses, arguments, points of agreement and disagreement; and d) shared understandings of the situation or issue, and if any decisions were made during this process, if any. In addition, I recoded these episodes of assessment talk to track the extent to which educators moved through different stages of data-driven decision making in their collective discourse: starting with *data*, providing forms of evidence to create *information*, and then using this information to create actionable *knowledge* for instructional improvement (Ikemoto & Marsh, 2007). Finally,

I coded the forms of inherent assessments made by practitioners to the extent to which these comments identified relationships and interactions among teachers, students, and content, in the various environments of schools (see Cohen, Raudenbush, & Ball, 2003; see also Ball & Forzani, 2007). I also tracked my own perceptions and explanations of the assessment talk episode in context of the study.

I then compared these episodes of data talk across different contexts of data use (i.e., instructional leadership team meeting, PLCs) to look for any differences or similarities in process and outcomes. I also looked for patterns of documentary, inherent and discursive assessment for themes of data use in practice. Finally, I also explored outcomes by type of evidence utilized to explore the extent to which the types of evidence source used (i.e., data displays, personal understandings of the situation) contributed to particular outcomes of data use practices.

Strengths and Limitations of the Study Design

Case study research provides both affordances and constraints to inquiry into situated data use practices. Due to my interest in exploring the intersections of ELL policies and accountability contexts on local data use practices, case study design “offers a means of investigating complex social units consisting of multiple variables of potential importance in understanding the phenomenon” (Merriam, 2009, p. 50). Given all of the potential variables and influences to situated data use practices, case study research allowed me to focus on a more holistic account of one particular school’s data use practices. While case study brings up issues of generalizability to other school sites,

this particular study of situated data use practices within Latino and ELL serving schools provides what Davies (1999) refers to as “analytic generalization” or “theoretical inference” wherein “the conclusions of [a qualitative study] are seen to be generalizable in the context of a particular theoretical debate rather than being primarily concerned to extend them to a larger collectivity” (p.91; as cited by Eisenhart, 2009, p.59). Thus, this study may provide theoretical insight into how context informs data use practices.

Within this particular context, it was important to focus on how an “...individual ‘becomes attuned to’ the affordance and constraints of a given context and, if successful, regulates her motivation and learning strategies to succeed there” (Volet, 2001, p. 425; as cited by Nolen & Ward, 2008). This case study design allowed me to help unpack personal understandings of affordances and constraints that educators practice, and consider how knowledge of these dynamics may serve as an important reflective tool of data use practices for instructional improvement within particular contexts. This site was selected for its potential to reveal something new or different about data use practices in this particular context in order to further refine it for extension into similar school contexts surrounding rapid Latino and/or ELL demographic change (Eisenhart, 2009).

However this case study is not without limitations. As mentioned previously, focusing primarily on instances of data use practices within leadership team meetings conceptually bounded exploring in depth how the data for instructional improvement was utilized within the school, and it ignores the contribution of other forms of evidence use throughout the school. Also, within these episodes of data talk, there is the potential to over extrapolate the contribution of the instructional leadership team as facilitator and

driver of data use on practitioner learning outcomes. For example, this study did not examine to what extent teacher leaders within their monthly constituent meetings used data to frame and explore problems of practice. Yet, my analysis did focus on how assessment practice within instructional leadership team meetings created opportunities for developing shared meaning for individual instructional leaders and the team as a whole, as evidenced by a historical analysis of discourse on assessment and follow up interviews proceeding team meetings. In addition, by extending my observations to PLCs, all-staff meetings, and site council meetings, I was able to triangulate findings surrounding how data and assessments were used with teachers throughout the school.

This analysis of instructional leadership team learning was only able to focus on a limited number of meeting observations of data use practices. Without a team of researchers to assist in data gathering, my inquiry was focused on data that I was able to collect within a shortened time frame and geographical considerations. Therefore, I focused on specific interactions between leaders, their followers, and the conceptual and physical tools of assessment, and analyzed micro-interactions of data use practices in order to stay within my own personal constraints and affordances as a researcher.

Finally, as the primary instrument for data collection, it is possible that despite methodological efforts to consider alternate meanings and representations of the case study through triangulation and member checks, my own assumptions, biases, and concerns surrounding the ways in which Latino, emergent bilingual, and low SES children and their families have been negatively categorized and portrayed through deficit orientations may color my interpretations. Therefore it was important to reflect

critically on my own identity as a Latina who has focused on exploring how Latino and emergent bilingual children have been chronically underserved by educators (Merriam, 2009).

As a daughter of an educator, it was extremely important for me to withhold judgment of resources or practices that seemed to encourage—or at the very least uphold— limited or deficit orientations and perceptions of low income students of color. For example, they used Ruby Payne’s *Framework for Understanding Poverty* as professional development for staff to understand their students, backgrounds. Payne’s notions of a “poverty culture mindset” could contribute to educators’ low expectations of students and their families and communities as primary sources of student failure (Bomer, Dworin, May, & Semingson, 2008). In instances such as this, I tried to assume positive intent and saw these forms of observations as sources of “inherent assessments” or understandings of the situation given available information. In her introduction to *Everyday Anti-Racism*, Mica Pollock (2008) explains that “many such acts taken in educational settings harm children of color, or privilege and value some children or communities over others in racial terms, without educators meaning to do this at all,” and she cites Don DeLillo to remind us that ultimately, “Everyday things represent the most overlooked knowledge.” (p.xvii)

Chapter 5:

Data-Focused Instructional Leadership and Perceptions of Assessment

“Fit” Within Collective Data-Use Practices

This chapter provides an overview of how an instructional leadership team along with their constituents interpreted and responded to assessment data, and planned improvement activities, in the context of their data use practices. In particular, it provides a glimpse at the variety of ways instructional and school leaders use data to inform their decision making and to improve instruction for their culturally and linguistically diverse schools. This chapter is not exhaustive of the entirety of collective instructional leadership or the use of assessments throughout the school—particularly within classrooms. But it does illustrate the ways in which multiple measures are used to inform school and classroom-level decisions to improve instruction within Crossroads Elementary, and how ELL support structures and accountability pressures can impact opportunities to learn within data use practices.

This chapter is divided into two sections, beginning with how Crossroads Elementary designed its ELL student and teacher support infrastructure to respond to accountability expectations of student achievement. I then examine how assessment routines helped to facilitate teachers' opportunities to learn during data focused decision making surrounding pedagogy, student support, and interventions. Lastly, I highlight the tensions surrounding problems of practice identified during data use practices, and

consider affordances and constraints to deep engagement with data and decision making within their particular context.

The Accountability Context and Organizational Structures for Student Improvement

The leaders' use of data within assessment practice can only be understood within the larger accountability context and the organizational structures created to respond to it, including routines that are specifically designed to encourage data-based solutions and practices. Designing student and teacher supports to respond to accountability mandates for achievement can be challenging for any school; yet, this can be particularly challenging for schools that have seen rapid demographic changes (Hopkins, Lowenhaupt, & Sweet, 2015). In particular, Hamman and Harklau (2015) identify the tendency for schools within these areas to apply "improvisational" methods as they respond to their ELL student support and language issues throughout the year. That is, many schools within areas not traditionally settled by Latinos (i.e., new areas of Latino population growth, such as within the south or particular rural areas) often lack bilingual educators, and their ELL students may be mismatched with the school and district's expected learner profile for tracking student achievement (Bruening, 2015). That could cause tension in designing programs and interventions for ELL student improvement. Therefore, in this section, I describe Crossroads Elementary system of student supports and interventions for ELL students, and outline the variety of

measurements they used for placement of students within these programs during organizational routines.

The Crossroads Continuum of Student Support Structures

While the district supported a Transitional Bilingual Program for their English Learners, there was no set language curriculum for ELL students in the district or at Crossroads Elementary, despite the growing number of ELL students within the district and particularly within this school¹¹. Therefore, Crossroads Elementary created its own continuum of supports for ELL students based on their English language development, performance on assessments, and their use of the district-supported Walk to Read model, which grouped students based on ability level for reading instruction (See Table 1). Officially, Crossroads Elementary provides their growing population of ELL students with a Content-Based ELL program. The principal said this program would allow all students to get a “crack at grade-level instruction curriculum,” with an expectation that all teachers provide differentiated instruction. However, this program was not without its own limitations. Rebecca, the ELL Support Specialist, explains:

In our district we have a content-based ESL program, which really means classroom teachers are given the job of developing language in their classrooms as they teach math, reading, science, and social studies. But without giving them the skills and the support to do that, it doesn't always happen.

¹¹ There was, however, one dual language elementary school within the district.

Thus, a majority of professional training and development for ESL strategies relied heavily on individual teacher interest and on informal professional development with the ELL Support Specialist Rebecca.

Within this general structure of ELL programming, Crossroads created a variety of student supports and interventions throughout the school day. As part of the legacy of Washington State’s K-12 Reading Model¹², grades 3-5 implemented the Walk to Read

Table 1. Crossroads Elementary Continuum of ELL Student Supports

<i>Support</i>	<i>Description</i>
Content-Based ELL	Students are taught English using English instruction of grade-level material. Grades K-2 and 6th all do homeroom reading and math. Grades 3-5 do Walk to Read for Reading and homeroom math.
Walk To Read	Implementation of the Walk to Read model involved grouping students by ability level within separate classrooms, each with either a “benchmark,” ELL student, or Special Education focus in grades 3-5 for a 90 minute block of reading instruction. Within the ELL student-focused leveled Reading group, ELL-specific Instructional Assistants provided targeted instructional support through in-class.
Enrichment and Intervention (E&I)	During the 30-minute Enrichment and Intervention (E&I) block, homeroom teachers work on targeted student needs in Math and Reading, sometimes with the help of instructional assistants.
School-Day Tutoring	In 20 min blocks in the afternoon, Instructional Assistants use pull-out tutoring groups for targeted student needs in grades 1-6 based on teacher assessment. Level 1 students work with the Literacy and ELL Support Specialists during these times.
After School Programming	After school programming for targeted students: - ASPIRE 3000 ESL Math Group - ACHIEVE/Kid Biz Reading Group - Washington Reading Corps - Local Church-led programming - MSP Blitz (1 month only)

¹² Developed in 2005, and since revised with the arrival of the Common Core in the 2013-2014 school year.

program in order to group students based on their ability level. For example, MSP “benchmark group,” a predominantly ELL group, and the “highest need” learner group, often fell below standard and/or with some Special Needs. During their 90-minute block of reading instruction, the students “walk” to their assigned classroom for reading, often with a different teacher in their grade level. Within two of these classrooms, instructional assistants (either ELL or Title 1-funded) are placed within these rooms to provide in-classroom support for learning. It is within these classrooms that ELL language support services are implemented, as the Literacy Specialist explained in an interview:

We know what is good for those ELL kids is good for all students, but you will see almost all teachers implementing a lot of language strategies. So we have learning targets that are to be posted...or if they are not posted, at least to have language targets for each lesson as well...And then the GLAD strategies or the SIOP strategies, you’ll see that implemented across the board in all the grades, in all the classes.

She comments that while some teachers do better than others, based on their training, she feels that there is a “real effort” made within the classrooms. For example, working on oral language within classrooms was a school-wide goal during the school year. Within Crossroads, each grade level was given the option to decide if they wanted to participate in Walk to Read. Therefore, only grades 3-5 participated in Walk to Read, with K-2 and 6th grade having math and reading in their homeroom. Therefore, the type and consistency of ELL language support and curriculum for an ELL student depended

on a student's homeroom teacher or Walk to Read teacher's training and experience with those strategies.

Throughout first and sixth grades, they also had a 30-minute Enrichment and Intervention (E&I) block for homeroom teachers to work on targeted student needs in reading and math, sometimes with the help of instructional assistants. They also had a series of 20-minute afternoon tutoring blocks for pull-out instructional support for targeted groups with instructional assistants, and one-on-one tutoring with support specialists with the highest need students. While these students work with instructional assistants and specialists, only the bottom 5 percent of students get to work with the ELL Support Specialist on a regular basis.

Finally, several students were targeted for after school programming. Academic-focused programming involved the ASPIRE 3000 ESL Math focused group, the ACHIEVE or Kid Biz computer based Reading support program, and a volunteer-based homework and academic support group implemented by a local church group¹³. Lastly, as one of the leadership team initiatives, a one-month MSP Blitz afternoon program was created to focus on targeted students for test prep and a focus on reading and math. The majority of these after school programs were implemented by instructional assistants, who received training and supervision by either the Literacy Support Specialist, Patty, or the ELL Support Specialist, Rebecca.

¹³ In addition, there was also programming provided by the Washington Reading Corps in History and Dance.

Testing and Assessment Routines

Crossroads Elementary's ability to identify, track, and monitor student growth within these programs and interventions relied heavily on the ways in which they utilized multiple measures for decision making throughout the year. Mountain Valley School District provided a series of interim and end-of-year summative assessments for data driven decision making for instructional improvement, as well as language assessments for identifying and providing services for ELL students (See Table 3). In 2012-2013, the MSP (Measurement of Student Progress) was the Washington state test for 3rd through 6th grade, administered once a year in May, and provided a benchmark for teachers in Math, Reading, Writing, and Science when they received their scores late in the summer. Performance on these benchmarks¹⁴ was the basis by which the state held districts and schools accountable for student performance through Annual Measurable Objectives.

To measure their own growth within schools, the district provided two additional interim tests, the Measures of Academic Progress (MAP) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), to be administered three times a year. The MAP test is a computer-based assessment program aimed at monitoring student growth in Math and Reading, while providing information for teachers to improve student learning. The DIBELS test provided measures for assessing the acquisition of early

¹⁴ They are also used for school, district and student accountability. At the school and district levels, state tests are used to help determine Annual Measurable Objectives (AMOs) and close proficiency gaps as a part of the Elementary and Secondary Education Act.

literacy and reading comprehension skills, and also to provide teachers information about how to adjust instructional support to meet the needs of students. These assessments served as “normed” data for the school to use to track student growth and to compare across schools within the district.

Additionally, the Washington English Language Proficiency Assessment (WELPA)—administered annually in February—was given to students either previously identified as ELL, or to newcomer students who identified a language other than English spoken within their home for placement. The WELPA assessed growth in English language development and determined eligibility for English Language Development (ELD) services, commonly referred to as ELL or ESL services, within Crossroads Elementary. During the 2012-2013 school year, 390 out of 610 students were eligible, or roughly 64 percent of their students. The majority of students were at Level 2 (Intermediate) but often flat lining at Level 3 (Advanced) and unable to advance to Level 4 (Transitional) often “until high school,” the ELL specialist noted.

Together, these tests provided multiple measures of data for staff and school leadership within Crossroads Elementary to identify and target students for support services and interventions throughout the year¹⁵. These decisions were made in a variety of organizational opportunities for collaboration and professional learning.

¹⁵ In this year, the 4th grade was also selected to participate for the NEAP, however that was not a focus for data inquiry for this school.

Organizational Routines for Data Use

Crossroads Elementary provided many opportunities for collective instructional leadership, decision making, and professional learning for student improvement, all informed to varying degrees by data, especially the data emerging from the assessment measures described above. These opportunities for collaboration often covered a variety of management, training, and coaching topics, as well as the prioritization of student resources. Examining multiple measures of student achievement was part of the collaborative culture within the school.

Data-based deliberations in the Instructional Leadership Team. The instructional leadership team (ILT) was comprised of the Instructional Support Specialist, Patty, the ELL Support Specialist, Rebecca, and the school principal Edward. Charged with planning and managing whole-school programming and interventions, this team also handles the day-to-day management issues through a variety of feedback channels throughout the school, such as staff meetings or constituent meeting feedback. During their weekly morning meetings, the ILT often took the opportunity to look at whole school or whole grade level data to track progress and to make decisions surrounding the efficacy of interventions, such as after-school programming. During these data inquiry sessions, they would take these forms of feedback to plan and make decisions related to instructional improvement.

These leaders also gathered and made sense of data separately—often through the use of their own data displays—to aid in collective sense-making during their hour-

long team meetings and within their respective professional learning communities (PLCs). For example, Edward describes his process of individual data analysis:

First I look at grade-level growth. Then I try to compare it to past years. And then I try to scan the individual growth, and I try to, in my own mind, try to think of how we are serving the kids who did poorly, and how are we serving the kids who did well. And then, are we serving those kids in the best way possible?

Patty sees looking at data in this way as a large part of her job in order to “communicate what we’re doing well and what maybe we don’t need to do as much” to the rest of the staff. As the first line of defense, these leaders actively interpreted and created displays of student interim assessment data for their staff, and external stakeholders.

Data-focused conversation in Professional Learning Communities (PLCs).

