

UNPACKING ASSESSMENT:
UNDERSTANDING TEACHER AGENCY IN THE
CREATION AND USE OF LOCAL ASSESSMENTS

by

Duncan Wilson

Dissertation Committee:

Professor Thomas Hatch, Sponsor
Professor Daniel Friedrich

Approved by the Committee on
the Degree of Doctor of Education

Date 13 February 2019

Submitted in partial fulfillment of the
requirements for the Degree of Doctor of Education in
Teachers College, Columbia University

2019

ABSTRACT

UNPACKING ASSESSMENT: UNDERSTANDING TEACHER AGENCY IN THE CREATION AND USE OF LOCAL ASSESSMENTS

Duncan Wilson

Current research into data use suggests that the intersection of assessment, reform, and public accountability and their impact on teacher practice are under-conceptualized and not well researched, especially when the data are derived from assessments that are locally produced by teachers. This qualitative case study investigates the development and use of local assessments in public schools in order to better understand how the act of creating local assessments and using data derived from local assessments influences teacher practice within the current state accountability policy environment. The conceptual frame and methodology of the study draw from existing literature on data use, teacher agency, as well as sense making theory. In particular, data use literature and sense making theory view teacher practice and teacher agency as situated in social and institutional contexts. Data for this study are derived from a year-long study of small groups of teachers in one suburban New York district who are engaged in professional development to develop and use assessments to inform their practice. The teams were trained through the Performance Assessment Design Initiative (PADI) where they developed learning objectives, rubrics as well as pre- and post-assessments based on state and local standards. The data in the study include assessment materials, non-participant observations, and individual interviews of teachers and team facilitators.

The findings are threefold: (1) Teachers engaged in local assessment design tended to favor data derived immediately from classroom observations over waiting for results from summative performances. (2) Even when state testing is not perceived as a barrier, indirect influences of accountability still constrain local assessment creation and use. (3) Despite many benefits described, “localness” was not demonstrated to be a mechanism of individual or organizational change. In addition to informing practitioners engaged in the creation and use of local assessments, findings also have the potential to inform policymakers by better understanding the link between classroom formative assessment and high stakes accountability. Findings also inform data use scholarship by examining the creation and use of a kind of assessment not previously studied. They also reveal the need for more research into the collective practices of teachers and facilitators in the creation and use of assessments and assessment data.

© Copyright Duncan Wilson 2019

All Rights Reserved

ACKNOWLEDGMENTS

I would like to acknowledge and thank the following people who supported me, not only through this project, but during my entire Doctoral Degree Program:

First, my family. To my parents Roger and Margaret. Thank you for allowing me to take 47 years from when you first put me in school until I finally finished! What can I say, some people just take a little longer than others! To my wife Mary Elizabeth and son Aidan. Thank you for putting up with many nights and weekends behind a computer or away from birthday parties, swim meets, and projects around the house. You inspired me to keep going!

Second, to Marty Brooks, Kathleen Reilly, and Giselle Martin-Kniep. Thank you for bringing me into the PADI Project at Tri-States. Your trust in me was invaluable. You showed me that it is possible to blend the world of research with the world of practice to help us all learn and grow.

Third, to the leaders and teachers at Suburban Woods. Thank you for letting me share your work for over a year and for allowing me to observe great teaching and facilitation in action. My hope is that the experience was as transformational for you as it was for me.

Finally, thank you to my committee, especially to my Sponsor, Tom Hatch, and my Second Reader, Daniel Friedrich. You pushed me farther than I could have ever pushed myself. Plus you were patient when the pace of my work took a backseat to career and family. You have made me a better thinker, a better writer, and a better practitioner. Thank you, Lucy Calkins, for gently pushing me whenever you ran into me at the Reading and Writing Project. Your belief in me was an inspiration. Thank you, James Purpura, for your insight and articles during and after my defense. I hope that we can continue the assessment "conversation" that we agree is still greatly needed in this country.

D. W.

TABLE OF CONTENTS

	Page
Chapter I—INTRODUCTION	1
Assessment, Reform, and Accountability	1
Two Theories of Action	2
Background: Into the Black Box	8
Teacher Agency: A Conceptual Frame	13
Sense Making Theory: A Theoretical and Methodological Frame	15
Problem.....	17
Purpose and Research Questions	18
Research Questions.....	10
Rationale.....	19
Significance.....	20
Chapter II—LITERATURE REVIEW	23
Introduction	23
Methodology	24
Agency	25
Data Use Research and Methodologies	36
Data Use as a Situated Phenomenon	37
Two Paths Emerge.....	43
Classroom Based Formative and Performance Assessment: "To Use Their Minds Well"	45
Synthesis.....	57
Chapter III—METHODOLOGY	60
Introduction	60
The Role of Sense Making in My Methodology	61
PADI: The Case Defined.....	63
A Summary of the PADI Process.....	64
Research Questions.....	66
Research Sample	67
The Shore School.....	68
The Hill School	70
Selection Considerations	71
Introducing the Participants.....	73
Site #1: The Shore School.....	74
Lily.....	74
Paula	75
Site #2: The Hill School	76
Amy	76
Tina	78
Data Collection	79
Interview Protocols	83
Data Analysis	86
Pilot Work	91
Validity and Limitations.....	92
Conclusions	94

Chapter IV—FINDINGS	95
Research Questions Introduction: Research Focus Changes from the Original Plan.....	96
Research Question 1	99
Finding One	99
Finding Two	103
Finding Three	110
Research Question 2	115
Finding One	116
Finding Two	120
Finding Three	126
Research Questions 3 and 4: Barriers and Support.....	131
Barriers	131
Finding #1: Mandated curriculum requirements (sometimes referred to as the "pacing calendar") were described as a barrier.....	131
Finding #2: State assessments were described as a barrier	134
Support.....	135
Finding #1: Both teams expressed feelings of support from the district culture and from leadership that helped make the challenges of the work achievable and worthwhile	135
Finding #2: Both teams saw the facilitator as taking actions to provide support, but in different ways.....	136
Finding #3: In contrast to the teachers, both facilitators had intention to share the PADI work with other teachers in their schools.....	138
Conclusion.....	142
Chapter V—DISCUSSION OF FINDINGS	144
Discussion of Findings	145
Data Use Assumptions Challenged	145
Where and When Data are Used: in Classrooms	146
How Data are Used by Teachers.....	147
What Constitutes Data: Observations Trusted over Numbers.....	148
Influences of Accountability on Teacher Agency	150
The Limited Influences of High-stakes Testing	151
Indirect Influences of Accountability	151
Formative Assessment Processes Explored: The Impact of "Localness"	154
The Positive Impacts of "Localness"	154
"Using their minds well" and validity	154
Authentic tasks	155
Empowerment from design and making judgments	156
"Localness" as a Mechanism for Changes in Practice and Outcomes	156
Influences of Environment on Teacher Agency.....	158
Culture and agency.....	158
Impact of differences in facilitation on agency	159
Limited Impact of "Professional Accountability" on Collectives and Organizations.....	161
Limitations.....	163

Limitations of Design	163
Limits to Generalizability	165
The Discourse of Data and Change: Limitations as Seen Through a Post-structuralist Lens	166
Pushing Back on the Post-structural Critique	168
Conclusion.....	169
Data Use Assumptions Challenged	169
Influences of Accountability of Formative Assessment Processes.....	170
Influences of Facilitation of Formative Assessment Processes	172
Professional Accountability and the Role of "Localness"	172
Positive impacts of localness	173
Localness as a mechanism for change	173
Implications.....	175
Implications for Practitioners	176
Implications for Research	178
Implications for Policy as it Relates to School Reform	180
REFERENCES	181
APPENDICES	
Appendix A—Informed Consent.....	205
Appendix B—Accountability Assessment: Sorting the Bad Apples.....	207
Appendix C—PADI Performance Task Template A	212
Appendix D—PADI Performance Task Template B—Pre/Post	224
Appendix E—Pilot Interview, April 1, 2015.....	238

LIST OF TABLES

Table		Page
1	Team Goals and Plans Summarized	65
2	Team Members and Goals Compared	74
3	Overview of Information Needed	80
4	Timeline of Data Gathering.....	82
5	Interview Questions	84
6	Initial Codes Compared to Revised Codes and Sub-codes	89
7	Code Use and Code Co-occurrence.....	90
8	Team Original Plans and Modifications Summarized	97

Chapter I

INTRODUCTION

Assessment, Reform, and Accountability

The nature and role of assessment and assessment data use have been at the center of debates about how to reform public education for well over two decades, both in the United States and internationally. At the same time, data use scholars (Coburn & Turner, 2011; Little, 2012; Spillane, 2002, 2012) posit that the intersection of assessment, reform, and public accountability and their impact on teacher practice are under-conceptualized and not well researched. Little (2012) adds,

The rhetoric of “data-driven decision making” and “evidenced-based practice” has gained widespread purchase among education leaders and policy makers but has generated little research that investigates how this rational discourse plays out in the system of everyday practice that makes up schooling. (p. 143)

More broadly, Kliebard (1993) said the story of public education is not one of slow and steady progress but is rather the story of multiple points of view competing for the attention of teachers, administrators, policymakers, and the public at large in our quest to make public education work. Kliebard (2002) later wrote of assessment in the form of testing specifically: “The problem is that, in and of itself, testing is not a reform at all; it is at best a measure of success and in the right circumstances may become a spur to reform” (p. 1). More recently, scholars have suggested that current reform environment and its emphasis on assessment in the form of large-scale standardized tests as well as the engagement with evidence of student achievement is an increasingly public process

(Coburn & Turner, 2011; Little, 2012; Rothstein, 2006). Just as Kliebard argues that multiple points of view compete within the larger story of public education, the story of assessment's role whether in the form of standardized test or other public measures of performance to "measure" and to "spur" reform has also been contested.

Two Theories of Action

James Popham's (2009) simple bifurcation of assessment into two "camps" or approaches also suggests two competing theories of action that reformers are currently taking. The first approach named by Popham, "accountability assessment," stresses the importance of annual high-stakes testing as manifested in the No Child Left Behind legislation in 2001 and the Race To The Top legislation in 2010, as well as New York State statute 3012-C in 2011. In this first approach, assessments are "outsourced" or created externally from schools and reflect externally generated standards. Standardized testing and its emphasis on external measures of reliability create conditions where notions of transparency and accountability have been transformed into what Ball and others describe as an "audit society" (Ball, 2009, p. 678). In an audit society, accountability is focused on items that can be measured—on outcomes as opposed to inputs and processes (O'Day, 2002). Truth is found in numbers; therefore, a person's worth—a student's or a teacher's—is seen through the lens of measurement (Porter, 1995; Taubman, 2009). Over the past two decades, accountability assessment has been adopted by policymakers who are attracted to cost efficiency and scalability, which allow for the evaluation of large numbers of students as well as the comparison of results across classrooms, schools, and even nations (Hamilton, 2003; Linn, 2000). In addition, it has become an increasingly global phenomenon (Kellaghan, 2001; Steiner-Khamisi & Waldow, 2012). Many nations see assessment as a way to standardize curriculum and bring educational practices and outcomes in line with global practice to be competitive in the globalized labor market (Kellaghan, 2001; Mourished, Chijioke, & Barber, 2010).

Michael Apple has analyzed and critiqued public school accountability policy for several decades. His work ties this trend in accountability to the larger trend of neoliberalism. Apple (1997) problematizes the assumptions and power structures that created current reform policy as well as its impact on students from many social classes. Apple was critical of the current metanarrative even before current federal and state policies were enacted: "It is becoming ever more clear that—although couched in democratic rhetoric—the ultimate effect of 'reforms' such as privatized voucher and choice plans, national curricula, and national testing will be to exacerbate existing inequalities" (p. 167). In this same article, he searches for answers as to why education policy has taken such a conservative turn. Apple laments, "The social democratic goal of expanding equality of opportunity (itself a rather limited reform) has lost much of its political potency" (p. 169). He points out that hidden in the current policy environment is a "we/they" binary that "distance[s] most people of color, women (i.e. feminists), gays, and lesbians, and other from the community of worthy individuals" (p. 170). Seven years later, just three years after No Child Left Behind was passed, Apple (2004) solidifies his critique of the current policy environment by describing the results he predicted. He says that the policies that create accountability assessment "represent a subtle but crucial shift in emphasis ... from student needs to student performance and from what the school does for the student to what the student [and teacher] does for the school" (p. 20). Here students are seen almost as commodities capable of earning for the school test results that will reward the school or at least prevent sanctions. Perhaps Apple (2004) says it best here: "The seemingly contradictory discourse of competition, markets, and choice on the one hand and accountability, performance objectives, standards, national testing, and national curriculum on the other hand have created such a din that it is hard to hear anything else" (p. 15). What is perhaps most insidious then about the current policy environment emphasizing "accountability assessment" is its ability to silence dissent and therefore many attempts to unravel it.