With the support of the district and additional planning time during the day for teachers, the instructional leadership team was able to set up a PLC system on Tuesday mornings, which Edward feels helped to unify “what goes on classroom-to-classroom at specific grade levels.” Instructional leadership of these groups was handled by each member of the instructional leadership team: Patty on Literacy, Rebecca on ELL Language and Strategies, and Edward with Math, Science, and sometimes grade-level management issues on a weekly basis rotating with other topical meetings (See Table 2).

Table 2. PLC and Staff Meeting Rotation Schedule

Meeting Type	Facilitated By	Topic Areas to be Covered
Staff Meetings	Rotating Instructional Leaders	<ul style="list-style-type: none"> - Differentiation - Science - Common Core Standards - Language Arts - Character Traits - Teaching with Poverty in Mind (Jenson) - TPEP - Emerging School Requirements
Constituent Meetings	Rotating Grade Level Leaders	<ul style="list-style-type: none"> - Design, assign interventions and enrichment - Collaboratively plan lessons and look at student work - Align curriculum, instruction, pacing - Align instruction with Common Core Standards
Literacy	Patty	<ul style="list-style-type: none"> - Design, assign interventions and enrichment - Student data analysis - Literacy professional development - Tier 1 reading and math interventions
ELL Support	Rebecca	<ul style="list-style-type: none"> - Developing English Language Standards across curriculum - Ongoing professional development in ESL strategies - Tier 1 ESL reading and math interventions - Student data analysis - Language targets
Math, Science and Management	Edward	<ul style="list-style-type: none"> - SIP implementation - Science and math alignment - Grade level management
Note: List of topic areas is from a school planning excel sheet. Not listed is a Technology and School Environment PLC.		

In addition to a PLC focused on technology with a district specialist and one on the School Environment where they plan monthly assemblies, and monitor student attendance and discipline data, they also have their own grade-level constituent meetings where they are able to set their own agenda to work on things like rewriting assessments, examining common core standards, aligning practice and instruction, etc. As such, depending on the focus of the meeting, there were often formal and informal professional development opportunities for staff. This time was also when staff were able to dig into classroom and student level data for decision making surrounding classroom and targeted interventions for their students. Due to the range of PLC

programming, each grade level is only able to meet with each member of the ILT twice a year.

Opportunities for staff communication and collaboration. Monthly staff meetings, Site Council, and ESEA reauthorization teams also created opportunities for whole staff and grade-level leadership communication surrounding whole school and grade level achievement. Staff meetings provided key opportunities for whole staff communication surrounding points of business and interventions, as well as providing opportunities for professional development, such as learning about students through a common book. Site Council provided rotating grade level membership for staff to participate in day-to-day business and planning around budget oversight and school improvement priorities. Meanwhile, the ESEA Reauthorization team, made up of volunteer staff and the ILT, helped to shape their school improvement plan for the following school year. Similar to staff meetings, student data was often explored in the grade-level aggregate. While not central to decision making, as within instructional team meetings or PLCs, forms of data and data displays were used as part of their culture of utilizing multiple measures for student achievement.

Thus leadership and staff within Crossroads Elementary utilized a variety of data use practices during their organizational routines for decision making and professional development. Yet, given the complexity of their accountability status as a “Priority School” and their growing ELL population, it is important to delve into how these multiple end-of-year and interim assessments were processed and utilized for decision making surrounding targeted intervention and tracking student achievement in these

different spaces. In particular, how they labeled and tracked students, what forms of evidence and data they relied on, and what problems of practice and tensions surrounding instructional improvement surfaced during data use practices.

Local Meanings in Data-Use Practices

The cycling of interim and end-of-year assessment data triggered a variety of feedback and sense-making opportunities for school leadership and staff to consider their curriculum, pedagogy, and interventions for instructional improvement. Yet Crossroads Elementary was like many schools that were “data rich and information poor”—or, more appropriately put, “insight poor”—in that, despite the proliferation of multiple forms of assessment data to help educators make informed decisions surrounding their practice, much of the information provided by these forms of evidence was confusing, contradictory, or ill-suited to their student population. The situation left many educators drawing from their own knowledge bases to “fill in the gaps” left by assessment data in order to help them make sense of student achievement outcomes and define their role in instructional improvement.

What follows is a snapshot of data use practices in which school leaders and staff engaged to help them make sense of assessment data. It shows how they utilized multiple measures to shape organizational goals and provide targeted assistance to struggling students, and the variety of challenges to effective data use and instructional improvement as seen by leadership and staff. Chapter 6 provides a more detailed look into data use in practice to explore what forms of inherent, discursive, and documentary

forms of evidence were used within episodes of data talk, and how specifically their use shaped the leaders’ thinking and actions.

The “Mountain Valley Way”: Use of Multiple Measures for Student Tracking and Interventions

In order to “conduct, monitor, and evaluate formative and summative testing results...[and] analyze human, financial, and material resources for efficiency and effectiveness” (2012-2013 SIP Plan), the instructional leadership team needed to find a way to synthesize the variety of MSP, DIBELS, MAP, end-of-unit assessments in Math and Reading, content based assessments, and fluency screenings into actionable knowledge for decision making. Within this study, the analysis and dissemination of assessment data into data displays for collective decision making created a variety of

Table 3. Local Framework for Organizing Student Levels in Crossroads Elementary

MSP/HSPE	WELPA	DIBELS	MAP	Enrichment Intervention	K-12 Model for Instruction in Math and Reading	Standard Level	
Level 4 Exceeds Standard			Level 4 RIT= or > 95 percentile	Level 4 Beyond Grade Level Expectation	Core Curriculum and Enrichment and/or acceleration	Meets Standard	
Level 3 Met Standard	Level 4 Transitional	Level 3 Low Risk Benchmark	Level 3 RIT 50 th – 95 th percentile	Level 3 on grade level goal is approximately 75-80%	Core Curriculum with scaffolding, intervention, and/or enrichment		Formerly Tier 1
Level 2 Below Standard	Level 2 Intermediate or Level 3 Advanced	Level 2 Some Risk Strategic	Level 2 RIT 25 th -49 th percentile	Level 2 Below Grade Level Goal is approximately 10-15%	Core Curriculum and 30 minutes targeted intervention	Does Not Meet Standard	Formerly Tier 2
Level 1 Well Below Standard	Level 1 Beginning or Advanced Beginning	Level 1 At Risk Intensive	Level 1 RIT <24 percentile	Level 1 Well Below Grade Level is <5%	Core curriculum plus 60 mins targeted intervention OR replacement curriculum		Formerly Tier 3

“unofficial” labels to refer to students’ academic and English proficiency levels.

Systems of tracking and student labeling. During instructional leadership team meetings, they often utilized what principal Edward called the “Mountain Valley Way,” a localized, unofficial name for a form of shorthand to label and categorize students into “Levels” that utilized the various assessment measures at their disposal alongside the tier-based system within the previous Washington K-12 Model for Instruction in Math & Reading (See Table 3). Given their school’s limited resources for interventions, Edward explains how this form of student labeling helps them to make decisions surrounding target student supports:

We refer to it in a pretty, I would say, unrefined way. Level 1 kids are kids who probably are not going to make grade level standard, despite many interventions, or at least not immediately...Level 2 kids, with some extra interventions and responding to those interventions the way you want them to, have a good chance of [meeting standard].

Within this framework of student achievement, a student may be labeled a “Level 1” student if they score “Well Below Standard” on the MSP, if they are labeled a “beginner or advanced beginner” in English Proficiency on the WELPA, score in below the 25 percentile for the MAP *RIT* score for achievement and growth over time, or are in the bottom 5% in meeting grade level goals.

As the former Success for All coordinator within the school, Patty took charge in managing student assessment data into the multi-variable data display of summative assessments that would ultimately be referred to as student “Levels” within the school.

Since these assessments arrived at various times throughout the year, she organized these assessments and evaluated student levels differently for decision making:

Normally at the beginning of the year, [I] give the MSP the most weight, and then the MAP score. And you know we have all of the guidelines, all the cut scores for everything. And then the DIBELS is the last one that I look at to see. And then if the kids are really grey, then I go to the classroom assessments, the most recent assessments and take a look.

From there, she created a colored spreadsheet to be used for decision making within the ILT and PLC meetings. As the trimester MAP and DIBELS data arrives, she adjusted each student's level or "color" based on this new assessment data.

I observed during ILT meetings and interviews that individual assessment terms were sometimes used interchangeably as a heuristic for talking about particular groups of students; for example, referring to "Tier 3" or "Level 1 kids" in the same sentence. Due to the color coding surrounding these student levels, students were also often referred to by color, such as "pink" (level 1, well below standard), "yellow" (level 2, below standard) or "green" (level 3 or at benchmark) kids during ILT team meetings, interviews, and sometimes within data-focused PLCs with staff. At other times, an accountability-focused vernacular would be used to refer to students as "bubble kids" or kids just below standard (Level 2 or "green kids") in terms of targeting them for meeting standards. Taken together, the use of these terms created a usable shorthand in the face of a complex assortment of summative assessments for decision-making.

Targeting students for interventions. During instructional leadership team and PLC meetings, multiple forms of assessment and data displays found within the Mountain Valley Way were used to identify and target students for interventions, and to evaluate their effectiveness. This was particularly important due to the high need but limited resources for targeted support such as instructional assistants and after-school programming slots. Therefore categorizing students as Level 1, 2, and 3 allowed leaders to focus on particular groups of students instead of making decisions on a case-by-case basis. For example, Edward explained how Level 1 students may receive targeted one-on-one support with the Instructional Support or ELL Specialist and tutoring interventions throughout the day, while Level 2 students may only get extra support during their leveled Walk to Read classroom so they would be targeted for after school programming.

Within instructional leadership team data-focused meetings, the spreadsheet of Leveled students was used to make initial placements in leveled reading groups for Walk to Read. After this, the instructional leadership team members worked with teachers during PLCs to pinpoint Level 1 and 2 student “skill gaps” and find ways to provide interventions to address those gaps so that, as Patty explains, “we’re always moving them to be green kids...the pink up to the yellow, and the yellow up to the green.” For example, Patty provided a display of DIBELS scores during her PLC with the 1st grade team in order to walk them through writing up their own planning sheet to form student groups to work on aspects of literacy indicated by DIBELS indicators, such as Phoneme Segmentation Fluency. Within this PLC, teachers appropriated this

language of student labeling by referring to a “yellow” student as she inquired about their test score.

Lastly, outside of the regular school-year student support and intervention programming, the instructional leadership created an “intense four-week skill-building tutoring plan for selected low income/SPED students [that] would enhance their ability to meet grade level standards,” or the MSP Blitz as it was called during the school year. Driven by the ILT, they took on the responsibility in identifying what they often called “bubble kids” for targeted MSP testing preparation.

“Doing” Data: Challenges to Utilizing High-Stakes and Interim Summary Assessments for Decision Making

Other than within PLCs, summative student assessment data was primarily categorized and used as a working heuristic for school-level decision making by the instructional leadership team. Perhaps due to their position as instructional leaders, this use of available assessments as an overall heuristic may also be due, in part, to their beliefs surrounding the limitations of these forms of assessment to inform their practice. As a “big believer” in using multiple forms of assessments for decision-making, Patty felt that each form of data only provided one part of the picture:

We don't get results back from the MSP until the end of the summer. So the MSP results...[don't] really give us any good data for informing instruction. In fact, very few of those tests give us any good data for informing instruction. Sure, if they are low in fluency we can do

something to boost up their fluency instructional practice, but fluency's not all that reading is. So, there again, the multiple measures help me.

While these multiple measures provided more than one indicator to make decisions to support student learning, it was unclear how these high-stakes tests provided more "bang for their buck" in terms of explanatory power for instructional planning and direction for decision making. -Overall, it was felt that, in general, neither the MAPS or DIBELS provided accessible and usable data for decision making. For example, the instructional leaders commented on the lack of alignment of available assessments within their team and site council meetings. Within these meetings, leadership and staff alike noted issues within their data use practices surrounding these summative assessment tools' lack of alignment, connection to formative assessment, and ability to plan and set clear instructional goals throughout the year.

Lack of alignment. The ILT team felt there was an overall lack of alignment and opportunities for backwards curriculum planning based on these interim tests, or that the test results came back too late to inform instruction. As tests designed to track student growth throughout the year, these summative tests did not seem to align with standards within the MSP, which often provided confusion and frustration amongst leadership and staff. For example, comments during ILT team meetings, PLCs, and staff meetings indicated that the MSP, MAPS, and DIBELS scores did not correlate with each other, making it difficult to make sense of and evaluate instructional programming for student improvement.

With the upcoming adoption of the new Smarter Balanced Common Core Curriculum, Patty felt okay transitioning out of using the MAPS, despite its use within the school for showing growth, because she felt it did not provide “formative data to inform instruction.” Likewise, Edward agreed that the MAPS was not helpful for formative assessment, particularly, he felt, because it lacked alignment with the Common Core or their standard-based test. Similarly, he felt that, like the MAPS test, the DIBELS test provided only a narrow view of student improvement, yet was useful as a tool for students to track their own growth:

DIBELS more so with the younger kids. I think the oral reading fluency is another easy one for kids to hold on to and to measure their growth and be able to understand what we’re talking about, what we’d like them to grow. But I think there’s a lot of hit and miss with oral fluency, too. It doesn’t measure comprehension one bit...So it’s a predictor but it’s not the only predictor. You don’t want to box kids in with one piece of data...

Yet without clear assessments to provide actionable information for teacher practice, there may be misalignment between how teachers use this information alongside their own formative assessments within classrooms to make decisions for instructional improvement.

Lack of connection to formative assessment. Being able to draw from both summative and formative assessments of student learning is an important way to drive meaningful reform for improvement. Particularly with the emphasis on data and evidence found in summative tests—such as high stakes tests like the MSP— it is

essential that educators are able to make connections to the instructional core. Yet, within this context, Patty felt that they were still in the process of training some teachers in how to use their formative assessments as sources of evidence when interpreting summative assessment in data use practices, and that they needed assessments that would help them “dig down” into student outcomes. She saw glimmers of this within her PLC that centered on exploring DIBELS scores to create individual targets for students, when one teacher questioned the utility of particular indicators:

That’s a very good formative assessment because she’s saying this child’s mastering what I am asking them to do, oral reading-wise. They don’t have this one step. Why is that? And the best thing she said is, “How necessary is that? Do they really need that if they’re already here?” That’s a very good observation.

Additionally, a lack of connection between assessment and practice can create cognitive dissonance between the impact a teacher feels they are making within the classroom and what feedback summative assessments provide. Rebecca provides her insight and perspective on using tests as an indicator of student growth during an interview:

...When you make growth about a number, it doesn’t always mean anything. And it’s the only way we have to assess it, but I think we need to be really careful about saying something was effective or not effective based on just those test scores...And it’s like I know I worked hard, and I know that our kids made progress. And yet, it’s this one little number that—We need to be careful of that, because our kids aren’t just numbers and neither is our teaching.

She highlights a key issue surrounding the pressure to use data to drive decision making, but at the same time potentially undermining a teacher's ability to formatively assess student growth by providing confusing or misleading measurements of student achievement. Essentially she—along with the rest of the ILT— felt that the “classroom-based piece of teachers really being aware of where students are at [was] huge.”

As such, the instructional leadership team members felt tension to support teachers to make gains on these summative measures in order to show growth within the district, with using data-focused inquiry to support teachers to make formative assessments to improve instruction. During her feedback session on ELL student support services, one teacher commented to her that she would really appreciate more coaching and professional development for ELL classroom strategies, rather than having PLCs more focused on “testing type stuff” as it had evolved into. Torn between needing to support teachers and students with testing, and providing formal professional development during their limited PLC-time with grade level teams, the focus of the meetings often became working on examining student data and planning for their numerous tests per year.

Lack of clear goal setting. Lack of alignment between interim and high-stakes assessment benchmarks created a disjunction between focusing on overall student achievement and growth, and increasing performance on particular indicators with the interim assessments. Patty voiced frustration with their current assessment system:

And I was thinking about that in the sense that so often we have a goal here...[and] the district has something over here. And sometimes we get the cart

before the horse and then the district puts the brakes on, whoa, and we end up having to go in a different direction. So then we've wasted that time...

The misalignment of goals across state, district, and school assessments created additional challenges for data use practices aimed at improving instruction. As seen within this school, making data-informed decisions with forms of assessment that do not provide aligned and accessible forms of information for teachers leads to more shallow forms of data analysis that were perhaps not fully aligned with district goals for measurable growth. In essence, while using data pointed them in one direction, it may not have been the “right” direction as defined by state and district goals for improvement. Robin, a 3rd grade teacher, recalls how one trainer's visit to their school summed up this issue within the district:

He looked at us and said, “Oh, Mountain Valley School District. It's all Ready, Fire, Aim with you people.” Like not knowing your purpose before you go forward. And I think that has been—it causes a lot of anxiety, I think, within our building...And it's a strange time in education because of that, because now we have this golden ticket with the Common Core, and I think that's where a lot of my trepidation comes a bit because I've been in the game now a little bit and have seen things come and go with not a lot of thought.