The second approach Popham names, “classroom assessment,” has stressed inputs and processes over outcomes (O’Day, 2002)—the importance of assessments and assessment-based decisions made by teachers as they support and help define the learning that occurs in classrooms. Included in classroom assessments are formative assessments designed to give students and teachers feedback as well as performance assessments that can provide students with opportunities to demonstrate their learning to a larger audience. It is the use of many forms of assessment over time that causes Martin-Kneip (2009) to suggest that assessment should be thought of as a process not an event. Unlike “outsourced” accountability testing that usually occurs annually, classroom assessments are often created by teachers within schools, reflecting a combination of local standards as well as externally generated standards (Martin-Kneip, 2009; Wiliam, 2010a). So, for the purpose of this study, I would change the name of Popham’s second camp from “classroom assessment” to be “formative assessment processes,” which include: the processes used by teachers and undertaken by students to both evaluate and support learning as it occurs in classrooms as well as processes that demonstrate learning over time to a larger audience. Whatever the actual assessment format, scholars who study formative assessment processes would agree with LeMahieu and Eresh’s (1994) call for researchers and reformers to focus their attention on work that originates in classrooms, not on outsourced standardized tests:

The classroom is the crucible. It is the place where every idea about the best or most powerful form of education finds its expression.... In short, all the elements that define the educational experience—curriculum, instruction, environment, and assessment—are inextricably intertwined. (p. 126)

Scholars who study formative assessment processes do see that assessment must have a measurement component as well as a concern for reliability in order to allow for comparisons across students, classrooms, and schools. However, many scholars (Baird, 2013; Elwood 2013; Glaser & Silver, 1994; Kellaghan, 2006; Popham, 2009) also cite the work of Messick (1989), who suggests that the validity of assessments cannot be judged by numbers alone. Instead, assessments must be judged by their uses and by the impact they

have on the outcomes of teacher practice and student learning. Like accountability assessment, the development and use classroom assessment processes such as formative assessments and performance assessments to demonstrate learning have also become a global phenomenon. In alignment with the larger neoliberal agenda, supporters of formative assessment processes are equally concerned with economic competitiveness (Darling-Hammond, 2010; Wiliam, 2010a), albeit in a smaller number of nations where policymakers are focused less on centralized testing systems and more on decentralized professional development systems to improve student performance (Mourished et al., 2010).

Both those who champion the use of assessments for accountability and those who focus on using assessments for formative purposes recognize the need for reform in public education in that they would agree on the need to improve all three elements of Coburn and Turner's (2012) "meaningful outcomes: "(a) outcomes related to student learning; (b) those related to changes in teacher and administrative practice; and, (c) those related to organizational or systemic change" (p. 177). However, the two camps do so with two very different theories of action as to the role that assessments and teachers play in the reform process. These assumptions were described two decades ago:

One view [accountability assessment] seeks to induce change through extrinsic rewards and sanctions for both schools and students on the assumptions that the fundamental problem is a lack of will to change on the part of educators. The other view seeks to induce change by building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches. (Ancess & Darling-Hammond, 1996, p. 57)

Even though this description is 20 years old, it still accurately describes practices around the world. In a global study of 20 districts engaged in system-wide reform, McKinsey and Company found that the majority of systems in the study rely on close, centralized systems of control that included standardized curriculum and "outsourced"

testing, while a small number of systems are more decentralized and emphasize teacher practice and student outcomes centered around a common set of standards (Mourished et al., 2010). O'Day (2002) frames these two theories of action as follows: On the one hand, "bureaucratic accountability" akin to accountability assessment assumes that schools, teachers, and students are accountable to a state mandated system of regulation. Furthermore, "bureaucratic accountability ... hold[s] schools and school personnel accountable not for delivery designated educational inputs and processes but for producing specific levels or improvements in student learning outcomes" (p. 11). On the other hand, "professional accountability"—akin to formative assessment processes—has three parts: "First it is centered on the process of instruction.... Second, much of the focus of professional accountability concerns ensuring that educators acquire and apply the knowledge and skills needed for effective practice.... Third, professional accountability involves the norms of professional interchange" (p. 20). More recently, Weiss (2012) also framed two approaches or "directions" of data use in school reform when she said of data systems, "[Some] ... are intended principally to improve the performance of school staff, whereas other data systems are intended principally to hold schools and districts accountable for outcomes" (p. 3).

In sum, two views of assessment emerge. One view focuses on using standardized tests for accountability purposes to incentivize individual and organizational improvement through rewards and sanctions and the other uses formative assessment processes that originate in classrooms to provide, feedback, support learning, and promote continuous improvement. Vidovich (2009) describes the tension between these views when she talks about the need to address the "current imbalance between 'prove' and 'improve' forms of accountability" (p. 564). On the one hand, accountability assessment demands that schools and teachers "prove" through student performance on externally produced standardized assessments that they are either performing to a required standard or working to increase student performance toward that standard (or both) in measurable ways. On the other

hand, those who study formative assessment processes posit that teachers must learn to use assessment data gathered from a variety of assessments to “improve” their practice (Hattie, 2011; Kim, 2010; Supovitz, 2012; Wiliam, 2010a). These assessments would include regular classroom observations produced internally to help students, as well as summative performances of learning that not only demonstrate learning but also influence what skills are learned (Darling-Hammond, 2008; Martin-Kniep, 2009). Globally, more and more scholars are questioning the dominance of accountability assessment in terms of its efficacy, as well as its impact on students and schools (Brookhart, 2013; Dominguez, Vieira, & Vidal, 2012; Gebril & Brown, 2014; Howie, 2012; Klenowski, 2012; Scott, 2007; Smaill, 2012). As recently as the spring of 2018, the role of public accountability through testing is evolving. The New York State Legislature passed bill (a.10475) in the spring of 2018 planning “to eliminate the mandate that state created or administered tests be used to determine teacher or principal’s evaluation.” But there is still a plan going forward to collect growth scores for teachers, even though there is no clear plan for how those results will be used.

What has yet to emerge in the research is any understanding as to how the two approaches and their corresponding theories of action may influence each other. In fact, there has been a tendency for educators, policymakers, and researchers to take sides and claim that one approach to assessment is better than the other. In the case of accountability assessment, there has been a tendency to take an “all good/all bad” stance (Scheulrich, Skrla, & Johnson, 2000). But simply to take sides would miss the mark in terms of understanding how both theories of action approach the relationship between assessment, reform, and teacher practice and how the two theories of action may, in fact, co-exist in schools.

In the discussion that follows, I will describe how scholars have framed the issues surrounding the uses of assessment and assessment data, demonstrating first that current policy has shaped evidence-based practice, not as a slow march toward progress, but as a process where one theory of action—using assessment in the form of testing for

accountability —has been privileged; and second, that alternative approaches to assessment-driven reform, including the need for assessments created and used locally by teachers, have been de-emphasized by policymakers and by many districts when these alternatives could present significant opportunities for school reform that creates more meaningful outcomes for students. I will also show how current scholarship into evidence-based practice has demonstrated gaps in research as to our understanding of teachers' uses of evidence derived from assessments (Coburn & Turner, 2011; Spillane, 2012). In addition, these scholars point to several methodological and theoretical approaches that might better inform research into “everyday practice that makes up schooling.” In so doing, I will demonstrate the need for research into specific cases where teachers are actively engaged in the development and use of locally created assessments processes to inform their practice and improve outcomes for students within the context of high-stakes accountability in New York State.

Background: Into the Black Box

In what is now a well-known and widely cited article on assessment, Black and Wiliam (1998) introduce the metaphor of the “Black box” to assessment theory—a metaphor that captures the theory of action behind accountability assessment. “Certain *inputs* from the outside—pupils, teachers, other resources, management rules and requirements, parental anxieties, standards, tests with high stakes, and so on—are fed into the box. Some *outputs* are supposed to follow” (p. 10). Black and Wiliam suggest that the current accountability assessment discourse treats the classroom as a black box where reform can occur without ever knowing what happens in classrooms. With the right standards and assessments to monitor that they are met—“inputs,” reformers need only apply the right combination of incentives and sanctions in order to reform the system and improve outcomes—“outputs.” The black box also suggests that the performance indicators

found in assessments are all that are needed to ensure that learning has occurred. In the words of one recent study, by following the black box assumptions, those who support the accountability assessment discourse have been “seduced by elegant simplicity” (Baker et al., 2010, p. 20). Baker et al. are critics of the current assessment discourse and of New York’s system of assessment and teacher evaluation found in the statute NY 3012-C. They argue that the current assessment discourse and its subsequent laws have over-simplified the complex issues of assessment, accountability, and teacher evaluation. Furthermore, the current discourse ignores possible unintended consequences of accountability assessment’s “inputs.” “Adopting an invalid teacher evaluation system and tying it to reward and sanctions is likely to lead to inaccurate personnel decisions and to demoralize teachers, causing talented teachers to avoid high needs students and schools” (p. 4). And they conclude, “Legislatures should not mandate a test-based evaluation approach to teacher evaluation that is unproven and likely to harm not only teachers, but also the children they instruct” (p. 4). This conclusion is reinforced by a study conducted in England by Gilbourne and Youdell (2000) of the impact of decisions in school-based high-stakes testing. “Our study analysis shows the multiple and active ways in which teachers [who are evaluated based on test scores] are engaged in the rationing of education” (p. 206). By “rationing education,” Gilbourne and Youdell describe a process by which teachers and schools in England have actually decreased educational opportunities for minorities in response to the pressure placed on them to increase test scores.

Just as Black and Wiliam advise us to look inside the black box to see how assessments work in classrooms, other scholars have posited that assessment, like most aspects of teaching, is a social activity and must therefore be seen through a sociocultural lens (Gipps, 1999). Gipps suggests that power and control have historically been an important consideration in assessment systems. In the evolution of state-run assessment systems, she cites three stages of control: first for selection and certification, second for monitoring, and third to exert policy influence on curriculum and teaching. This

sociocultural lens encourages that we look not just at the intended purposes of assessment but also at its impact on teachers and students. In a similar way, the poststructuralist frame of critical theory can be a helpful tool in looking into the current assessment discourse, beyond the “common sense assumptions” (Apple, 2004) and the apparent “simplicity” (Baker et al., 2010, p. 20) of the current accountability assessment discourse. Like Gipps, critical theorists have also examined the role of assessment and accountability through the lens of power and empowerment. “Discourse systematically constructs versions of the social and natural worlds and positions subjects in relations of power” (Williams, 1992, as cited in Luke, 1996, p. 21). Luke explains that any text shapes our thinking and positions subjects in relation to power. Assessments as “texts,” therefore, create discourses that define “what counts” and what constitutes “success and failure” in schools. (Luke, 1996). Furthermore, teachers both willingly and unwillingly participate in the creation of assessment discourses that, in turn, define and regulate their identity as professionals. In this sense, teachers create assessments, and assessments create teachers.

When turning to the role of teachers’ use of assessments, poststructuralists ask who decides the categories and conditions by which accountability exists (Cruikshank, 1999). In particular, the work of Apple (2004) and Ball (2009) is strongly critical of current policies that create conditions where notions of transparency and accountability have been transformed into what Ball and others describe as an “audit society” (Ball, 2009, p. 678). It is in the context of an audit society that two visions of teachers emerge. In Ball’s vision, teachers are evaluated based on their ability to perform a set of predetermined tasks, specifically helping students perform on “outsourced” externally generated tests. Outsourced assessments exert control on curriculum and teaching—Gipps’s third stage of control. Teachers implement a predetermined curriculum and are deemed effective only by measurement of outcomes that are imposed from without. As Ball (2009) describes the theory of action behind accountability assessment,

The practitioner is left or held responsible for their performance, but not for the judgment as to whether that performance is "right" or "appropriate," but rather whether it meets audit criteria.... Within all this teachers have lost the possibility of claims to respect except in terms of performance. (p. 669)

Ball and others have termed this vision of teaching as “post-professional,” which is similar to what O’Day (2002) discusses in the frame of “bureaucratic accountability.” This is what Aneess and Darling-Hammond (1996) mean by “change through intrinsic rewards and sanctions” (p. 57). The vision represents a drastic departure from earlier reforms where school improvement was to be achieved by improving and empowering teachers, or what Aneess and Darling-Hammond mean by “change by building knowledge among school practitioners” (p. 57). Yet despite rhetoric such as “a high quality teacher in every classroom” (NCLB, 2001), current policy, at its heart, is based on the assumption that schools will improve—“tops” will be raced to and “gaps” will be closed—only through rigorous external accountability systems that force teachers to conform to a single externally derived standard.

In contrast, to the “post-professional” teacher, Ball (1999) puts forward an alternate vision for teachers as professionals akin to the theory of action behind formative assessment processes. Ball offers the term “authentic professional” to describe the role along with some specific attributes of teachers as professionals in an age of accountability. Authentic professionals are practitioners who organize their work based on their own judgments of what is right or best for students. Ball describes specific values and behaviors of teachers, such as the ability to define standards, to make judgments and “right” decisions about teaching and learning in the context of ambiguity and pluralism. These decisions often “involve issues of moral purpose, emotional investment, and political awareness” (p. 677). This is not to say that external judgment or accountability has no place in practice. Or as one scholar states, “It also seems naïve to believe that educators can successfully improve the current scene without leaving room for quantitative measures of academic learning that can be meaningfully aggregated, and for general purposes serve to guide

educational practice” (Sanger, 2012, p. 303). Assessment is the heart of good practice, where teachers use short-cycle formative assessment processes to adjust their practice and help students learn everyday (Lee, 2010; Supovitz 2012). They also use longer-cycle summative assessments and performance assessments, both to demonstrate student progress over time but also to shape that progress toward skills that are required to complete the task. State assessments can be used as one way to calibrate local assessments and to demonstrate first to the teachers, then to the community, that the teacher or the school is making the good decisions that support student learning. Again, on the global stage, a small number of school systems are described in Ball’s “authentic professional” terms where scholars describe systems that “move the locus of improvement from the center to the schools themselves; the focus is on introducing peer-based learning through school-based and system side interaction ... as well as innovation and experimentation” (Mourished et al., 2010, p. 26).