Indeed, while there was hope and anticipation for a new system of assessment that would be better aligned to teachers' practice, it was also felt that it could also end up being more of the same challenges they currently face surrounding the use of data for student improvement. Yet, despite the identified limitations to available assessment tools

by staff and school leadership, an additional explanatory factor for issues with their research-based, normed interim assessments may have involved issues surrounding language acquisition for their emergent bilingual students.

Perceptions of Assessment “Fit” for Interpreting

ELL-Student Learning

Two primary tensions for supporting ELL student learning within Crossroads Elementary involved the use of the DIBELS assessment with emerging English language learners, and teacher and staff expectations for ELL academic language use, particularly during administration of the WELPA assessment. Throughout the school year, Rebecca, the ELL Support Specialist, highlighted on numerous occasions what she saw as problems surrounding the district’s use of the DIBELS for tracking student growth in literacy for ELLs:

I’ve voiced my opinion a lot of times about the DIBELS, especially the nonsense word piece, and just feeling like for ELL— I don’t know what else I would use...but I do have concerns about using the fluency only as an indicator for reading ability because I don’t think that’s a primary indicator for someone learning a second language, especially when you throw in nonsense words in there, as we’re really trying to teach them words have meaning.

Patty made a similar connection in that she noticed that their ELL students were able to get high scores on DIBELS, but that didn’t necessarily correlate with reading comprehension:

Often our ELL kids are really good at decoding. We've created some very good little decoders. But when it comes to comprehension sometimes especially when they are in 3rd and 4th grade, it's the comprehension piece they don't have.

Focusing on skill gaps identified by the DIBELS may be problematic due to the tendency to teach students "testing" skills in order to bolster achievement scores. For example, Rebecca brought up within an ILT meeting that she noticed within her one-on-one tutoring groups that her students were "reading and just continuing to read, even though they didn't know what a number of the words meant." She goes on to explain her student response to her observation:

...And then we said, so what is your strategy when you're reading and you don't know a word? "Well they tell us just to skip it." And then when they go take fluency, we say if you get to a hard word just skip it and move on, because it's all about words per minute. And so I think some of it's just establishing that purpose for reading when its not just for a minute-timed test.

This observation signifies a potential troubling tension between preparing students to be successful on a test, and teaching students lasting strategies for engaging with learning. While the ILT felt that this was a "mixed message" for students and that the proper times to use these test taking strategies should be better communicated. It was unclear if students would be able to differentiate between these contexts.

Conversely, Rebecca also identified issues surrounding the administration of assessments for ELL students. For example, during the proctor trainings with the instructional assistants for one of the Speaking rubrics from the WELPA, she got

pushback from one IA in scoring a 3 versus a 4, due to their individual perception of “that’s just how our kids talk.” Therefore, differing expectations surrounding ELL student academic speech may potentially skew student performance on the WELPA. Rebecca also highlighted how unfamiliar staff were with the variety of ELL accommodations for tests such as the MAPS or MSP, that may also potentially create discrepancies in student achievement scores. Thus it is possible that a combination of assessment “fit” for ELL students, alongside unfamiliarity with appropriate assessments and/or accommodations may have also contributed to leadership and staff perceptions of the usability of their multiple measures for decision-making.

**Summary: Context for Data Use Practices
in ELL-Serving Schools**

This chapter provided a broad picture of the range of available assessments, how leaders interpreted and utilized multiple measures for decision-making surrounding supporting student outcomes, and of the variety of barriers surrounding their use of data. In this ELL-serving school, leaders and staff faced the necessity of using data for decision-making and program evaluation as part of their ongoing assessment routines, despite any reservations they may have had surrounding their available assessments and the lack of any viable alternatives. Therefore, they used high-stakes and interim summative assessments to create a categorization system and heuristic shorthand in order to refer to and place large groups of students into targeted interventions. However, while the limitations of individual assessments were acknowledged by leaders, it is unclear if utilizing interim assessments for potentially high-stakes decision making is

beyond the scope and intended purpose of these often diagnostic and formatively-driven assessments.

Using their forms of student labeling and tracking, school leaders worked with staff during grade-level PLCs to provide support for teachers and students to show growth on their district's interim assessments. Within these opportunities for data inquiry, there were reported barriers to data use practices, such as the lack of alignment between high-stakes and interim assessment outcomes, little connection to formative assessment within the classroom, lack of clear and consistent goals within the variety of assessment indicators, and applicability for using these measures with their large population of emergent bilingual students. And, despite these issues, Crossroads still needed to show improvement on these tests, as the principal disclosed, "We still have to take the test. We're still accountable. We still need to show growth, or we are kicked for it." Therefore within the very limited amount of time the leadership team had one-on-one with grade level teams, they often focused on examining data and test preparation, instead of focusing on longer term investments within professional development in these PLC spaces.

Finally, the proliferation of assessment measures from the state and district to hold schools accountable alongside their particular context of developing capacity to teach and assess ELL student learning may have driven many of these barriers to deep data inquiry. Unable to deeply engage in data inquiry, data use practices were often focused on outcomes of student achievement, as opposed to inputs, such as curriculum and pedagogy. Often triggers or "problem spaces" for group discourse on student

achievement, these forms of data and evidence derived from summative assessments provided opportunities for the instructional leaders and staff within PLCs to draw from personal perceptions, understandings of the situation, and belief systems, to interpret student achievement outcomes. These discursive spaces surrounding student achievement and intervention efficacy provided opportunities for leadership and staff to bring in their own inherent assessments of student achievement to collective problem solving and decision making. However, when searching for the antecedents to Latino and ELL student assessment outcomes, practitioners tended to predominantly identify external influences to student achievement—such as poverty or community cultural factors— instead of engaging in discourse surrounding curricular or pedagogical improvement.

In the next chapter, I explore in more detail and in a more fine-grained way how forms of inherent and documentary assessments—such as test scores and data displays—were used within discursive spaces for data inquiry surrounding particular problems of practice. I consider the extent to which various forms of assessment are taken up (or not) as participants weigh different forms of evidence to create shared understanding or decisions surrounding instructional improvement, and the situational factors that may influence particular outcomes of data use practices.

Chapter 6:

Exploring Data Use in Practice and Sources of Evidence for Instructional Improvement within Collective Data-Focused Routines

The previous chapter provided a glimpse into the ways in which summative assessments were used by instructional leaders and staff to help identify, track, and label students for targeted support and instruction. It also highlighted the perceived barriers to data use practices that derived from the external accountability and ELL language policies structures within the district. This chapter takes a different analytic angle by examining in depth how data use practices shaped collective discourse for instructional improvement and planning, the influence of documentary assessments on practitioner discourse and forms of inherent assessments of the situation, and how their contextual influences impacted this process. To do so, I searched within instructional leadership team and PLC meetings to find instances where problems of practice and documentary assessments—such as test score data displays—triggered extended discourse and sense-making surrounding evaluation and planning for instructional improvement efforts. For the purposes of this paper, I will refer to these instances as episodes of *assessment talk*.

As mentioned previously, I examined the extent to which practitioners moved through different stages of data-driven decision making in their collective discourse: starting with *data*, providing forms of evidence to create *information*, and then using this information to create actionable *knowledge* for instructional improvement (Ikemoto & Marsh, 2007). Finally, since a key assumption of data systems for accountability and

instructional improvement is that educators are able to use this information to make informed decisions to improve student outcomes, I also examined the forms of evidence—or inherent assessments—practitioners provided during these episodes for identified relationships and interactions among teachers, students, and content, in the various environments of schools (see Cohen, Raudenbush, & Ball, 2003; see also Ball & Forzani, 2007). Within these episodes, four key themes emerged surrounding the depth with which participants engaged with data and forms of evidence, and the forms of information practitioners drew from during their collective assessments of the problem of practice. These themes highlighted: (a) student growth as evaluation; (b) the role of practitioners on student test scores; (c) aligning expectations to meet standards; and (d) connections between ELL language proficiency and assessment performance. These themes characterize patterns of data and evidence use and discursive topics within these episodes of assessment talk.

Within this chapter, I argue that an overreliance on whole-school and student data displays surrounding evaluation and progress monitoring within this context created a tendency for practitioners to draw from their own understandings of student and contextual issues to interpret test score performance. In contrast, grounding the discourse on student achievement or other aspects of summative assessments (such as assessment standards or alignment of curriculum) created more opportunities for practitioners to investigate relationships between student learning, curriculum, and teacher pedagogy. In essence, a focus on student data—particularly for evaluation or targeting students for intervention—often involved practitioners drawing from their

knowledge base on the antecedents of student achievement. While at the same time focusing on aligning and preparing practitioners to understand the linkages between the classroom and summative assessments created more opportunities for collective discourse surrounding negotiating actionable responses, implications and understandings of how to improve student outcomes on summative tests. Lastly, within a majority of the episodes of assessment talk, there was a lack of instances of inherent assessments surrounding understanding or professional training related to their Content-based ESL programming and the needs of emergent bilingual students. This was a factor in how they planned and understood instructional improvement within this school. While these episodes were situated within a particular school at a particular time, examining data-focused episodes of assessment talk in this way provides a lens to explore the active and collective processes of practitioner learning with data in the various environments of schools (Ball & Forzani, 2007).

Patterns of Evidence-Use and DDDM in Episodes of Assessment Talk

As mentioned in chapter 4, I identified instances of assessment talk surrounding instructional improvement to track the forms of documentary, inherent, and discursive assessments surrounding persistent problems of practice (see Table 4). Within these episodes, I categorized data displays and handouts within instructional leadership teams and PLC meetings as documentary assessments, which served as the “problem space” for collective discourse on aspects of instructional improvement. Instances where practitioners either drew from their own interpretation of the documentary assessments

or from their personal understanding of the situation were labeled inherent assessments, while topics of discussion within the episodes were identified as discursive assessments.

Table 4. Examples of Assessment Practice in Data Use Routines

Assessment Type	Definition	Examples
Documentary Assessment	Stable representation of assessment and theory of learning, often surrounding student assessments of learning	<ul style="list-style-type: none"> - Displays of student standardized test scores (student-level, grade-level, school-level) <ul style="list-style-type: none"> o Student performance on particular indicators - Scoring rubrics - Interim assessment benchmark goals and indicators of risk - Documentation of student interventions and supports - “Hot Lists” of students targeted for classroom assessment intervention - Frameworks for student learning
Inherent Assessment	An internal activity of individuals made public within collective discourse; forms of evidence or information understanding the situation	<ul style="list-style-type: none"> - Descriptions of student and family/community character, motivation, and circumstances - Shorthand labels of students (i.e., “bubble kids,” “Level 1’s”) - Aspects of professional expertise and knowledge surrounding pedagogy, curriculum, and student learning - Understandings of testing and accountability context - Understandings of ELL language proficiency and Content-based ESL curriculum - Understandings of school and district contexts (i.e., goals, expectations, available resources)
Discursive Assessment	Shared and reflective talk generated by a group of individual engaged in joint activity around a particular topic or problem of practice	<ul style="list-style-type: none"> - Factors or sources of evidence to use for decision making - Explaining lack of correlation between summative assessments - Exploring factors surrounding lack of student growth/achievement - Using summative assessments to target and support students to make standard - Expectations for summative assessment benchmarks and standards - Progress monitoring with assessments - Connections between ELL student language proficiency and assessments - Evaluation of student support programs/interventions - Exploring frameworks for learning

Documentary assessments used within episodes of assessment talk predominantly focused on examining student, grade-level, or school-level output of student achievement. However, during PLCs practitioners also focused on other aspects of student assessment, such as understanding testing preparation, student

accommodations, standards and goals of particular indicators, and aligning frameworks for student learning with high-stakes and interim tests. As practitioners engaged in “problem spaces,” either initiated or grounded by these documentary assessments, they drew from a variety of sources of personal knowledge to interpret and help form shared understanding within their collective discourse around assessment and student learning. Their efforts here ranged from utilizing their own shorthand labels of students to talk about how students were doing, to employing their expertise surrounding student learning, pedagogy, and curriculum tools. Finally, while much of the joint activity surrounding assessment practice centered on progress monitoring of student outcomes and targeting students for intervention, there were also opportunities—particularly within PLCs—for practitioners to discuss other aspects of professional learning, such as aligning expectations for literacy within and across grade levels.

I created matrices to organize these forms of assessment practice and find patterns in the forms of evidence, interpretations, and negotiated understandings of the problem within data use practices (See Table 5). As seen in Table 5, the majority of the collective discourse surrounding problems of practice and data focused on individual interpretations of student outcome data for decision-making. Student outcome data was often presented to instructional leadership team members or practitioners within PLC meetings, often leading to an extended conversation surrounding why they felt a student or grade-level score was one way or the other. Within instructional leadership team meetings, there was a higher likelihood of simply examining and noticing changes in student outcomes and moving to decision, such as categorizing or targeting students for

intervention. This practice was found within instances where an immediate decision was needed, such as evaluating an afterschool program for their School Improvement Plan or targeting students to participate in a particular intervention. Finally, within ILT and PLC meetings there seemed to be little opportunity to deeply engage with data and assessment. As such, there were only a few instances—primarily within PLCs— where practitioners spent time collectively developing shared understandings and claims about the situation, and negotiated possible solutions and strategies to the problem of practice. Finally, it is important to note that these opportunities to develop shared knowledge for decision making also happened to center around aspects of assessment other than looking at student outcome data, such as assessment rubrics.

Table 5. Connections to the Instructional Core within Data-Driven Decision Making Processes

Connections to the Instructional Core	Progression through DDDM Processes		
	Data	Data + Information	Data + Information + Knowledge
Student and Context	1, 3, 6, 7c	2a, 2b, 4a, 4b, 5 , 7a, 9a, 9b, 9c	7b, 8b*
Teacher and Context		4c ,	7b
Content and Context		2a, 2b, 4a, 4c, 5 , 7a	8b*
Student, Teacher, and Content		7a, 9a	7b, 8a* , 8c* , 10*

Note: Episodes of assessment talk are numbered. Meetings where there was more than one episode has been designated by letter. Bolded numbers are PLC meetings; non-bolded are instructional leadership team meetings. Instances where the documentary assessment in focus was not student outcome data have asterisks.

Additionally, I was able to draw linkages to practitioners' inherent assessments of the situation and interpretations of data with aspects of the instructional core. Overall, when faced with interpreting student outcome data, practitioners often focused on their understandings of issues outside of the instructional core of student learning, pedagogy, and curriculum context. Instead they often made connections between individual aspects of the instructional core and their wider context; for example, identifying parents as non-supportive, students lacking motivation, or issues with lack of alignment or applicability of assessment to their population. Identified interactions between pedagogy and their context of accountability was not commented on as much, perhaps due to the teaching staff's reluctance to "teach to the test" within their classrooms, as evidenced by teacher pushback on the ILT's plan for the MSP Blitz. Finally, on a few occasions practitioners made commentary specific to interactions and connections between students, teacher, and their curricular content, primarily within PLC meetings.

The next section focuses on these episodes in depth and explores how assessment practices within data-driven decision making helped to shape the discourse on program evaluation, student achievement, and testing preparation. In particular, I explore thematic characterizations within these episodes of assessment talk and consider their implications in context. I provide examples of these data use trends and contextualize the episode by providing background to the problem of practice and trigger of the particular episode of data inquiry. In addition, I pay particular attention to the forms of inherent assessments practitioners used during these episodes of assessment talk. Then I synthesize these examples of assessment talk and consider the implications of data-

focused inquiry given identified tensions or barriers to decision making as outlined in the previous chapter.

Focusing on Outputs: Student Data and Practitioner Inquiry

Tracking student growth and performance on summative assessments is a central task of instructional leaders within “priority” schools. As schools that struggle to make AYP (adequate yearly progress) or AMO (annual measurable objectives) from district or state expectations, they must show how they use district and state provided summative assessments to make decisions and demonstrate student progress over time. Therefore, as mentioned previously, the vast majority of data-focused discourse surrounding problems of practice centered on student outcome data. As such, discourse centered on examining school- and student-level data as feedback loops on decision outcomes, targeting students for interventions, attempting to interpret the variety of student summative assessment data to track progress throughout the year.

Student Growth as Success

As highlighted in the previous chapter, instructional leaders used their multiple measures to create a spreadsheet that categorized students as Level 1 through 4 based on their performance on the MSP, MAP, DIBELS, and WELPA primarily. Decisions were then made at the beginning of the year for Walk to Read classroom placement, targeted intervention, and after school participation. At the end of the year and as part of their

school improvement planning for the following school year, the instructional leaders would examine the Spring scores of MAP and DIBELS student data who had participated in their after school programming.

With after school programming evaluation as a backdrop, displays of student data served as the problem space for whether or not an after school program was effective. In their evaluation, grade- and student-level data were addressed systematically, going through each grade level or student and seeing if they showed growth over the year or not. Exploration as to why there was growth or not was often minimal, with the culmination of this data inquiry process concluding if a particular program was successful and should be continued for next year as part of their school improvement plan.