Vidovich (2009) puts it this way: “We need some new ways of thinking about productive rather than destructive accountability relationships,” when she posits the need to address the “current imbalance between ‘prove’ and ‘improve’ forms of accountability” (p. 564). Accountability exists in both a post-professional and authentic professional world, but it takes different forms. Where “post-professional” accountability is reduced to proving performance through acts of conformity and surveillance (Foucault, 1979), “authentic professional” accountability involves choice and moral purpose as well as collective judgment aimed at improving practice and improving student performance. While current policy uses the language of public accountability, Biesta (2009) argues that this is an illusion: “Underlying this ostensible aim of accountability to the public the real requirements are for accountability to regulators” (p. 655). In the end, many scholars are left concluding that the accountability assessment approach leaves teacher practice severely limited or constrained (Darling-Hammond, 1995; Hamilton, 2002; Linn, 2000; O’Day, 2002; Sanger, 2012; Wills & Stanholtz, 2009). Taken further, this conclusion suggests

that the impact of accountability has entered and perhaps altered the potential of the classroom as a “crucible” for reform. The challenge, then, for researchers is to find ways to observe and understand the impact of accountability assessment policies on teachers in classrooms. By what means can the contrasts and potential conflicts between accountability assessment and formative assessment processes be observed empirically and better understood? How could this understanding impact our understanding of assessment reform and classroom practice?

Teacher Agency: A Conceptual Frame

One way to describe and observe the overarching contrasts and potential conflict between the “post-professional” accountability assessment and the “authentic professional” classroom formative assessment is with the concept of “agency.” Many scholars offer definitions that are potentially relevant for this study (Campbell, 2012; Elmore, 2009; Popkewitz, 2008). Most recently, in a study of teacher agency in the context of urban public school reform, Birdwell-Mitchell (2015) cites Emirbayer and Mische (1998), who define agency as “the capacity of actors to make practical and normative judgments among alternative possible trajectories of action, in response to emerging demands, dilemmas, and ambiguities of presently evolving situations” (p. 141). This definition is significant because of its direct reference to the relationship between agency and “evolving situations” such as the current reform environment. Birdwell-Mitchell (2015) contends that during times of change, teacher agency can act to either support or resist attempts at institutional change.

For the purposes of this study, teacher agency can be defined as a teacher’s sense or belief that they have the ability and authority to make changes in their practice based on their own judgment. In the case of assessment driven reform, this would mean the sense that teachers have the ability and authority to use assessment data to make instructional decisions and change their curriculum.

Aligning the concept of teacher agency to the second theory of action described earlier, Sloan (2006) describes an “agency orientation” reminiscent of O’Day’s “professional accountability” in looking at reform efforts that assume a certain degree of teacher input and teacher judgment in the process of school change. In several descriptions of classroom formative assessment, the concept of teacher agency as “agency orientation” can be seen. With its focus on learning, not on standardized test performance, formative assessment processes shift the use of assessment away from performance indicators to actual performances, from measurement toward “shaping instruction as it unfolds, gauging student achievement, and evaluating curriculum” (Young & Kim, 2010, p. 5). In this sense, the actions that teachers know how to take and feel that they have the authority to take are essential.

Birdwell-Mitchell (2015) also argues that, for researchers, the concept of agency has an important role to play in helping us observe and understand the relationship among assessment, accountability, and reform. He suggests that an understanding of the conditions that create teacher agency can in turn provide a “generalizable framework for the activities that help teachers drive outcomes in their schools” (p. 140). As such, teacher agency is a useful lens through which researchers can describe and observe the conditions in classrooms that can either inhibit or create meaningful outcomes for teachers and students.

With some grounding into the ways scholars have framed the uses of assessment and assessment data through the lens of agency, we have revealed that the concept of agency may be the common phenomenon through which these competing theories of action can be explored. Furthermore, we have seen the importance of looking into the black box, of seeing that current policy ignores the complexity of the relationship and situated nature of assessment, reform, and teacher agency within classrooms. This suggests that current policy has shaped evidence-based practice, not as a slow march toward progress, but as a process where the theory of action found in accountability assessment has dominated policy and practice. These conditions, in turn, may actually inhibit alternative approaches to

reform, including the possibility of assessments created and used locally by teachers and school districts when such assessments could present significant opportunities for reform.

What we have not examined is how one might approach these contested issues of accountability, practice, and agency empirically. Current scholarship into evidence-based practice has demonstrated gaps in research as to how teachers use evidence derived from assessments, be they outsourced standardized tests or locally developed formative assessment processes. Focusing not on the macro level of policy but rather on the micro level of practice, they, too, uncover the complex and situated nature of the “everyday practice that makes up schooling.” It is in that scholarship that a theoretical and methodological frame around accountability, practice, and agency emerges.

Sense Making Theory: A Theoretical and Methodological Frame

Both accountability assessment and formative assessment processes assume that assessments produce data and that practitioners in turn make sense of—examine, interpret, and use—assessment data to create outcomes—changes in practice, improvement in learning (Coburn, 2012). A great deal of the literature on assessment in school reform is described as normative and not analytic, where assessment data use is accepted as “an article of faith” (Coburn, 2006), when, in fact, both Spillane (2012) and Little (2012) point out that data use by teachers and school leaders is both under-researched and under-conceptualized. Moss’s (2012) question illustrates the need to better understand the policy-practice connection: “How do policies, social structures, artifacts, and attendant classifications that span multiple organizational contexts get taken up in local practice, and how, in turn are they generated from local practice?” (p. 223). Moss suggests a way to answer her question when she introduces Spillane’s ideas in her introductory essay to a collection of data use articles. “As Spillane (2012) states, one can design for practice, but one cannot design practice. People will make sense of formalisms in their own situated way”

(p. 224). It is the act of “making sense” in the situated context of organizations that lies at the heart of sense making as a theory, as a conceptual frame, and as a research methodology. Coburn and Turner (2011) justify the use of sense making theory because it acknowledges the multiple steps between assessment data and practice—steps that can be influenced by beliefs, prior experiences, social interactions, and institutional pressures.

Little (2012) frames the problem of research into assessment data use as the need to connect the “macro” world of policy with the “micro” level of practice. Spillane (2012) explains that data do not spontaneously make or even guide decisions on their own, people do (p. 114). Furthermore, Spillane describes data use as “ostensive” and “performative,” suggesting that what people say they do with assessment data is potentially different from what they actually do with data. In sum, there is a need for theory to help better conceptualize data and assessment use by teachers.

Central to this dissertation is the idea that sense making is a social activity (Weick, 1995). Coburn (2006) emphasizes that sense making is both collective and situated.