In this episode, the instructional leadership team is examining a projected display of individual student DIBELS data to get a look at the particular students in their church- and community-based afterschool programming.

Episode 7c: Examining student grades for after school improvement

Patty: She was in the fall at 55, and in the spring at 67. She went up to 85, so there's some good growth.

Edward: 55, 67, 85.

Patty: And the next 3rd grader I have is [Richard¹⁶]. He was mostly in math, he was already at benchmark in DIBELS. [Colt] went from in the fall 20 words per minute in the fall---

Edward: He went from 20--

Patty: 20 in the fall to 49 in the winter, and then—

Edward: 46 in the fall—or no, spring.

¹⁶ All student names are pseudonyms

Patty: [Kathleen] was 50 in the fall, 60 in the winter—
Edward: 92 in the spring.
Rebecca: Oh that's good growth.

In this case and in similar episodes, looking at student after-school participant growth in relation to whole-grade level growth was a primary indicator of programmatic success. For example, Patty runs down each student on the data display and highlights that “Kathleen’s” DIBELS scores had improved from 50 to 92, fall to spring, and declared that “good growth.” While there were a few utterances of why a particular student succeeded or not (such as perceptions of behavior or motivational factors), overall the discourse surrounding whether or not a program was successful centered squarely on student growth performance, despite any of their previous reservations on the limitations of the MAP and DIBELS for tracking student performance. It’s important to point out that this form of assessment talk was within the context of evaluating programs so that they could show progress and justify continuing with these programs within their school improvement plan. Therefore, just showing yearly student growth on MAP and DIBELS assessments appeared to be the main goal of this particular inquiry for accountability purposes.

This pattern of data use highlights a particular structure of decision making with data that perhaps has external influences. Within the structure of data-driven decision making processes, these leaders moved directly from examining displays of student data to a particular decision outcome—such as the continuation of a program—bypassing interpretation or collective meaning-making surrounding why the program worked for

their students. In this case, it is possible that meeting the state or district goals for student improvement in participation with these programs did not require any further analysis into what was specifically helpful for students or what other factors may have led to these outcomes. Within this context, it was often the absence of student growth that led to further engagement with assessment data for student improvement.

Targeting Students for Intervention

In cases where student outcome data showed particular issues for meeting standards or benchmarks in summative assessments, displays of student data served as the predominant, but not sole, source of information for instructional leaders and practitioners to make decisions about how to target groups or particular students for intervention. In certain cases during the year (targeting students for the MSP Blitz, focusing on particular indicators of assessments in the classroom), leaders and practitioners drew from a variety of inherent and documentary assessments to make decisions about who to target for intervention. Edward outlines some of his reasoning as his team considers what factors to use in determining who gets to participate in the MSP Blitz:

Do we want to—we could just let the teacher choose—let's say [Joe] wants to do it in 5th grade. Let him choose 10 or 12 kids in his class who he thinks, with an extra hour after school, he could push them over the top. Or do we want to say, hey, here's a group that we think, based on their conglomerations of scores here,

we think we can push them over in Math with a little bit of extra help....Or do we want to stay with the same ASPIRE group all in one classroom....

In this case, they consider two primary factors: inherent assessments of teachers surrounding who would benefit most from this intervention, and documentary assessments of student data displays. They take into consideration the “bubble kids” who are already participating in the after school ASPIRE program that were initially targeted via their student spreadsheet of multiple measures. Patty likes the idea of teacher nominations, as long as they are using a targeted “hot list” of students derived from MAP and DIBELS scores alongside their formative assessment of student needs.

This blending of formative and summative assessment data for decision making was seen as a desired practice by the instructional leadership team. However, in supporting practitioners to show growth on interim assessments, there was a tension between training teachers to utilize the theory of learning found within benchmark assessments—such as DIBELS indicators— in contrast to their understanding of literacy training for emerging readers. In this episode, Patty is working with her 1st grade teachers within their PLC-time to help them identify their “hot list” of students from student-level DIBELS assessment output in order to have them focus on particular indicators within the DIBELS. Here, individual student DIBELS scores served as the “problem space” for discussing how to boost student outcomes.

Episode 4a: “Hot Lists” and “Going Through the Hoops”

[The 1st grade team is looking at a handout of student-level DIBELS output with individual scores for specific ¹⁷indicators]

Patty: ...If you look right now and you look at that column over to the left, the first score column, that's where they were in winter. And in winter they had to have 35, and now in Spring it's still 35 on that one. So I have highlighters here too, so feel free to grab a highlighter, marker, whatever you want. And that's the Phoneme Segmentation. We used to called that "break-it-down" in SFA (Success for All)...(1)

And now the next target that they're going to be assessed in the target is the Nonsense Word Fluency. And for the Nonsense Word Fluency you'll see two score columns. The first one is CLS, and that is Correct Letter Sounds. And then the one to the right of that is Words Read Correctly.

Here's the dirty little secret here. The Words Read Correctly doesn't count at all. It's the Correct Letter Sounds that matter. (2)

Teacher 1: So just really sounding it out is key?

Patty: Correct Letter Sounds. So the key is how many correct letter sounds can they get in one minute. That's the key. So if they're going [*verbalizing*] nuh-ah-guh, nug, buh-uh-m, bum, you're slowing them way down. So they need to be doing just sound it out and move one. [*Verbalizing quicker*] B-A-U, N-A-G. (3)

Teacher 1: Don't really make it a word?

Teacher 2: Don't make it a word.

¹⁷ Phoneme Segmentation Fluency, Nonsense Word Fluency, Nonsense Word Fluency - Correct Letter Sounds, Nonsense Word Fluency - Words Read Correctly, Oral Reading Fluency, Oral Reading Fluency - Words Correct, Oral Reading Fluency - Accuracy, Oral Reading Fluency - Retell Fluency, and Word Use Fluency.

Patty: And to me, as a reading teacher, that is counter-productive. But that's how we're going to play the game here. That's going through the hoops. (4)

The intention of these student “hot lists” derived from DIBELS performance data was for teachers to identify which students were at the cusp of passing a particular measure and work on specific testing strategies during class. So Patty focused the conversation on providing the teachers with a form to fill out of their targeted students derived from a look at their students’ scores on particular indicators (1). Then she pulls from her own sources of knowledge surrounding the DIBELS assessment to explain to teachers what indicators “matter” for meeting benchmark goals and how to prepare their students to meet those particular indicators (2 & 3). She comments that these practices seem “counter-productive” to her understanding of reading pedagogy, but that they need to do it in order to “go through the hoops” of passing the DIBELS assessment (4). While there was some discourse surrounding particular student DIBELS outliers (such as an ELL student being a “good decoder” or having someone unfamiliar to the student administering the assessment) during this meeting, student outcome data seemed to take priority for targeting particular “bubble kids” for support with particular DIBELS indicators. Patty called this “play[ing] the game,” a strategy that schools often find themselves employing in the classroom when under pressure to demonstrate performance and student growth by external accountability sources (Diamond & Spillane, 2004, Pandya, 2011, and others).

This episode suggests that when identifying students for targeted interventions, documentary assessments seemed to take precedence over teacher understanding of “best practice” for teaching students literacy. Yet, devoid of a particular immediate decision outcome (such as targeting students, or continuing with a program), examining student data more informally for progress monitoring allowed for more extended collective discourse on the potential sources of student outcomes.

Interpreting Student Outcomes

The arrival of new MAP and DIBELS scores in the Winter and Spring provided opportunities for the instructional leadership team to take a “first look” at the scores in order to interpret them, often separately, and then collectively during their weekly meetings. Within these more informal assessment routines, the instructional leaders drew from their own personal knowledge and expertise to consider the possible antecedents to student outcome data on interim assessments as part of their progress monitoring routines. Within these informal assessment routines, instructional leaders often identified barriers to effective data use practices, such as lack of alignment between the MSP, MAP, and DIBELS assessments; and issues of assessment “fit” with ELL student literacy comprehension and fluency indicators found within DIBELS. For example, upon examining the winter MAP and DIBELS scores with the previous MSP scores, Patty brought up to the team that one big reason why there was a “disturbing” trend in the upper grade levels was due to a lack of correlation between the two interim

assessments. While wanting to be on the same page as the district, she worried about where they should be focusing their energies given these discrepancies.

It was also within this meeting that Rebecca was able to draw from her own expertise as the ELL Specialist to highlight the tension between ELL student language comprehension and fluency indicators found in the DIBELS assessment. In particular, she also drew from her own experiences working with students during her small group intervention time to highlight how testing strategies with the DIBELS had manifested in everyday learning strategies for her lowest “Level 1” students. These two insights surrounding their available assessment tools—lack of alignment between assessments and “fit” for ELL students—were ideas that moved beyond the instructional leadership team meeting context and into other leadership arenas. For example, when going over student test scores during the Site Council meeting, Edward pointed out that their interim assessments may not be entirely correlated with each other, which prompted one of the grade level leaders to inquire why they continue to use them if they are not correlated with the “high-stakes” MSP test. Finally, when given the opportunity to reflect on the assessments themselves, Patty was also able to question if the assessments made sense for their particular student population. For example, Patty initially felt that there may still be some connection to fluency and comprehension due to her knowledge of the “research.” Yet by the end of the year she seemed convinced that there was a problematic connection for their ELL students, perhaps in part due to her interactions with Rebecca surrounding this issue.

Connections Outside the Instructional Core

However, it is important to note that when the “problem space” derived from student test score outcomes, the discourse tended to focus practitioner inquiry outside of the instructional core. Specifically, in episodes where there were opportunities to connect outcome data with practitioner interpretations or understandings of the situation, they related individual aspects of the instructional core (such as ELL student learning) to external contextual factors (DIBELS fluency measures), instead of making connections *between* student learning, pedagogy, and curriculum content. Perhaps due to the perceived lack of alignment between measures and their difficulty in deriving diagnostic implications from the interim assessments, the instructional leadership team reported sometimes feeling at a loss for how to work harder to improve student outcomes in their particular context.

In this episode, the instructional leadership team is taking a first look at Spring MAP and DIBELS data. After a little bit of discourse surrounding MAP score changes and possible influences of Enrichment and Intervention time, or Math teaching in general, the topic of ELL language challenges took focus into their interpretation and examination of DIBELS scores.

Episode 7b: Perceptions of Efficacy for Instructional Improvement

Patty: DIBELS. The DIBELS looks bad. There again, I just don't think we can-- it can't be an excuse, but I don't think we can discount the fact that we're “DIBBEL-ing” second language learners for the most part or for a large part, and that's not going to come as naturally to them. (1)

Rebecca: It's not. I'd like to see are there schools with our population that have really done a good job of getting those scores up, and has that affected comprehension, because those are the two pieces for me. That's always the... we need to have children reading fluently so that their comprehension is-- because it affects comprehension. And I'd love to see data on a school who did a really great job with fluency, and also having increased comprehension and see what they're doing. Because I'm not as concerned just about increased fluency. It is not impacting comprehension. (2)

Patty: I agree with you. And I find that a lot of our kids that are making the grade are not necessarily having any sort of increased comprehension because they're just so focused on that decoding piece. But I feel like I don't know what more to do. We've had [district consultant] come and do in-service. We've talked about the different kinds of things, we work with our kids along those lines. I've really pushed it all year long the last two years as a focus, and I think they almost look worse this year than they did last year, don't they? (3)

Edward: I think they're about the same ballpark.

Upon looking at their student outcome data, Patty brings in her understanding the being a second language learner makes it difficult to meet benchmark goals for the DIBELS (1). Rebecca clarifies this statement by using her expertise as the ELL Support Specialist to again comment on her identified issue surrounding English language learner comprehension alongside DIBELS fluency measure. She also brings up her frustration that they have no examples of other schools with similar student demographics actually performing well on this particular interim assessment (2). Agreeing with Rebecca, Patty shares her observations of the lack of student

comprehension alongside increased fluency scores and voices frustration that they are not able to improve both of these aspects of literacy despite their best efforts (3).

While the instructional leadership team was able to employ both documentary and inherent assessments (e.g., student data and personal knowledge of the interaction between ELL students and assessments) during their collective discourse surrounding poor DIBELS scores, they developed understandings about their situation that focused on factors they felt were out of their control, such as the district interim assessments they must show growth on. Additionally, Rebecca brings up an issue surrounding a lack of available models for “best practice” for their student population, which can present barriers for instructional improvement if they feel the resources were not made for “schools like us.” As a result, instructional leaders and staff alike were often frustrated and disappointed in their interim assessment student outcomes and struggled to connect what they were doing with students in the classroom to show growth on these assessments.

In this context, the “schools like us” problem likely derived from not being able to use assessments as diagnostic tools aimed at improving the instructional core, particularly with their ELL students. This outlook led to feelings that outside consultants or “best practices” within schools with different demographics were not particularly helpful to them, nor could they find high-performing examples of schools to give them ideas about how they might change their practices. Edward explains to the team his feelings on getting additional consultant help within the school:

And I just said I don't have time to tell you what—I don't feel like you're going to, to bring you in here and to consult with me is going to move us forward. I said I feel like we've, over the last couple of years, have kind of put the ideas that we have collectively either into place or have ruled them out. I said I kind of feel like we're at a point where if there are schools with our demographics that are doing way better than us, then I'd kind of like to know who they are and what they are doing so that we can further our ideas.

In this quote, Edward highlights the team's reluctance to use district-led supports that they feel have little relevance to improving outcomes for their student population, which may limit the additional forms of information and models for instructional improvement.

Therefore, within this school context, the overreliance on examining student assessment outcome data to respond to state and district expectations of growth created a “problem space” focused primarily on connections to external factors within this particular school context. This finding leads me to consider the extent to which “problem spaces” created by focusing on understanding testing standards, or aligning practices horizontally and vertically along scaffolding and meeting MSP grade-level expectations would lead to different forms of inquiry that were focused more internally on aspects of the instructional core.

Focusing on Inputs: Aligning Expectations to Teaching and Learning

Focusing on student outcome data as “problem spaces” for practitioner discourse surrounding instructional improvement seemed to create patterns of data use inquiry that focused on generating collective understandings of the situation, but seemed to stall on negotiating actionable responses and implications of knowledge gleaned from data focused inquiry. Therefore I extended my definition of “episodes of data use” to also include other aspects of assessment practice, such as understanding standards and aligning practice to standards, in order to consider the extent to which different types of documentary assessments—other than student data— influenced the nature of collective discourse surrounding instructional improvement. However, these forms of data were not initially a central focus of data collection, since these instances of assessment talk are derived from qualitatively different data: fieldnotes from Rebecca and Edward-led PLC meetings. These two cases provided discursive examples of assessment implications on teaching and learning.

Interrogating School Expectations for ELL Student Speech

In her PLC work with the 6th grade team, Rebecca focused the meeting on exploring and unpacking the WELPA rubrics for assessing forms of student speech. Due to the upcoming testing and to support teachers in preparing students for this test, she handed out two rubrics focused on WELPA expectations for “clear descriptions, explanations, and requests for information,” as well as the student’s ability to tell “a

coherent or cohesive narrative [that] is based on a logical, ordered narrative” surrounding a series of pictures. She explains that many of their ELL students are primarily 2’s and 3’s on the WELPA and that she wanted to look at the differences between a 2 and 3 score to discuss “how we are pushing our kids in terms of our expectations of them on a daily basis with their oral language.”

With the WELPA assessment rubric as a “problem space” for discourse surrounding oral language expectations in day-to-day student interactions, the staff participated in an extended discussion that helped to build their understandings and negotiate actionable responses surrounding expectations for student speech in their classrooms. For example, they considered tensions in correcting students’ English within the classroom, the overall vernacular of the school surrounding “natural speech” and its implications on ELL language learning, and how issues of vocabulary and language acquisition may be interacting with students’ connection to expository stories within the classroom. In essence, they were able to consider aspects of the situation that were directly tied to the instructional core. In this episode, Rebecca is outlining how they could be maintaining their oral language expectations within the classroom.

Episode 8a: “Natural Speech” and Expectations for Oral Language

Rebecca says, “I think as whatever content area we are working on, we can push the major information that’s more of a content-related piece, but [also] the language piece of it.” She goes on to explain their ELL students have good sentence structure but with several grammatical errors. And that they should be “honing in” on that as the teachers are doing grammar and reading, especially during their mini-lessons in writing. She says, “a lot of times I think it kind of slides by and we do not take the time to correct it, but if we don’t, they are not going to.” (1)

A teacher comments on how when she tries to get them to correct their grammar in the classroom they seem to get embarrassed [*holds up hands and waves them in front of her face*] and don't want to talk about it anymore. (2)

They talk about not “harping” on academic English so much that they don't talk at all. Rebecca provides some example to the teacher surrounding how to use language targets and expectations in the classroom, “This is how we are going to start your sentence...” and provide other reminders to the students surrounding “when you saying this make sure you...” (3)

Simone chimes in saying, “I feel like when I am in my Reading classroom, or even my classroom in general, the vernacular of our school even if it's a native speaker is [*lowers voice for emphasis*] low.” She focuses in on the example of using “like” repeatedly throughout speech and comments that she notices when a particular native speaking student in Reading class uses this, a non-native speaker in the classroom picks up on it and starts using it. She emphasizes that this should not be happening in the classroom. (4)

After handing out the WELPA rubrics, Rebecca begins to unpack the differences between what “numerous,” “some” and “few” errors in grammar look like in context and how might teachers be modeling and monitoring that in the classroom. She draws from her observations of classroom practice that students are allowed to make obvious grammatical errors in the classroom without being corrected (1). When a teacher talks about her interactions with students in her classroom surrounding correcting their speech, Rebecca is able to provide concrete examples surrounding how to help students with these expectations (2 & 3). Finally, 6th grade teacher Simone provides her assessment that the expectations for student speech in their school, as a whole, was “low” and that is a problematic example to set for their second language learners (4).

This episode provides an example of how WELPA standards (as displayed as a rubric) served as a jumping off point for some informal professional development on oral language expectations for ELL (and Native speaking) students within the classroom. Rebecca's focus on language targets and oral language expectations created an opportunity for teachers to discuss their understandings of the situation (challenges to correcting students and overall school "natural speech" usage) as it related to teacher pedagogical choices and interactions with students in the classroom. The rubric served as a touchstone for discussion as teacher discourse moved from focusing in on interactions between teacher and students towards teacher engagement and understandings of specific aspects of the WELPA rubric in their school context.

Horizontal and Vertical Alignment

Finally, Edward's PLC with the 4th grade team differed from all of the other episodes in that the episode was triggered by exploration of a new framework for aligned K-4 literacy (David Matteson) being used within the district. This open space for practitioner input allowed one teacher to bring in her own MSP score sheet as a way to center their conversations surrounding aligning K-6 literacy expectations with an eye on MSP grade level expectations. In this episode, Edward starts the meeting by explaining how part of their needs assessment involves creating horizontal and vertical alignment of their literacy instruction. He mentions in the past how they had been "on their own" for implementing something like that, and explains how the David Matteson developmental framework for literacy trainings will be coming from the district shortly. He distributes

several framework handouts¹⁸, and asks the teachers an open-ended question surrounding how they might go about aligning grade levels surrounding literacy and writing.

Episode 10: Aligning writing standards and the MSP

Teacher 1 says, “one of the biggest issues is the fact that many students come with very little writing experience” and she points out [referring to the handout] that “right here” we are seeing a large amount of writing experience and opportunities for writing throughout the day. She emphasized that how the 4th grade team has a better chance at getting their students to write a 3-5 page paper with some continuity in the prior grades. (1)

Edward discusses where 1st and 2nd grades are in terms of providing beginning literacy and writing experiences.

Teacher 1 continues the conversation by also noting a strong spelling component, and says, “that’s the second big battle we fight, you cannot pass state tests if you don’t have conventions.” She goes into how she explains the MSP scoring for writing to her students and highlights how students can get a particular score for ideas and organization, but could potentially not pass due to the use of incorrect conventions such as using lowercase “i” or spelling errors on basic words like “because.” Mentions that going back to a strong spelling component would be “powerful” for her students. (2)

Edward asks the group where conventions are located on the Common Core.

Teacher 2 comments how they are similar to what the current ones are on the MSP.

Teacher 1 then brings out her MSP “cheat sheet” on scoring sheet and continues to break down scoring for the group surrounding writing. There is a discussion between Edward, Teacher 1, and Teacher 2 surrounding how it is the

¹⁸ The David Matteson handouts were called "The Early Literacy Continuum for Writing" on one side and the Developmental Perspective of Literacy Development on the other. They also his Reading Response Journal available online.

conventions piece that Teacher 1 explains “breaks a lot of them. Any of our 8’s coming through, that’s our 8’s.” (3)

Edward asks the group if they “see this happening in a school-wide way, instead of an isolated grade [level] way.”

Teacher 2 explains her perspective as a former 3rd grade teacher and explains that K-3 is all focused on the ultimate goal for making MSP 4th grade level expectations, and 5th and 6th grade goals are moving past 4th grade expectations towards the next state test. (4)

By opening up the floor for teacher input, the veteran teacher of the group was able to bring in her own deep understanding of students’ current levels of writing and literacy and consider the extent to which they would need to be scaffolded in prior grade levels in order for 4th grade to meet their MSP expectations in writing (1). She hones in on one particular aspect of the MSP 4th grade writing standards (spelling) to talk about her observations of students not meeting conventions for writing and how that impacts their MSP score (2). She then provides her “cheat sheet” of MSP standards as a way to ground their conversation surrounding how the MSP scores get calculated (3). Finally, a different teacher connects the conversation about aligning writing and literacy standards across grade levels with her perception that all grade-level goals should be focused on the “high-stakes” MSP tests during 4th and 7th grades (5).

In this episode, teachers were able to deeply engage with developing shared understandings surrounding the implications of aligning literacy and writing throughout grade levels, while keeping an eye on being able to successfully pass the MSP during this process. By starting with an open-ended discussion on how to align curriculum

(grounded by a framework handout), the teachers were able to draw on their understandings of current student writing development within the school, particular aspects of writing that challenge students in passing the MSP writing portion, and how it's important to delineate by each grade level the sorts of writing scaffolds 4th grade teachers need prior to students entering their classrooms. In this example, teachers are engaging with the problem of practice by considering student and teacher connections to their current curriculum, and to external curriculum connections with the MSP standards.

As a result of focusing on aspects of assessment beyond student outcome data, practitioners were able to focus on practitioner and curricular inputs that have direct connection to the instructional core versus aspects of the situation that they feel they have little control over (such as district and state choice of assessment tools). However, it is important to note that given their student population and identified needs, it was surprising to see very little discussion (other than that prompted by Rebecca) surrounding these issues. Therefore, the final finding of data use practices within this school context concerns examining how discussion surrounding ELL language supports and issues occurred and how educators were able to engage with those ideas collectively.

ELL Practitioner Knowledge and Analysis

As highlighted in the previous examples of data use discourse, the majority of discussions surrounding ELL student language needs and potential issues with

curriculum, testing, and teacher expectations were often driven by specific comments by Rebecca to bring up these issues for discussion. This may be due, in part, to an emphasis toward focusing on student data and testing over coaching ELL support strategies during PLC time. For example, she commented that she feels guilty that the staff comment on how there is no training for ELL classroom support strategies, and that when she does do it, it is mostly “handing out information” instead of taking it to the next level. She explains how this hasn’t been a priority during PLCs and other opportunities for professional learning:

I used to do after school groups where we’d do a book study or go through something, that was completely on my own, and on their own time. It wasn’t necessarily like a set up PLC structure. But this year with the language objectives, which I see the importance of, I do. And I think if we could get that going in the classrooms it would be wonderful. But it’s not really a school-wide initiative. It’s been a “Rebecca, please support that in your two PLCs that you get with the grade level all year.” And there’s not been the expectation that it’s happening in classrooms. Nobody’s going around to check.

In addition to this limited time coaching teachers on ELL support strategies, there were tensions surrounding providing support for language needs versus academic needs, particularly for new students with very little English language exposure. In discussion with the instructional leadership team, Rebecca explains that she is “more of an academic support specialist now than I am a language support specialist, at least in the E&I times, and what I am getting from the teachers,” and that they need to hire an

additional instructional support specialist so that she can focus on more intensive language instruction for ELL students. She acknowledges that this current structure was “out of necessity...of truly meeting student needs,” but wishes there were dedicated times to work with WELPA Level 1 students. Patty confirms this within her interview in that she feels that they don’t have a set curriculum for ELL students and that they mostly focus on oral language and some GLAD strategies in the classroom. That said, while the district is looking to develop a more specific language curriculum (for example, one school in the district has a dual language curriculum), there are currently not a lot of opportunities for Rebecca to be freed from teaching reading groups to focus on language-specific groups.

These factors created a school context that lacked ongoing professional development for ELL student support strategies and provided less time to work on language strategies with newcomer students. It is possible that a combination of these factors perhaps led to the lack of practitioner inherent assessments surrounding ELL language issues during opportunities for collective discourse. In this final episode, we see Rebecca working with the 3rd grade team during PLC-time to collect feedback on their understanding of the Content-based ESL model that they use by asking how things are going in their classrooms and if they need any additional support.

Episode 6: Practitioner Knowledge and ELL Student Expectations

Teacher 1: Maybe I just don’t know enough about what to look for in helping ELL students, but I don’t see in my classroom, I think any concerns that could be related to ELL concerns. Does that make sense? (1)

- Teacher 2: I think so. I feel the same way.
- Teacher 1: I see behavioral concerns, I see other things that I could use help with, but I have kids that are struggling in Math but I don't think it's because of ELL. (2)
- Teacher 2: And I feel like this group in particular, they've all been here.
- Robin: Yeah, we have more settled out—
- Teacher 2: So they know how to assist each other. Like if you don't get them the directions, they'll real quick in Spanish, "blahblablahbla. Oh!" without us prompting them. They kind of know how to support themselves. My thing is more with the home piece, like with homework. (3)
- Teacher 3: Oh god. This group has no home support.
- Teacher 2: If the teachers aren't able to do it, I don't have a good way of communicating with parents.
- Teacher 1: Yeah.
- Teacher 3: Well it's not even the fact that they don't do it, it's just the fact—like we color-code our homework, all of us do. They're not checking to see if that yellow paper is done. You don't have to understand it. Is it done or not? I mean really, that's why we color-code it so you don't have to figure out what the paper says or anything like that. And we made all that very clear at the beginning of the year at curriculum night, at conferences.
- Rebecca: And how many people were at curriculum night?
- Teacher 3: Well yeah. (4)

Different than what we saw in episode 10, opening up the floor for feedback on their Content-based ESL programming led to practitioners initially focusing on their

perceptions of outside and external influences on their ELL student outcomes. To his credit, Teacher 1 starts out the conversation stating that he “just [doesn’t] know enough about what to look for in helping ELL students;” however, he ends that statement with his general perception that the ELL students within his classrooms do not have any challenges based on their language proficiencies (1). Instead he focuses on his inherent assessment that his ELL students suffer more from behavior issues, rather than language issues (2). Teacher 2 supports his statement to claim that their students “have been more settled out” or have been within the school district for many years, and that she feels that they can support each other in the classroom when they are not sure about what was said or what to do (3). Finally, Teacher 3 brings in her own lens to focus on the lack of parental support for this student population and her frustration that they do not check their homework despite having it “color-coded” so the parents “don’t have to figure out what the paper says” in order to check to see if the homework was completed (4).

This episode highlights several themes specific to this school context, but also within schools new to serving a critical mass of Latino and emergent bilingual students. First, Teacher 1 admits that while he may not know what to look for in helping ELL students accessing mainstreamed curriculum, he expressed that overall he doesn’t feel that issues within his classroom are “ELL concerns.” In a post-meeting reflection, Rebecca commented that in this meeting she was trying to listen and “not push my perspective or agenda” in order to see what teachers’ perceptions are surrounding ELL student learning needs. In this case, she found it interesting that he stated he didn’t have

any ELL language issues within his class yet he has 8 ELL students, and commented “even your Level Four’s still have language issues, you know.”

Secondly, the lack of practitioner insight surrounding ELL student language needs in their classrooms may have contributed to an extended discussion of their own inherent assessments of student issues, all of which focused on aspects outside of the instructional core, with a somewhat deficit orientation. For example, after beginning with a conversation surrounding needing larger class sizes in order to have support to have ELL students pass the MSP, they then discussed students’ behavioral issues, lack of parent support, and eventually ADD-issues with students. However, later when Rebecca re-focused the discourse to look at support structures for teachers, they began to highlight potential issues with the leveled-grouped Walk to Read students forming shared understandings of where students are at in terms of reading and language issues. Finally, the meeting ended with an ask from teachers to have more ongoing professional development on ELL language support.

This episode—and episodes of data use without an ELL focus—suggests that without access to regular professional development, training, and language curriculum frameworks to draw from to interpret student outcomes, practitioners in Latino and emergent bilingual schools like this may be missing vital elements in their sense making to improve instructional outcomes. In particular, without professional knowledge surrounding language acquisition and student engagement with curriculum materials—such as Rebecca bringing up that sometimes what seems like ADD are symptoms of ELL student comprehension issues in the classroom—practitioners may focus on deficit

assumptions about students and their families, instead of working to build their own capacity to educate second language learners.

Summary: Sources of Information for Collective Inquiry

The forms of evidence used within organizational routines to make sense of student outcomes and plan for instructional improvement were found to be influenced both by the context (such as for evaluation versus open inquiry) and the “problem space” created by the documentary assessments used to explore problems of practice (student output data versus standards found in assessments). While not exhaustive of all of the ways in which data was used to form decisions throughout the year, these episodes highlighted particular dynamics of interaction that may contribute to the often shallow depth of analysis with data and to the limited forms of connections practitioners are able to make with the instructional core.

The school’s accountability context seemed to play a large role by creating opportunities for data -focused inquiry (e.g., assessing program interventions for efficacy), but was also seen as a possible barrier to effective data use practices to improve student outcomes. Language acquisition issues for their ELL students were not perceived as a barrier for student success, except when highlighted by their ELL Support Specialist. Finally, while it may go beyond the scope of this study, these findings suggest implications for what instructional leaders and practitioners potentially learn in the pursuit of using data to bring insight into persistent problems of practice in this kind of school context.

Chapter 7:

Discussion, Conclusions, and Implications: Expanding the Forms of Evidence in Data Driven Decision Making

The systematic use of data to inform decision making and instructional reform remains complex, yet recent developments in policy and within the literature hold promise for improving student outcomes. This study took place at the tail end of Wave 2 of standards-based reform, where, under No Child Left Behind federal Title III, aid to English language learners added assessment requirements for student English language proficiency as a requirement, and data systems and accountability measures were created to set performance targets, monitor progress, and to provide incentives for meeting standards (Hakuta, 2015). While Wave 3 of the Common Core State Standards shows promise for new assessment systems and standards for the English language proficiency of ELLs to be aligned with language demands in content knowledge, this study has illustrated the challenges in using summative assessments to create the sea change necessary to shift interactions between student learning, pedagogy, and curriculum material—that is, to change the instructional core itself. This is due, in part, to the *retrospective* assessment of learning that interim, benchmark, and state “high-stakes” summative assessments provide educators. That’s in contrast with formative assessments that allow educators real-time assessments of student learning (Linquanti, 2014).

However, within schools that are under accountability pressures to show immediate growth on student standards and benchmarks, it is often a focus on

summative tests rather than formative assessment that take center stage within organizational routines for instructional improvement (Diamond & Spillane, 2004, Pandya, 2011, and others). Additionally, case study within Ohio schools in the “Latino Diaspora,” found an increased percentage of ELL students in these schools are taught in ways that display a *decrease* in research-based strategies known to address language-related issues effectively (Rader-Brown & Howley, 2014). These contexts were found to be exerting particular pressures on effective data use within this study.

In this chapter, I will briefly summarize the main findings from my study, and comment on what they contribute to the emerging lines of research concerned with data use practices aimed at instructional improvement. Following that, I note important unanswered questions and aspects of the research problem that my study was unable to explore satisfactorily. I then underscore possible studies that could investigate those matters more fully. Finally, I consider what these findings mean for the practical work of teaching a linguistically diverse population of students, and for the policy designs that, while purporting to support this instruction, yet which have often failed to do so.

Summary of Findings

Within this particular school during this particular time, I found that instructional leaders and practitioners used their own combination of “high-stakes” and interim summative assessments alongside assessments of language proficiency to efficiently and easily categorize students for targeted intervention. Examining student output data and preparing for testing was a common focus in many of the instructional leadership team

meetings, and within the few PLC meetings I observed. Within these meetings, difficulties surrounding assessment and data use focused on practitioner perceptions of a lack of alignment between “high-stakes” and interim summative assessments, little connection to or priority on formative assessment, and a lack of assessment “fit” for their ELL students. Despite these issues, the leadership team felt they had to focus on showing improvement on these assessments in order to avoid punitive measures from the state and district.