[Sense making] is collective in that it is shaped by interaction, signaling, and negotiation.... It is situated in sense-makers embedded contexts. Individuals and groups draw on ideas or approaches available to them in their proximal communities as they make sense of their situation.... In this respect, sense making theorists recognize the ways in which existing social structures and cultures shape interpretation. (p. 345)

Individuals in organizations and organizations themselves undertake acts of sense making (Weick, 1995). In particular, Weick describes the concept of “generic subjectivity” as a kind of collective sense making undertaken by organizations. An example of generic subjectivities in the context of assessment use is when a group of teachers works with a common rubric to create consistencies in expectations and practice. But Coburn (2006) takes the collective aspect of sense making farther than Weick when she describes how teachers connect policy and practice in organizations.

There is a growing body of evidence that these sense making processes play a central role in how people in schools implement policies. The ways in

which teachers enact policies are shaped by what they understand the meanings and implications of these policies to be. Teachers construct this understanding by drawing on their preexisting beliefs and practices in a process that is shaped by patterns of interaction with colleagues and the social and structural conditions of their workplace. (p. 346)

Sense making, as a theory and as a conceptual frame, is well suited to investigate the “everyday practice that makes up schooling.” Coburn’s (2006) quotation suggests that sense making theory may help researchers shed light on the complexity of the relationship and situated nature of assessment, reform, and teacher agency within classrooms. If accountability assessment has dominated policy and practice with the result of creating conditions that limit the agency of teachers in the reform process, this would impact how teachers make sense of assessments and assessment practices in their classrooms. These same conditions, in turn, may actually inhibit alternative approaches to reform, including the possibility that teachers and school districts can create and use assessments locally, when these approaches to assessment could present significant opportunities for reform.

Problem

In the field of data use research, the majority of studies have been done in contexts where the data teachers are expected to use have already been collected. (Wallenstein & Hatch, 2014)—primarily data derived from standardized tests or what Smith et al. (2014) have termed “outsourced assessments.” What’s more, these studies demonstrate that teachers do not seem to use data very effectively to make instructional decisions for two reasons. First, the data used are given to them from outside their classrooms. Second, new accountability pressures may be creating conditions that limit how or to what degree teachers change their practice. Only a small number of preliminary studies have looked at teacher-created assessments in the context of high-stakes accountability. This research project builds on the work of Wallenstein and Hatch (2014), who have looked at the teacher’s role in creating assessments as a factor in how data derived from those

assessments impact instructional decision making. This study examines further the potential impact of locally created and implemented assessment practices on data use by examining the influences on teacher practice of assessment data derived from formative assessment processes that are: (1) locally produced and strive to align with local and state learning standards; (2) produced in facilitated and collaborative efforts where teacher teams work together to inform their practice; and (3) created within the context of state and local accountability structures. By doing so, this study looks more deeply into the second of the two theories of action described in this chapter, namely, the “professional accountability” model, where reform is induced by “building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches” (Ancess & Darling-Hammond, 1996, p. 57). In doing so, this study attempts to understand two problems not currently addressed in the literature. First, it is not clear how data derived from locally created performance assessments actually influence teacher practice. While the second theory of action claims that knowledge can be built and changes in practice can be induced through local, collaborative assessment practices, the theory has not been well explored empirically. Second, while formative assessments, performance tasks, and performance assessments have been developed and researched for two decades, only recently is this work being attempted within the context of high-stakes accountability. Therefore, the impact of the current policy context on teachers’ ability and tendency to use data derived from locally created formative assessment processes is not well understood.

Purpose and Research Questions

The purpose of this research project is to examine: (1) how teachers use the data from locally-created assessments to inform their instruction, change their curriculum,

and/or change their assessment process; (2) teachers' sense of their ability to change and modify their instruction/assessment practice; and (3) the factors that limit/support their use of assessment data to change their practice.

Research Questions

I am conducting this qualitative case study in order to understand:

- How does creating and using local performance assessments influence teacher practice?
- How do teachers who participate in the work of creating local performance assessments both as individuals and as members of a group use the data produced from these assessments to inform their practice or change their curriculum?
- How does this process influence their understanding of the role(s) that assessments play in their work? What changes do they make in their uses of assessment and in their curriculum?
- What factors support or limit teachers' participation in the creation of local assessments?
- What factors support or limit the changes teachers make in their instructional practices or their curriculum?

Rationale

Since the processes behind the second theory of action are not well understood, there is a need to study places where that theory is being put into practice. The Performance Assessment Development Initiative (PADI) provides one such opportunity. PADI was created by Giselle Martin-Kniep in conjunction with teachers and administrators from 14

districts in the greater New York area. Approaching reform with an “agency orientation” (Sloan, 2006), Martin-Kniep (2009) suggests that assessment should be seen not as an event but as a process. To that end, teachers involved in PADI are trained in developing and implementing performance assessments that align to existing state and local standards. The work calls for a balance between traditional tests and performance tasks that serve both formative and summative purposes. Traditional instruction and assessment used in concert with performance tasks help students learn new content, then make meaning of their work, while also giving feedback to teachers, who can support the learning (Martin-Kniep, 2009). Over time, the goal of PADI is for teachers not only to reform their own practice, but for them to influence the practice of colleagues in their districts.

Since schools that have been working with PADI are looking for alternatives to the current accountability testing system in New York to evaluate and improve student outcomes and also to improve the practice of their teachers, these PADI schools provide a rich opportunity to observe the impact of local assessment development in an effort to understand the conditions where the impact can lead to “meaningful outcomes.”

Significance

Twenty years ago, Haney (1991) warned us that policy might emphasize the wrong types of assessments. It seems now, a generation later, that similar concerns are being raised about the implementation of the Common Core Learning Standards and the new wave of outsourced high-stakes tests aligned to those standards. In a recent publication that both defends the CCLS as well as raises concerns about how they are being implemented and evaluated, three education scholars raise similar concerns (Smith et al., 2015). They refer to the “outsourcing” of assessment in and around the common core as one key problem. They credit Paul Zavitovsky (2012) with the term and quote him directly as follows:

Given what we know about the culture of American teaching and the power of high quality classroom assessment, the troubling thing about current work on Common Core assessment is that we seem to be doubling down again on outsourcing, this time with tests that are being developed by teachers by the PARCC multi-state consortia. (p. 8)

At a time when the efficacy of “outsourced” or externally produced and implemented assessments is in question, the need to understand the potential of local assessments is significant. This qualitative case study that examines the impact of locally created and implemented performance-based assessments developed in classrooms by the teachers who created them has potential significance on three levels:

First, it has the potential to inform practitioners: To give subjects the opportunity to reflect on their sense making or how they are making sense of their work is one way this study may benefit the teachers and school districts being studied. One way that I want to judge the validity of this work is whether or not the teachers in the study see benefit in participating in the study. It is my belief that the act of reflection on practice can be somewhat transformational (Cho & Trent, 2006; Marshall & Rossman, 2010) for practitioners. Teachers will gain insight into their role in the processes that are happening to them and around them (Transactional validity), gain “voice” in a process, and gain some sense of the technical language of their work. The district(s) being studied will gain understanding of the complexities of implementing a vision of “assessment culture” and promote improvement—and potentially gain public evidence of this district work as an alternative to the accountability assessment of state policy.