As a result of this context, observations of practitioner “assessment talk” found that focusing the inquiry on student outcome data created a tendency for them to bring their own understandings of *external* influences on student achievement into consideration, such as focusing on student motivation or parental support, or focusing on the limitations of their summative assessments. In contrast, focusing on other aspects of assessment not concerned with summative measurement, such as exploring rubrics or aligning developmental learning models to assessment expectations, provided more opportunities for practitioners to draw from their own understandings of what was or could be taking place in instruction itself—the interactions between student learning, pedagogy, and curriculum materials, and the ways these interactions might be influencing student learning. In addition, the purpose of data-based inquiry or deliberation mattered. Cases in which an immediate decision was needed, as with an evaluation or targeted intervention, are juxtaposed with situations in which the goal was more informal progress monitoring. This dynamic impacted the degree to which practitioners moved from examining “data” for a quick decision, to more open

interpretation of what the data might be saying, and ultimately to developing shared understandings and actionable responses to the problem of practice. Finally, despite the large number of English language learners within the school, there were very few conversations surrounding language proficiency and ELL student interactions with aspects of the instructional core, except when prompted by the ELL Support Specialist.

Regarding my research questions, I found that practitioners within this rural, Latino and emergent bilingual serving school did not feel their current forms of summative assessments “fit” the needs of their students, particularly their ELL students, and that these perceptions of “fit” often created contradictory or confusing evidence for developing understandings of how to improve student outcomes. In their assessment routines, practitioners regularly used their own understandings of the variables that impact student achievement outcomes, primarily focusing on aspects outside of teacher control.

Furthermore, it was also clear that these practitioners’ focus on student data output was primarily driven by their designation as a “priority” school within the district, and as such their need to show improvement with particular student sub-groups such as ELL students. However, the district’s lack of coherent ELL language policies, curriculum, and professional development for staff created few sources of research-based insight for interpretation of ELL student assessment outcomes. As a result, school planning for instructional improvement was seen as “Ready, Fire, Aim.” The lack of diagnostic information derived from summative assessments, competing goals from misaligned assessments, coupled with the lack of training or examples of “best practice” for

serving ELL students, created a lack of clarity surrounding the purposes and goals for instructional improvement. In short, their engagement with data yielded far less insight into instructional improvement than the rhetoric of data-based practice would have assumed.

What the Study Contributes to Research on Data Use Practice for Instructional Improvement

This case study highlights how data and assessment use is an embedded and situated process where the particular contexts of data use practice and processes impacts opportunities for practitioner learning. In particular, this research provides particular insights into the intersections of competing “logic models” of accountability, and using data to inform practice (Orland, 2015).

First, assumptions that the availability of assessments and student data within data systems will yield instructional improvement fails to take into account local uses of data, and potential unintended consequences. Specifically, how data is used and interpreted for instructional improvement in local contexts has implications for test score validity. Haladyna, Nolen, and Haas (1991) refer to two major aspects of “test score pollution”: how educators prepare students to take test (such as the MSP blitz found in this study), and the various non-standard practices (such as focusing on “bubble kids”) and conditions (such as being a “Priority” school) under which tests are administered within local contexts. Moss (2013) argues that even with specified uses or “decision rules” accompanied with assessment tools, the emphasis on the use of tests as primary

sources of evidence creates a context for data use that “under anticipates the complexity of how test scores are being used locally, in action, by teachers and other education professionals in different contexts for their own purposes” (p.92). In this study, I suggest that the use of multiple measures may not be particularly useful for educators when there are clear identified issues with available assessments, such as lack of alignment or “fit” or validity of use for particular student populations such as emergent bilingual students.

The focus on testing within my school site was consistent with the research on how testing programs can often reduce time for instruction, create pedagogical strategies that “teach to the test” or narrow modes of instruction, and possibly reduce opportunities for teachers to build capacity to improve student learning and not just test score outcomes (Smith, 1991). For example, a study on school microprocesses of data use (Hubbard, Datnow, & Pruyne, 2014) found— similar to my findings—that teachers often used assessment data to focus on specific skills or benchmarks found in assessments. They also identified various incompatibilities of multiple reform initiatives within data use, and felt that their assessments did little to inform teachers how to improve ELL student outcomes within their specific curriculum, such as reporting issues such as too complex vocabulary, or not having specific information on how to scaffold ELL student learning. A particular contribution of my research highlights how educator learning for ELL support strategies within assessment talk was mediated by the ELL Support Specialist, particularly when not focused on interpreting student test scores. These findings underscore Datnow and Hubbard’s (2015) call for more research on the context

of teacher's use of interim benchmark data, particularly how teachers connect data to instruction.

Second, this line of research contributes to our emerging understanding of the complex contexts of data use for instructional reform and the opportunities for practitioner learning. This case study sits within the tradition of studies focused *inside* educational transactions between students, teachers, and curriculum content that constitute teaching and learning. Specifically, it is an example of data use in practice (Spillane, 2012) that focuses inside data-focused organizational routines for instructional improvement. As part of a line of research that seeks to understand the phenomenon of data use within the context of data use interventions, I sought to explore how data was interpreted within specific “processes, conditions, and contexts” (Coburn & Turner, 2011). In particular, I took a systems approach to studying data use (Honig & Venkateswaran, 2012) to consider how my site's embedded context of accountability testing and pressures alongside ELL language policies and district support structures impacted the goals and structure of exploring data within instructional leadership team meetings, and available practitioner expertise on ELL student support strategies. This body of research calls for continued examination of the nuances within data and evidence use practices for instructional improvement alongside an exploration of how specific contextual variables influence this phenomenon.

The study is connected to an emerging line of research within data and evidence use that seeks to unpack the microprocesses of data use (Little, 2012) by conceptually unpacking and investigating what and how educators learn when engaging with data and

with each other surrounding problems of practice. Schldkamp and Poortman (2015) highlight three primary factors that influence data use within teams: characteristics of data (such as access, availability; school organizational factors (such as goals, forms of leadership, support for teachers); and aspects of individual and team characteristics (such as knowledge and expertise, attitudes, collaboration, data literacy, knowledge of assessment, etc.). A particular theoretical strength of my case study focused on how exploring a particular school's organizational context allowed me to utilize assessment-as-practice theory (Jordan & Putz, 2004) within episodes of data talk to consider and compare forms of *documentary assessments* (i.e. sources of data and assessment in meetings) with the forms *inherent assessments* (i.e. perceptions of the situation, worldviews, expertise and pedagogical knowledge) used in data-focused conversations. Focusing on these various forms of "evidence" used within assessment talk highlighted the complexity of the sources of information and "data" that are brought into these spaces (Rainey, 2014), and the depth of interpretation of data (Gannon-Slater, La Londe, Crenshaw, Evans, Greene, & Schawandt, 2014) for decisions closely related to instruction versus a more compliance use of data for accountability.

My study also speaks to understandings of how educators use data to learn and make decisions for instructional improvement in two theoretical ways. One, it draws from current conceptions of data-driven decision making (Ikemoto & Marsh, 2007, and others) to consider the extent to which the types of "problem spaces" posed within forms of data impact how practitioners move from looking at data, to forming and sharing interpretations of data, to creating shared understandings and actionable decisions

intended for (or assumed to contribute to) instructional improvement. Secondly, I use the concept of *assessment-as-practice* (Jordan & Putz, 2004) to integrate sociocultural aspects of teacher learning and knowledge (Wenger, 1998; Star & Griesmer, 1989; Nolen, Horn, Ward, & Childers, 2011; Horn & Little, 2009; Jordan & Putz, 2004) within these DDDM processes by reframing these steps to include additional forms of “data” or evidence beyond summative assessment output, a range of practitioner interpretation in relation to the available evidence, I then begin to consider the forms of shared knowledge that may occur as an outcome of this process beyond an actionable decision. My conceptual integration of the ostensive and performative aspects of data use (Spillane, 2012), particularly a focus on aspects of communities of practice, allows for future research on how an educator’s vertical expertise (i.e., individual knowledge, skills, expertise) within data use contexts creates horizontal expertise (i.e., co-constructed knowledge through interaction) (Marsh, Bertrand, & Huguet, 2015). In addition, by using the concept of *inherent assessments*, I am able to consider how cognitive aspects of practitioner knowledge and skills alongside their beliefs, value systems, identity, and other epistemic elements are used to interpret assessment data or evidence into actionable knowledge and practices (Gummer & Mandinach, 2015).

My use of *boundary objects* within my study was analytically useful for identifying and investigating assessment talk. As stable representations of language and evaluation within assessments, boundary objects serve as the “problems of practice” or focus of assessment talk, that is particularly useful for examining the use of the same boundary object (such as a specific assessment, or test) across school, district, and

professional learning community contexts for instructional improvement (Nolen, Horn, Ward, & Childers, 2011). For example, a study on two middle school math teachers' interpretations of the same district assessment found that differing organizational cultures surrounding the notions of data and evidence use—data use for instructional management versus instructional improvement—created distinct framing and use of data within teacher workgroups that either enhanced or inhibited opportunities for professional learning (Horn, Kane, & Wilson, 2015). By being able to “hold constant” the district assessment boundary object across sites, these researchers were able to identify other important variables and dynamics within data use practice.

Finally, a notable contribution of this study is its focus on the intersection of accountability pressures and ELL language policies and curriculum, and how these exert an influence on data use practices within a school going through large demographic changes. Through my site selection, I was able to investigate the extent to which educators within a “Latino Diaspora” context were able to have access to and draw from research-based knowledge for the teaching and learning of English language learners. My finding of a lack of engagement with these ideas within organizational routines for instructional improvement is of particular import for educational leaders and policy makers within non-traditional areas of Latino and emergent bilingual settlement. For example, Turner's (2015) analysis of policymakers responses to demographic shifts in their communities found that their understandings of race, class, and immigration were often woven together and lead them to disregard key influences to student learning, such as ELL language issues. These findings highlight the importance of how the framing of

an activity can influence the policies, practices and processes aimed at instructional improvement (Coburn et al., 2009). Therefore, it is important to consider alternative framings and investigations of data use phenomena as part of the limitations of this research.

Unanswered Questions and Directions for Further Research

The impact of framing and facilitation of data use serves as a limitation to this particular study. For example, Jennings (2012) highlights how educators may have particular notions surrounding the uses of particular forms of data. While one could argue that data systems for accountability were meant to serve as diagnostic tools for improvement, she argues that data can also serve as a compass for particular changes or maintaining the status quo, monitoring student performance on goals, or as a legitimizer of decisions previously made. These patterns and uses of data were found in this study, and individual educators' perceptions of the uses and meanings of data-driven decision making may have been another contributing factor to the patterns found in my study. In essence, how someone understands the uses of data, and the goals of particular task is of particular significance in the investigation of data use in practice.

An educator's capacity to use data for decision making is also a contributing factor that was not fully explored in this study. While all of the educational leaders reported getting little to no training or support on using data, data literacy could have contributed to ways in which they were able to engage and interpret data and facilitate

discussion on the possible sources and actions for instructional improvement. Gummer and Mandinach (2015) highlight the emerging nature of the field's understanding of data literacy and outline a complex framework for understanding that considers three different domains for data use (teaching, content knowledge, and pedagogical content knowledge), alongside aspects of the inquiry cycle, which encompasses 59 elements of knowledge and skill in those components. This work suggests a more complex understanding of what educators are being asked to do, and what competencies they must develop when they engage with data to form actionable decisions. Additionally, the positioning of the facilitator, and the organizational cultures of the grade level and instructional leadership teams may have also impacted data use. Due to my limited engagement within the day-to-day goings-on within my school context, I was unable to comment directly on particular norms of interaction or conversational routines outside of the instructional leadership team that would have informed or shed light on particular PLC observations surrounding how the grade-level teacher teams interacted with each other. Finally, I was unable to track how particular understandings of the situation derived from collective interactions with data contributed to individual teacher learning and practice within the classroom.

These additional aspects of data use in practice that have not been well explored in my study highlight the need for building upon current emerging research in these areas that can examine data use across a variety of contexts to assess the extent to which different interactional variables impact data and evidence use practices. Important questions surrounding the nature of the influences that documentary assessments (i.e.,

student data, rubrics, etc.) have on the forms of discourse and meaning making within data use practices remain largely unanswered (Horn, Kane, & Wilson, 2015 is a notable exception). Specifically, further studies are needed that investigate how notions of learning found within tests and assessments trickle down and are translated and incorporated into the instructional core.

Further work in this line, including extensions of my own research, would make a continuing contribution to this area. For example, I visualize a next stage of my research that will utilize research methodology found within studies of actor-network theory (Latour, 1987; Fenwick, Edwards, & Sawchuck, 2011; Fenwick, 2012). In education, this methodology would examine how various actors and activities within assessment and data use “materialize or animate knowledges, identities, and action” (Koyama, 2013, p. 3; see also Koyama & Menken, 2013). In addition, I plan to further refine my conceptual framework to consider how affordances and constraints to discourse within data use practices can contribute to opportunities for teacher learning, as described in theories of distributed cognition and sociocultural aspects of conceptual understanding and knowledge through interaction (Greeno & van de Sande, 2007). Finally, future research in this area would also explore which forms of information, evidence, and knowledge surrounding ELL student learning need to be taken into consideration during data use practices in order to integrate what we know about supporting ELLs’ development of content, analytical practices, and language with formative and summative assessments, particularly within areas with rapid Latino and emergent bilingual student growth.

Implications for Practice and Policy

Understanding which forms of data and evidence educators need during cycles of instructional improvement is important for the practice of leading with data. While examining student output data did not provide this particular school with many opportunities to deeply engage with their assumptions and interpretations of data for instructional improvement, we know that access to particular tools or practices to support deeper inquiry with student output data can improve practitioner learning. For example, scholarship has pointed to the use of equity audits (Scheurich & Skrla, 2003) to uncover systemic inequities within a school, or to specific data use frameworks that integrate summative and formative assessments—Data Wise is one such approach as are aspects of “learning-focused” leadership (Knapp, Honig, Plecki, Portin, & Copland, 2014).

Of particular importance, given what did and did not happen in the data use routines I studied is that leaders need to find ways of bringing data on the instructional core itself into view alongside the student outcome data. A case in point: leaders using data for instructional improvement may also want to include more systematic collection of practitioners’ formative and inherent assessments of student learning found within instructional rounds, a process aimed at adult learning that focuses in on problems of practice and asks educators to focus on that particular issue during classroom observations that are then used for an evidenced-based discussion surrounding instructional improvement (Roberts, 2012; City, Elmore, Fiarman, & Teitel, 2009). By

utilizing an expanded form of “data” to include assessments of learning found in tests, and within teacher observation of practice and student learning, data use practices can then use “multiple-measures” aimed at improving the instructional core. In addition, leaders must enable teachers to have more flexibility to make instructional changes based on what they learn within these data and evidence-focused conversations, while also considering teacher feedback on the usefulness of available assessments (Datnow, 2011).

This work also has implications for policy surrounding punitive accountability measures and ELL language policies, curriculum and professional development. Educator data use practices are often “silo-ed” from each other (Thron, Meyer, & Gamoran, 2007). My study found that within their particular accountability context, there was pressure to use summative assessments for decision making despite their lack of perceived usefulness, and on helping students “pass” the test instead of focusing on improving overall classroom pedagogy. Lessening the imperative of meeting accountability mandates and instead building a learning culture within schools could help to shift the purposes and goals of data use from finding ways for students to meet standards toward building practitioner knowledge about supporting the learning of all students within their school. This helps to relieve the “problem of presentism” (Hargreaves & Shirley, 2009) by shifting short term goal-setting within schools towards longer term strategies for change, such as improving professional development and ongoing learning. In addition, districts should create systems across schools to help educators build connections across data sources, as well as transmit promising practices.

Shifting the purpose and scope of data use practices in these ways is particularly important within schools serving increasing numbers of English language learners and emergent bilingual students, as educators must make deep pedagogical shifts to support ambitious learning for ELLs (Heritage, Walqui, & Linqanti, 2015). Being new to their growing student population, it is difficult for educators to know what they don't know, which has significant implications for the types of problem framing and solving found within data use practice. In particular, Heritage, Walqui, and Linqanti (2015) argue that the role of assessment policies and data driven decision making within ELL-serving schools must privilege informed formative assessments as part of their comprehensive assessment system, which they view is essential in teaching ELLs due to its connections to teacher-student interactions around learning. Thus, educators within the "Latino Diaspora" need not only the time to develop capacity to serve ELL students (such as obtaining certification, building on-going PD and coaching strategies), but to also build capacity to bring in their expertise into data-focused spaces.

Closing Observations

In summary, the study provides insights into how aspects of NCLB accountability policies impact data use practices that can potentially lead to stagnant accountability gains within Latino and emergent bilingual-serving schools. Many studies of leadership take into consideration how the cultural diversity of a school impacts decision-making; yet, the impact of linguistic diversity in high-stakes environments is ignored or serves as an afterthought. These issues took center stage within this study and

my findings highlight the need for continued research within situated data use practices and microprocesses to consider the additional complexity to reform that these contexts hold. Furthermore, practitioners serving rapidly changing schools need better tools and understanding of how to apply effective leadership practices for Latino students, rather than a list of effective leadership traits or capacities.