Second, the study has the potential to inform policymakers who claim to value a combination of classroom performance and standardized testing. For policymakers, this study has the potential to improve our understanding of the efficacy, impact, and unintended consequences of current accountability policy.

Finally, it is significant to the scholarship on data use because it builds on existing studies by looking at a kind of assessment that has yet to be studied. The majority of data use research has looked at data created outside the classroom—specifically, the

“outsourced data” from standardized testing. Few studies have looked at the quality of “localness”—the production and use of assessment data locally as a factor in the data use process.

Chapter II

LITERATURE REVIEW

Introduction

In the field of education, the topic of assessment is vast and has been written about extensively in both the research literature and in the popular press, especially in the last 25 years (Koretz, 2009; Linn, 2000; O'Day, 1993, 2002; Ravitch, 2011; Rothstein et al., 2008; Black and Wiliam, 1998; Wiliam, 2010a). One challenge in confronting a review of literature on the topic of assessment has to do with questions of scale and scope. Even the two sub-topics of accountability assessment and formative assessment processes that often occur in classrooms outlined in the first chapter are vast fields that have been written about through lenses of research, reviews of research, theory, and education history. The goal of this literature review into the intersection of assessment, accountability, teacher agency, and the reform of practice is to follow the advice of Boote and Beile (2005), who suggest some criteria for an effective review of literature. "It sets the broad context of the study, clearly demarcates what is and what is not within the scope of the investigation, and justifies those decisions" (p. 4). So while the broad context of this study is assessment and assessment policy in the last two decades, this review will focus on teacher agency and the reform of practice within the context of the evolving assessment climate. It will demonstrate the need for research into how teachers use data generated from local assessments to inform and improve their practice.

Methodology

For the broadest contextual framing, my literature search began by looking at several literature reviews that set out to survey the assessment field during the time period 1990 to the present. These reviews include work by Wolf, Bixby, Glenn, and Gardner (1991), Glaser and Silver (1994), Hamilton (2003), Darling-Hammond (1995, 2008), Black and Wiliam (1998), Klenowski (1995), Young and Kim (2010), and Bennett (2011). I conducted several citation searches based on initial texts and reviews. This process helped me develop and conduct key-word searches in several databases, including ERIC, EBSCO, Education Full Text (Wilson), and Proquest. This search included a specific search for dissertations. Key search terms included *assessment, evaluation, testing (formative, authentic, performance, classroom), accountability, agency, and teacher agency*. Using these terms alone yielded results in the tens of thousands, so I added additional limiting terms in a Boolean search to focus my findings, with a particular focus on assessment as it pertains to agency and teacher agency. This process helped me limit my focus to studies that looked at assessment and that focused on school-aged students and teachers of school-aged students. Finally, using the same key search terms, I searched several American and international online journals directly, including *Review of Research in Education, Curriculum Inquiry, American Journal of Education, and Measurement in Education*. I found this approach more efficient at times, especially since many articles refer to teaching conditions and teacher attitudes related to the concept of teacher agency without using the specific term “agency.” Out of approximately 40 books, 85 articles, and close to 60 dissertations found through these methods, I narrowed the field by focusing on work done during or after 1990 for the reasons described above.

The one specific addition to this process to help identify methodological considerations came in my approach to research on “data use” and its connection to assessment policy. My familiarity with some of the authors (Coburn, Turner, Spillane,

Supovich, Warren-Little) in this field through course readings on school change and policy implementation research helped me identify two special publications devoted to the topic of data use. They are *American Journal of Education* 118 and *Teachers College Record* 114, both published in 2012. These publications, which often consisted of several literature reviews joined by a common theme, yielded a second round of citation searches and key-word searches in and around the following terms: *data use, assessment policy and practice, sense making, policy implementation*, as well as *micro-process and macro-process in schools*.

From this point, I organized the research into four broad categories—teacher agency, data use, sense making, as well as accountability assessment and formative assessment processes—that reflect the four moves mentioned in the introduction. A few of the studies fit into more than one category; in those cases, the research is described in detail once and then referred to more briefly in later categories.

Agency

Birdwell-Mitchell (2015) conducted qualitative research in an urban public school undergoing state-mandated reform. The result of the study was to formulate a theory about the conditions under which teacher agency promotes reform and under which conditions teacher agency can act as a barrier to reform. The Birdwell-Mitchell study contributes significantly to this study's use of teacher agency as a lens through which we can describe and understand the relationship among assessment, accountability, and reform. Such a lens not only narrows the focus of this review to assessment research that focuses on the relationships between agency and assessment, but also sheds light on some of the limitations of that current research. In so doing, this section will argue that there is a need for more studies like Birdwell-Mitchell's where the phenomenon of agency is used as a means of understanding something else, such as the conditions under which policy implementation or the reform of practice actually occurs in classrooms.

One of the earlier studies that looked at the conditions surrounding assessment use and reform was conducted by Sloan (2006) in an article titled "Teacher Identity and Agency in School Worlds: Beyond the All-Good/All-Bad Discourse on Accountability Explicit Curriculum Policies." Sloan posits that researchers with an "agency orientation" tend to focus "on curricular and pedagogical practices related to the creation of socially productive and inclusive classrooms" (p. 121). She develops the agency stance as part of a larger point about the highly polarized nature of public discussions around accountability, which, according to Scheulrich et al. (2000, cited in Sloan, 2006), have "devolved into a dichotomy in which accountability is either 'all-good' or 'all-bad'" (p. 119). Sloan (2006) argues that the "all-good" side tends to focus on improvement of the quality and equity of instruction but ignores the increase of test-focused or "test-explicit" instruction. On the other hand, the "all-bad" side tends to focus on the fact that accountability has "forced teachers to teach to the test and has prevented them from delivering higher-quality, more child-focused instruction" (p. 120). At the same time, the "all-bad" discourse tends to ignore examples where accountability has led to high-quality and more equitable instruction.

It is in the complication of the "all-good/all bad" dichotomy that Sloan (2006) makes two important contributions to the theory behind research into assessment policy and practice. First, she suggests that one reason for the dichotomy is based less on local variation between districts and more on a "paradigmatic orientation" among educators and researchers. Similar to the two contrasting theories of action seen with accountability assessment and formative assessment processes, Sloan suggests that some scholars tend to view school reform and assessment policy through a "systems-world" paradigm that looks at the administrative and managerial questions around school reform and assessment, or a "life world" paradigm that looks at school reform and assessment through the lived experiences of teachers and students. It is within the second "life world" paradigm that she frames the idea of the "agency orientation."