For districts in states that have seen swift growth of their Latino student body, tools to implement accountability policy while building organizational capacity cannot come soon enough. These changes underscore our need for new models of accountability and improvement for Latino and emergent bilingual-serving schools. For that reason, this study of assessment practices of instructional leadership teams within Latino and emergent bilingual-serving elementary schools is an important first step in building integrated theory around organizational learning within data and evidence use practices for instructional improvement and capacity building.

REFERENCES

- Abedi, J., & Dietel, R. (2004). *Challenges in the No Child Left Behind Act for English Language Learners* (National Center for Research on Evaluation, and Student Testing (CRESST). Los Angeles: UCLA.
- Abedi, J., Bailey, A., Butler, F., Castellon-Wellington, M., Leon, S., & Mirocha, J. (2000/2005). *The Validity of Administering Large-Scale Content Assessments to English Language Learners: An Investigation From Three Perspectives*. (CSE Report 663). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Abedi, J., & Gándara, P. (2007). Performance of English Language Learners as a subgroup in large-scale assessment: Interaction of research and policy. *Educational Measurement: Issues and Practice*, 25(4), 36-36.
- Aguirre-Muñoz, Z., & Solano-Flores, G. (2010). Accountability and educational assessment for Latino English Language Learning students: Improving practices through multidisciplinary approaches. In E. Murillo Jr., S. Villenas, R. Trinidad Galvan, J. Sanchez Munoz & C. Martinez (Eds.), *Handbook of Latinos and Education: Theory, Research, and Practice* (pp. 503-522). New York: Routledge.
- Anderson, S. E., Leithwood, K., & Seashore Louis, K. (2012). Data use: An exploration from the district to the school. In K. Leithwood & K. Seashore Louis (Eds.), *Linking Leadership to Student Learning* (pp. 158-180). San Francisco, CA: Jossey-Bass.
- August, D., Goldenberg, C., & Rueda, R. (2010). Restrictive state language policies: Are they scientifically based. *Forbidden language: English learners and restrictive language policies*, 139-158.
- August, D. & Hakuta, K. (1997). *Improving schooling for language-minority children*. Washington, DC: National Academy Press.
- Ball, D. L., & Forzani, F. M. (2007). 2007 Wallace Foundation Distinguished Lecture-- What Makes Education "Research Educational"? *Educational Researcher*, 36(9), 529-540. doi: 10.3102/0013189x07312896
- Black, W. R. (2006). Constructing accountability performance for English Language Learner Students: An unfinished journey toward language minority rights. *Educational Policy*, 20(1), 197-224. doi: 10.1177/0895904805285948
- Brown, C.L. (2004). Content based ESL curriculum and academic language proficiency. *The Internet TESL Journal*, 10 (2). Retrieved April 23, 2015 from <http://iteslj.org/Techniques/Brown-CBEC.html>.
- Bruening, E. (2015). Doing it on their own: The experiences of two Latino English language learners in a low-incidence context. In E. T. Hamann, S. Wortham & E. Murillo Jr. (Eds.), *Revisiting Education in the New Latino Diaspora*. Charlotte, NC: Information Age Publishing.

- Bogdan, R., & Biklen, S. K. (1982). *Qualitative research for education: An introduction to theory and methods*. 2d ed. Boston: Allyn & Bacon.
- Bomer, R., Dworin, J. E., May, L., & Semingson, P. (2008). Miseducating teachers about the poor: A critical analysis of Ruby Payne's claims about poverty. *Teachers College Record*, 110(11).
- Cho, V., & Wayman, J. C. (2014). Districts' efforts for data use and computer data systems: The role of sensemaking in system use and implementation. *Teachers College Record*, 116(020206), 45.
- City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard Education Press.
- Coburn, C. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145.
- Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, 9(4), 173-206. doi: 10.1080/15366367.2011.626729
- Coburn, C. E., & Stein, M. K. (2006). Communities of practice theory and the role of teacher professional community in policy implementation. In M. Honig (Ed.), *New Directions in Education Policy Implementation: Confronting Complexity*. New York: State University of New York Press.
- Coburn, C. E., Toure', J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision making at the district central office. *The Teachers College Record*, 111(4), 1115-1161.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational evaluation and policy analysis*, 25(2), 119-142.
- Coldren, A. (2007). Spanning the boundary between school leadership and classroom instruction at Hillside Elementary School. In J. Spillane & J. Diamond (Eds.), *Distributed Leadership in Practice* (pp. 16-34). New York: Teachers College Press.
- Collier, V. P., & Thomas, W. P. (2004). The astounding effectiveness of dual language education for all. *NABE Journal of Research and Practice*, 2(1), 1-20.
- Contreras, F. E., & Stritikus, T. (2008). *Understanding opportunities to learn for Latino students in Washington*. Seattle WA: University of Washington, College of Education.
- Contreras, F. (2011). *Achieving equity for Latino students: Expanding the pathway to higher education through public policy*. New York: Teachers College Press.
- Datnow, A., Borman, G. D., Stringfield, S., Overman, L. T., & Castellano, M. (2003). Comprehensive school reform in culturally and linguistically diverse contexts: Implementation and outcomes from a four-year study. *Educational Evaluation and Policy Analysis*, 25(2), 143-170.
- Datnow, A. (2011). Collaboration and contrived collegiality: Revisiting Hargreaves in the age of accountability. *Journal of Educational Change*, 12(2), 147-158. doi: 10.1007/s10833-011-9154-1

- Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, 117(4).
- Darling-Hammond, L. (2010). *The flat world and education: How America's commitment to equity will determine our future*: Teachers College Press.
- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality? *Teachers College Record*, 106(6), 1145-1176. doi: 10.1111/j.1467-9620.2004.00375.x
- Dexter, S., Seashore Louis, K., & Anderson, R. E. (2009). The roles and practices of specialists in teamed institutional leadership. *Journal of School Leadership*, 19(4), 445-465
- Duran, R. P. (2008). Assessing English-Language Learners' achievement. *Review of Research in Education*, 32(1), 292-327. doi: 10.3102/0091732x07309372
- Elfers, A. M., & Stritikus, T. (2013). How school and district leaders support classroom teachers' work with English Language Learners. *Educational Administration Quarterly*. doi: 10.1177/0013161x13492797
- Elmore, R. F. (2000). *Building a new structure for school leadership*: The Albert Shanker Institute Washington, DC.
- Eckert, P., & McConnell-Ginet, S. (2013). *Language and gender*. Cambridge University Press.
- Erickson, F. (1985). *Qualitative methods in research on teaching* (pp. 119-62). E. Lansing MI: Institute for Research on Teaching, Michigan State University.
- Emerson, R.M., Fretz, R.I., & Shaw, L.L. (2011). *Writing ethnographic fieldnotes*. (2nd Ed). University of Chicago Press: London.
- Eisenhart, M. (2009). Generalization from qualitative inquiry. In Ercikan, K., & Roth, W-M. (Eds.) *Generalizing from educational research: Beyond quantitative and qualitative polarization*, 51-66. New York, NY: Routledge.
- Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly* 48(1): 94–118.
- Fenwick, T. (2007). Organizational learning in the "knots": Discursive capacities emerging in a school-university collaboration. *Journal of Educational Administration*, 45(2), 138-153.
- Fenwick, T. (2012). Co-production in professional practice: A sociomaterial analysis. *Professions and Professionalism*, 2(2). doi: 10.7577/pp/v2i1.323
- Fenwick, T., Edwards, R., & Sawchuk, P. (2015). *Emerging approaches to educational research: Tracing the socio-material*. Routledge: Chicago.
- Fink, E., & Resnick, L. B. (2001). Developing principals as instructional leaders. *Phi Delta Kappan*, 82(8), 598-606.
- Finnigan, K. (2010). Principal leadership and teacher motivation under high-stakes accountability policies. *Leadership and Policy in Schools*, 9(2), 161-189. doi: 10.1080/15700760903216174
- Fuhrman, S. H. (2003). *Redesigning accountability systems for education*. Philadelphia: Consortium for Policy Research in Education.

- Gallucci, C. (2008). Districtwide instructional reform: Using sociocultural theory to link professional learning to organizational support. *American Journal of Education*, 114, 541-581.
- Gallucci, C. (2003). Communities of practice and the mediation of teachers' responses to standards-based reform. *Education Policy Analysis Archives*, 11(35). Retrieved 6.6.15 from <http://epaa.asu.edu/epaa/v11n35>
- Gándara, P., & Conteras, F. (2009). *The Latino education crisis*. Cambridge, MA: Harvard University Press.
- Gándara, P., & Hopkins, M. (2010). *Forbidden Language: English Learners and restrictive language policies*. New York: Teachers College Press.
- Gee, J. P., & Green, J. L. (1998). Discourse analysis, learning, and social practice: A methodological study. *Review of Research in Education*, 23, 119-169.
- Gándara, P., & Baca, G. (2008). NCLB and California's English language learners: The perfect storm. *Language Policy*, 7(3), 201-216.
- Garcia, E., Arias, M. B., Harris Murri, N. J., & Serna, C. (2009). Developing responsive teachers: A challenge for a demographic reality. *Journal of Teacher Education*, 61(1-2), 132-142. doi: 10.1177/0022487109347878
- Glaser, B., & Strauss, A. (1967). *the discovery of grounded theory*. Aldine:Chicago
- Gonzalez, M. L. (2010). The critical role of all educators in the school success of Latino children. *Leadership and Policy in Schools*, 9(4), 479-494. doi: 10.1080/15700763.2010.493635
- Gonzalez, M. L., Huerta-Macias, A., & Tinajero, J. V. (1998). *Educating Latino Students: A Guide to Successful Practice*. Lanham MD: Rowman & Littlefield.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606-633.
- Greeno, J. G., & van de Sande, C. C. (2007). Perspectival understanding of conceptions and conceptual growth in interaction. *Educational Psychologist*, 42(1), 9-23.
- Gummer, E. S., & Mandinach, E. B. (2015). Building a conceptual framework for data literacy. *Teachers College Record*, 117(040305), 22.
- Hakuta, K. (2009). *Guidelines for the assessment of English Language Learners*. Princeton: Educational Testing Service.
- Hakuta, K. (2011). Educating language minority students and affirming their equal rights: Research and practical perspectives. *Educational Researcher*, 40(4), 163-174.
- Hakuta, K. (2015). Foreword. In M. Heritage, A. Walqui & R. Linn (Eds.), *English Language Learners and the New Standards*. Cambridge, MA: Harvard Education Press.
- Hakuta, K., Goto-Butler, Y., & Witt, D. (2000). *How long does it take English learners to attain proficiency?* The University of California Linguistic Minority Research Institute. Policy report 2000-1.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7.

- Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools, 4*(3), 221-239. doi: 10.1080/15700760500244793
- Halverson, R. R., & Clifford, M. A. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Educational Administration Quarterly, 42*(4), 578-619. doi: 10.1177/0013161x05285986
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The New instructional leadership: Creating data-driven instructional systems in school. [Feature]. *Journal of School Leadership, 17*(2), 159-194.
- Hamann, E. T. & Harklau, L. (2010). Education in the new Latino diaspora. Faculty Publications: Department of Teaching , Learning and Teacher Education. Paper 104. Retrieved from <http://digitalcommons.unl.edu/teachlearnfacpub/104>
- Hargreaves, A., & Shirley, D. (2009). The persistence of presentism. *Teachers College Record, 111*(11), 2505-2534.
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110. doi: 10.3102/0162373708317689
- Heilig, J. (2011). Understanding the interaction between high-stakes graduation tests and English language learners. *Teachers College Record, 113*(12).
- Heritage, M., Walqui, A., & Liguori, R. (2015). *English Language Learners and the new standards: Developing language, content knowledge, and analytical practices in the classroom*. Cambridge MA: Harvard Education Press.
- Honig, M. I. (2008). District central offices as learning organizations: How sociocultural and organizational learning theories elaborate district central office administrators' participation in teaching and learning improvement efforts. *American Journal of Education, 114*(4), 627-664.
- Honig, M., Copland, M., Rainey, L., Lorton, J. A., & Newton, M. (2010). *Central office transformation for district-wide teaching and learning improvement*. Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.
- Honig, M. I., & Louis, K. S. (2007). A new agenda for research in educational leadership: A conversational review. *Educational Administration Quarterly, 43*(1), 138-148. doi: 10.1177/0013161x06290646
- Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher, 33*(8), 16-30.
- 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.
- Hopkins, M., Lowenhaupt, R., & Sweet, T. M. (2015). Organizing English Learner instruction in new immigrant destinations district infrastructure and subject-specific school practice. *American Educational Research Journal, 52*(3), 408-439.

- Horn, I. S. (2007). Fast kids, slow kids, lazy kids: Framing the mismatch problem in mathematics teachers' conversations. *The Journal of the Learning Sciences*, 16(1), 37-79.
- Horn, I. S., Kane, B. D., & Wilson, J. (2015). Making sense of student performance data data use logics and mathematics teachers' learning opportunities. *American Educational Research Journal*, 0002831215573773.
- Hubbard, L., Datnow, A., & Pruyne, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation*, 42, 54-62.
- Huguet, A., Marsh, J.A. & Farrell, C. C. (2014). Building teachers' data-use capacity: Insights from strong and developing coaches. *Education Policy Analysis Archives*, 22 (52). Retrieved from <http://dx.doi.org/10.14507/epaa.v22n52.2014>.
- Hussar, W. J., & Bailey, T. M. (2011). *Projections of education statistics to 2019* (NCES 2011- 017). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision making. In P. A. Moss (Ed.), *Evidence and Decision Making*. Malden, MA: Blackwell Publishing.
- Ishimaru, A. (2014). Rewriting the rules of engagement: Elaborating a model of district community collaboration. *Harvard Educational Review*, 84(2), 188-216.
- Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record*, 114(110304), 23.
- Jordan, B., & Putz, P. (2004). Assessment as practice: Notes on measures, tests, and targets. *Human Organization*, 63(3), 346.
- Kaminiski, R., & Cummings, K.D. (2007). *DIBELS: Myths and Facts*. Dynamic Measurement Group, Eugene, OR. Retrieved from https://dibels.org/papers/Myths_0208.pdf
- Kantor, H., & Lowe, R. (2006). From new deal to no deal: No Child Left Behind and the devolution of responsibility for equal opportunity. *Harvard Educational Review*, 76(4), 474-502.
- Kowalski, T. J. (2010). *The school principal: vVisionary leadership and competent management*. New York: Routledge.
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., 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.
- Knapp, M. (2008). How can organizational and sociocultural learning theories shed light on district instructional reform? *American Journal of Education*, 114(4), 521-539.
- Knapp, M. (2012). Foreword. In K. Leithwood & K. S. Louis (Eds.), *Linking leadership to student learning* (pp. xix-xxi). San Francisco: Jossey-Bass.
- Knapp, M. S., Copland, M. A., & Swinnerton, J. A. (2007). Understanding the promise and dynamics of data-informed leadership. In P. A. Moss (Ed.), *Evidence and decision making* (pp. 74-104). Malden, MA: Blackwell Publishing.

- Knapp, M. S., Mkhwanazi, S. T., & Portin, B. S. (2011). School-based instructional leadership in demanding environments: New challenges, new practices. In A. R. Shoho, B. G. Barnett, & A. K. Tooms (Eds.), *The changing nature of the instructional leadership in the twenty-first century*. Charlotte, NC: Information Age Publishing.
- Koyama, J., & Menken, K. (2013). Emergent bilinguals: Framing students as statistical data? *Bilingual Research Journal*, 36(1), 82-99. doi: 10.1080/15235882.2013.778223
- Koyama, J. (2013). Principals as bricoleurs: Making sense and making do in an era of accountability. *Educational Administration Quarterly*. doi: 10.1177/0013161x13492796
- La Londe, P. G., Crenshaw, H., Evans, M., Gannon-Slater, N., Greene, J. C., & Schwandt, T. A. (2014). Considering equity and opportunity in teacher discourse on students in data-driven decision-making. Paper Presentation at University Council of Education Administration, Washington D.C.
- Lasky, S., Schaffer, G., & Hopkins, T. (2008) Learning to think and talk from evidence: Developing system-wide capacity for learning conversations. In L. M. Earl & H. Timperley (Eds.), *Professional learning conversations: Challenges in using evidence for improvement* (pp.95-108). Springer.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.
- Lave, J. & Wenger E. (1991). *Situated learning: Legitimate peripheral participation*. New York, NY: Cambridge University Press.
- Leithwood, K., & Riehl, C. (2005). What do we already know about educational leadership? In W. A. Firestone & C. Riehl (Eds.), *A new agenda for research in education leadership* (pp. 1-11). New York: Teachers College Press.
- Leithwood, K., Patten, S., & Jantzi, D. (2010). Testing a conception of how school leadership influences student learning. *Educational Administration Quarterly*, 46(5), 671-706. doi: 10.1177/0013161x10377347
- Leithwood, K., Steinbach, R., & Jantzi, D. (2002). School leadership and teachers' motivation to implement accountability policies. *Educational Administration Quarterly*, 38(1), 94-119. doi: 10.1177/0013161x02381005
- Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797-832. <http://www.jstor.org/stable/3699464>
- Levinson, B. A., Sutton, M., & Winstead, T. (2009). Education policy as a practice of power theoretical tools, ethnographic methods, democratic options. *Educational Policy*, 23(6), 767-795.
- Linn, R. R. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives*, 13, 33.

- Little, J. W. (2003). Constructions of teacher leadership in three periods of policy and reform activism. *School Leadership & Management*, 23(4), 401-419. doi: 10.1080/1363243032000150944
- Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, 118(2), 143-166.
- Lucas, T., Villegas, A. M., & Freedson-Gonzalez, M. (2008). Linguistically responsive teacher education: Preparing classroom teachers to teach English Language Learners. *Journal of Teacher Education*, 59(4), 361-373. doi: 10.1177/0022487108322110
- Malagon, H., McCold, P., & Hernandez, J. (2011). *Educating English Language Learners in Washington state 2009-10*. Retrieved from <http://www.k12.wa.us/MigrantBilingual/LegisReports.aspx>
- Mangin, M. M. (2007). Facilitating elementary principals' support for instructional teacher leadership. *Educational Administration Quarterly*, 43(3), 319-357. doi: 10.1177/0013161x07299438
- March, J.G. (1994). *A primer on decision making: How decisions happen*. New York: The Free Press.
- Marsh, J. A. (2012). Interventions promoting educators use of data: Research insights and gaps. *Teachers College Record*, 114, 48.
- Marsh, J. A., & Farrell, C. C. (2014). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*. doi: 10.1177/1741143214537229
- Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making A framework for understanding capacity building. *Educational Management Administration & Leadership*, 43(2), 269-289.
- Matsumura, L.C., Sartoris, M., Bickel, D.D., & Garnier, H.E. (2009) Leadership for literacy coaching: The principal's role in launching a new coaching program. *Educational Administration Quarterly*, 45(5), 655-693.
- Maxwell, J. A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage
- Means, B., Padilla, C., & Gallagher, L. (2010). *Use of education data at the local level: From accountability to instructional improvement*. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation, and Policy Development.
- Menken, K. (2006,). Teaching to the test: How standardized testing promoted by No Child Left Behind impacts language policy, curriculum, and instruction for English language learners. *Bilingual Research Journal*, 30(2), 521-546.
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon, UK: Multilingual Matters.
- Maxwell, J. A. (2005). *Qualitative research design : An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Mehan, H. (1992). Understanding Inequality in Schools: The Contribution of Interpretive Studies. *Sociology of Education*, 65(1), 1-20.

- Menken, K. (2006). Teaching to the test: How No Child Left Behind impacts language policy, curriculum, and instruction for English Language Learners. *Bilingual Research Journal*, 30(2), 521-546. doi: 10.1080/15235882.2006.10162888.
- Menken, K. (2010, Spring). No Child Left Behind and English language learners: The challenges and consequences of high-stakes testing. *Theory Into Practice*, 49: 121-128.
- Merriam, S. B. (2009). *Qualitative research*. San Francisco, CA: John Wiley & Sons, Inc.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis : An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2013). *Qualitative data analysis: A methods sourcebook*. SAGE Publications, Incorporated.
- Minaya-Rowe, L., & Ortiz, J. A. (2010). Boosting the professional development of teachers of Latino English Language Learners. In E. Murillo Jr., S. Villenas, R. Trinidad Galvan, J. Sanchez Munoz & C. Martinez (Eds.), *Handbook of Latinos and education: Theory, research, and practice* (pp. 523-530). New York: Routledge.
- Moss, P. A. (2012). Exploring the macro-micro dynamic in data use practice. *American Journal of Education*, 118(2), 223-232.
- Moss, P. A. (2013). Validity in action: Lessons from studies of data use. *Journal of Educational Measurement*, 50(1), 91-98.
- Nolen, S. B., Horn, I., Ward, C., & Childers, S. (2011). Novice teacher learning and motivation across contexts: Assessment tools as boundary objects. *Cognition and Instruction*, 29(1), 88-122. doi: 10.1080/07370008.2010.533221
- Nolen, S. B., & Ward, C. J. (2008). Sociocultural and situative approaches to studying motivation. In M. L. Maehr, S. Karabenick & T. Urdan (Eds.), *Social Psychological Perspective on Motivation and Achievement* (Vol. 15, pp. 425-460). London: Emerald Press.
- Pandya, J. Z. (2011). *Overtested: How high-stakes accountability fails English Language Learners*. New York: Teachers College Press.
- Perrone, T. (2014). *Score reports overview: Washington English Language Proficiency Assessment 2014* [pdf document]. Retrieved from: www.k12.wa.us/assessment/EL/pubdocs/ReportsOverviewWELPA6-4-2014.pdf.
- Pentland, B. T., & Feldman, M. S. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48(1), 94-118.
- Pollock, M. (Ed.). (2008). *Everyday antiracism: Getting real about race in school*. The New Press.
- Portin, B. S., Knapp, M. S., Dareff, S., Feldman, S., Russell, F. A., Samuelson, C., & Yeh, T. L. (2009). *Leadership for learning improvement in urban schools*. Seattle: University of Washington, Center for the Study of Teaching & Policy.
- Pullin, D. C., & Haertel, E. H. (2008). Assessment through the lens of "opportunity to learn". In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel & L. J. Young (Eds.),

- Assessment, equity, and opportunity to learn* (pp. 17-41). Cambridge: Cambridge University Press.
- Office of Public Instruction (2013). *Washington State Report Card*. Retrieved from <http://reportcard.ospi.k12.wa.us/summary.aspx?groupLevel=District&schoolId=1&reportLevel=State&year=2013-14>
- Office of Public Instruction (2014). *UPDATE: Transitional Bilingual Instruction Program (TBIP) 2013-14*. Retrieved from <http://www.k12.wa.us/LegisGov/2013documents/TransBilingual2013.pdf>
- Office of Public Instruction (2015). *Washington English Language Proficiency Assessment (WELPA)*. Retrieved April 23, 2015, from <http://www.k12.wa.us/Assessment/EL/default.aspx>.
- Rader-Brown, L., & Howley, A. (2014). Predictors of the instructional strategies that elementary school teachers use with English language learners. *Teachers College Record, 116*, 1-34.
- Rainey, L. (2014) What are you driving at?: How school leaders use data when making school-level decisions about instructional improvement. Unpublished doctoral dissertation. University of Washington, Seattle, WA.
- Reed, B., & Railsback, J. (2003) *Strategies and resources for mainstream teachers of English language learners*. Portland, OR: Northwest Regional Educational Laboratory.
- Reitzug, U. C., West, D. L., & Angel, R. (2008). Conceptualizing instructional leadership: The voices of principals. *Education and Urban Society, 40*(6), 694-714. doi: 10.1177/0013124508319583
- Riehl, C. (2005). Educational leadership in policy contexts that strive for equity. *International handbook of educational policy*, 421.
- Scanlan, M., & Lopez, F. (2012). ¡Vamos! How school leaders promote equity and excellence for bilingual students. *Educational Administration Quarterly, 48*(4), 583-625. doi: 10.1177/0013161x11436270
- Scheurich, J. J., & Skrla, L. (2003). *Leadership for equity and excellence: Creating high-achievement classrooms, schools, and districts*. Thousand Oaks, CA: Corwin Press.
- Schildkamp, K., & Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers College Record, 117*(040310), 42.
- Shin, H. B., & Kominski, R. A. (2010). *Language use in the United States: 2007 American Community Survey Reports, ACS-12*. Washington, DC.: U.S. Census Bureau.
- Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20*(5), 8-11.
- Smylie, M. A., & Denny, J. W. (1990). Teacher leadership: Tensions and ambiguities in organizational perspective. *Educational Administration Quarterly, 26*(3), 235-259.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professional in Berkeley's Museum of Vertebrate Zoology. *Social Studies of Science, 19*(3), 387-420.

- Solorzano, R. W. (2008). High stakes testing: issues, implications, and remedies for English Language Learners. *Review of Educational Research*, 78(2), 260-329. doi: 10.3102/0034654308317845
- Spillane, J., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 386-431.
- Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113-141.
- Supovitz, J. A. (2012). Getting at student understanding: The key to teachers' use of test data. *Teachers College Record*, 114(11), 29.
- Timperley, H. (2009). Evidence-informed conversations making a difference to student achievement. In L. Earl & H. Timperley (Eds.), *Professional learning conversations: Challenges in using evidence for improvement* (Vol. 1, pp. 69-79): Springer Netherlands.
- Tinajero, J. V., Munter, J. H., & Araujo, B. (2010). Best practices for teaching Latino English Learners in U.S. schools. In E. Murillo Jr., S. Villenas, R. Trinidad Galvan, J. Sanchez Munoz, & C. Martinez (Eds.), *Handbook of Latinos and education: Theory, research, and practice*. New York: Routledge.
- Thompson, M.S., DiCerbo, K.E., Mahoney, K. and MacSwan, J. (2002). *¿Exito en California? A validity critique of language program evaluations and analysis of English learner test scores*. *Education Policy Analysis Archives*, 10(7). Retrieved 6.6.15 from <http://epaa.asu.edu/epaa/v10n7/>.
- Thorn, C., Meyer, R. H., & Gamoran, A. (2007). Chapter 14: Evidence and decision making in education systems. *Yearbook of the National Society for the Study of Education*, 106(1), 340-361.
- Turner, E. O. (2014). Districts' responses to demographic change: Making sense of race, class, and immigration in political and organizational context. *American Educational Research Journal*, 52(1), 4-39. doi: 10.3102/0002831214561469
- U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2012). *National Evaluation of Title III Implementation—Report on State and Local Implementation*. Washington, DC.
- U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service (2012). *National Evaluation of Title III Implementation*, Washington, DC.
- Valli, L., Croninger, R., Chambliss, M., Graeber, A. O., & Buese, D. (2008). *Test driven: High-stakes accountability in elementary schools*. New York: Teachers College Press.
- Weiss, J. A. (2012). Data for improvement, data for accountability. *Teachers College Record*, 114, 1-7.
- Wayman, J. C., & Stringfield, S. (2006). Technology-supported involvement of entire faculties in examination of student instructional improvement. *American Journal of Education*, 112(4), 549-571.

- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. *School Effectiveness & School Improvement, 23*(2), 159-178. doi: 10.1080/09243453.2011.652124
- Wenger, E. (1998). *Communities of practice: Learning, meaning and identity*. Cambridge: Cambridge University Press.
- Wells, G. (1996). Using the tool-kit of discourse in the activity of learning and teaching. *Mind, Culture, and Activity, 3*(2), 74-.
- Wolcott, H. F. (2005). *The art of fieldwork*. Lanham MD: Rowman Altamira.
- Wortham, S., Hammond, E. T., & Murillo Jr., E. G. (2001). Education and policy in the new Latino Diaspora. In S. Wortham, E. T. Hammond & E. G. Murillo Jr. (Eds.), *New Latino diaspora: Policy and the politics of identity* (pp. 1-16). Westport, CT: Ablex.

APPENDIX A

Summary of Student Assessments during the 2012-2013 School Year

Acronym	Full Name	Description	Purpose of Assessment	Interpreting Results	Use in Assessment Practices
MAP (Source: Seattle Public Schools)	Measures of Academic Progress	Computer based assessment program produced by the Northwest Evaluation Association (NWEA). Each child spends about 45 minutes completing each test in reading and math, in the winter and spring with an option to test in the Fall. Aligned to WA state standards and is used in 131 districts.	Administered to 2-6 th students. Provides teachers with information to improve student learning Monitors academic growth Informs students, families about skills Makes data-driven decisions about instruction	2 scales — RIT and Percentiles RIT (Rasch units) show students' current achievement on a scale that is independent of grade level. Shows growth over time. Percentiles - Used to compare students of similar age and grade level. Lexile Scores - Used to show where students are at on a Lexile scale, which is a framework for reading to measure student's reading level and growth.	MAPS testing is done three times a year. Scores are shared with the School Board, periodically at Parent Group Meetings, parent-teacher conferences. Used in leadership team meetings and PLCs. Often compared against DIBELS scores.
DIBELS (Source: "DIBELS: Myths and Facts" by Dynamic Measurement Group)	Dynamic Indicators of Basic Early Literacy Skills	A set of procedures and measures for assessing the acquisition of early literacy and reading skills from K-6 th grades. Designed to use "within a formative assessment process to evaluate the effectiveness of interventions for those children receiving support in order to make changes when indicated to maximize student learning and growth"	Administered in grades K-6 th Provide teachers with information to match the amount and type of instructional support needed by students—not as a sole measure of student success. Can be aggregated to a systems level to be used formatively to identify needs for supports	Uses Outcome-Driven Model DIBELS Benchmark Goals Phoneme Segmentation Fluency (PSF) is designed to be an indicator of student progress toward long-term phonemic awareness. Initial Sound Fluency (ISF) indicator of child's knowledge and awareness of initial sounds in words Nonsense Word Fluency (NWF)	DIBELS testing is done three times a year Scores are shared with the School Board, periodically at Parent Group Meetings, parent-teacher conferences. Used in leadership team meetings and PLCs. Often compared against MAPs scores.

<p>WELPA</p> <p>Source: OSPI</p>	<p>Washington English Language Proficiency Assessment</p>	<p>Determines student eligibility for English language development (ELD) services.</p> <p>The WELPA annually assesses growth in English language development by the state's English language learners. This assessment tests reading, writing, listening and speaking knowledge and skills.</p>	<p>Consists of two tests: Placement Test to determine initial student eligibility for English language development (ELD) services.</p> <p>Annual Test is given to all students who qualified for ELD services with a Placement Test. It measures students' growth in English language knowledge and skills. Results from this test determine which students are eligible to continue to receive ELD services.</p>	<p>A score at Levels 1, 2, or 3 on the WELPA determines that the student will continue in program. Students successfully transition from the program when they meet the exit criteria on the WELPA.</p>	<p>Migrant/Bilingual students take the WELPA in February. Results post in the summer and teachers receive students current level before the school year begins.</p>
<p>MSP</p> <p>Source: OSPI</p>	<p>Measure of Student Proficiency</p>	<p>The Measurements of Student Progress (MSP) is the state's exam for students in grades 3-8.</p> <p>Students are tested in reading (grades 3-8), mathematics (grades 3-8) writing (grades 4 and 7) and science (grades 5 and 8).</p>	<p>Used for school, district and student accountability. At the school and district levels, state tests are used to help determine Annual Measurable Objectives (AMOs) and close proficiency gaps as a part of the Elementary and Secondary Education Act.</p> <p>Reported on Washington State Report Cards.</p>	<p>A student's performance on the MSP, HSPE and EOC is reported using "scale scores." Scale scores are three-digit numbers that are used to place the student into one of four levels:</p> <p>Level 4: Advanced (exceeding state standard)</p> <p>Level 3: Proficient (meeting state standard)</p> <p>Level 2: Basic (not meeting state standard)</p> <p>Level 1: Below Basic (not meeting state standard)</p>	<p>Used as benchmark for student learning for the school year and as indicators of student growth over the year.</p> <p>Used for external stakeholder progress monitoring.</p>

VITA

Kathryn E. Torres, Ph.D., obtained her doctorate at the Educational Leadership, Policy, and Organization studies program at the University of Washington. Her work focuses on school leadership, and the impact of accountability and ELL language contexts on educators opportunities to learn, particularly within data-focused inquiry.

While at the University of Washington, she was an IES Pre-doctoral Fellow in the Collaborative Researchers for Education Sciences Training Program (CREST), which prepares scholars to be equipped in interdisciplinary research and evaluation of K-20 policy interventions and associated programs through mixed-methods research. She is also affiliated with the University Council for Educational Administration (UCEA) as a Jackson Scholar Alumnus and former UCEA Program Center Fellow. Her work has appeared in the book *Revisiting Education in the New Latino Diaspora: One in Twelve and Rising*, and the *Journal of Social Issues*. She was previously a research fellow in the *Developing Leaders to Support Diverse Learners Research and Development Initiative* through UCEA. Currently, she is a part of the Gates-funded *Equitable Parent-School Collaboration* research project, headed by Drs. Ishimaru and Lott.

Prior to entering higher education, Kathryn served as an AmeriCorps member in Communities in Schools working as a tutor, mentor, and instructor to three 4th grade bilingual classrooms in Austin, TX. While at the University of Washington, her work has focused on student, teacher, and parent perceptions of Latino student experiences in WA State; exploring post-secondary achievement of Running Start students in WA State; and examining one district's use of longitudinal data, monitoring and assessment practices focused on reducing student drop-outs. She has also taught at the undergraduate level focused on contemporary issues in K-12 Education, and at the graduate level focused on an introduction to educational leadership. Kathryn received her M.Ed. at the Harvard Graduate School of Education.