Second, Sloan (2006) begins to complicate the notion of teacher agency in both paradigms. She posits, “Both the public discourse and education literature present teachers as mostly passive agents whose teaching behaviors are leveraged (negatively or positively) in seemingly predetermined ways” (p. 121). To this seemingly oversimplified concept of teacher agency, she adds, “Researchers are exploring and reporting the complexities and uncertainties involved in trying to understand the ways individual teachers experience and respond to test based systems of accountability” (p. 121). Pushing this point further, she goes on to suggest that “current understanding of teacher agency vis-à-vis accountability related curriculum policies as merely a capacity to resist and act otherwise—as does much of the literature critical of accountability—obfuscates important issues of teacher quality and equitability” (p. 123). In these two moves, Sloan opens the door for a more complex understanding of assessment practices as well as the role teacher agency can play in observing and understanding the impact of assessment policies on the practices of teachers in classrooms. She also raises a question about the nature of teacher agency itself, suggesting that a simple mechanistic view of agency is inadequate. It is this complication of the notion of teacher agency that other scholars take up as they explore teacher agency within the context of assessment and accountability.

While the term “agency” is not always used specifically, the idea of teacher agency and teachers as agents, particularly as agents of change, can be found throughout works on the sociology of teaching and professional practice (Lortie, 1975; Schön, 1984), on the moral nature of teaching (Buzzelli & Johnson, 2002; Campbell, 2012; Sanger, 2012), as well as school change (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Fullan, 2007; Hargreaves & Shirley, 2009; Wiliam, 2010). Many scholars offer definitions that are potentially relevant for this study and its focus on teacher practice. Popkewitz (2008) describes agency in contemporary life as “the individual as purposeful actor who produces change through intentional actions” (p. 18). City, Elmore, Fiarman, and Teitel (2009) define human agency as “the capacity to exert control over the terms and conditions of one’s

learning” (p. 189). The authors are concerned not only with student learning but also with teacher learning as an essential mechanism of school reform. Elizabeth Campbell (2012) offers this definition of agency: “The state of agency enables individuals (and, to some, collectives) to make free or independent choices, to engage in autonomous actions, and to exercise judgment in the interests of others and oneself” (p. 183). She also ties the notion of agency to a teacher’s role in education policy: “The capacity of teachers to use professional discretion in their pedagogical and curricular practices exists, not always easily, alongside their accountability to the state” (p. 183). Emirbayer and Mische (1998, in Birdwell-Michelle, 2015) define agency as “the capacity of actors to make practical and normative judgments among alternative possible trajectories of action, in response to emerging demands, dilemmas, and ambiguities of presently evolving situations” (p. 141). This definition is significant because of its direct reference to the relationship between agency and “evolving situations” such as the current reform environment. Birdwell-Mitchell (2015) contends that during times of change, teacher agency can act either to support or resist attempts at institutional change.

Specific to school reform, and to current accountability policy, teacher agency is often discussed as something that has been lost or severely limited (Ball, 2009; Biesta, 2009, 2010). In a recent publication of *Curriculum Inquiry* (2012), a series of research papers take on the notion of teacher agency directly as it pertains to accountability, testing, and school reform. These papers are helpful to this review because they provide important definitions of the term “agency,” they review recent research, and they present new research and research methodologies into teacher agency in schools. In doing so, they build on Sloan’s work and help construct a theoretical and methodological lens through which to examine other research into accountability assessment and formative assessment processes use by teachers at the level of practice.

Elizabeth Campbell (2012) offers this definition of agency: “The state of agency enables individuals (and, to some, collectives) to make free or independent choices, to

engage in autonomous actions, and to exercise judgment in the interests of others and oneself" (p. 183). She also ties the notion of agency to a teacher's role in education policy: "The capacity of teachers to use professional discretion in their pedagogical and curricular practices exists, not always easily, alongside their accountability to the state" (p. 183). From a moral perspective, Campbell cites her own work when she suggests a "double pronged" nature of agency where teachers have an obligation to conduct their work to exacting standards as well as serve to help students, through their actions and their example, learn to embody those same standards. In this double-pronged approach, Campbell acknowledges an important role of teachers with regard to agency in supporting the development of the agency of their students.

Priestley, Edwards, Priestley, and Miller (2012) take up the issue of agency as it relates to reform by pointing out that a common phrase in the current reform discourse is to describe teachers as "agents of change." The authors see three problems with what they call a "renewed emphasis" (p. 192) on the notion of teacher agency in conjunction with school reform. First, they question if agency always results in positive actions. Just as agency may play a role in the successful implementation of reform, it could act equally to prevent reform. Second, they question "to what extent teachers can achieve agency. There is arguably a low capacity for agency in terms of curriculum development within modern educational systems" (p. 192). Third, they point out that there is no clear understanding of how teacher agency actually works in school reform—or how assessment policy impacts teacher agency in the context of reform. "Teacher agency is often conceived as a slogan to support school based reform, despite attempts by researchers to locate it in relation to wider theoretical discussions of agency" (p. 193). As a result of these three points, Priestley et al. make two significant contributions, one theoretical and the other methodological.

First, in terms of bringing theories of agency in general to the specific question of teacher agency in reform, Priestley et al. (2012) draw on the work of several theorists to frame a more complex definition of teacher agency. Similar to Little's (2010) observation

about data use, that researchers must work in the intersection of the macro world of policy and the micro world of practice, Priestley et al. point out several theorists who claim that there is a connection between the agency of the individual and their environment, including Popkewitz (2008), Fuchs (2001), and Archer (2000). In particular, they describe the ecological view of agency described by Biesta and Tedder (2006):

The two key ideas that we have put forward for consideration are (1) the suggestion that agency should be understood in an ecological way, i.e. strongly connected to context, and (2) the idea that agency should not be seen as a capacity or possession of the individual, but as something that is achieved in particular (transactional) situations. (p. 27)

Biesta and Tedder also suggest implications for research:

All this implies that agency is not something that people *have*. It is, as we suggest, something that people can *achieve*, and they can only achieve it in transaction with a particular situation. This allows for the empirical possibility that in some cases the achievement of agency requires more effort from the individual than in other cases, something which is connected to the availability of resources. (p. 19)

Agency can be impacted by many factors and conditions in schools, suggesting that more research is needed to understand which factors and under which conditions agency can be achieved in the reform process.

Biesta and Tedder (2006) describe the situated nature of agency by suggesting that agency is achieved not just in certain places, but a “dynamic interplay” among three elements—“routine, purpose, and judgment” (p. 9). These three elements are described as the Cordial Triad, which Biesta and Tedder ascribe to the work of Emirbayer and Mische. The triad also suggests that agency has a temporal quality where “routines” focus on a person's sense of the past—what has always been done; where “purpose” focuses on the future of what could be or should be done; and where “judgment” focuses on the present—what is actually being done. Not only does the Cordial Triad further develop what agency is, but also suggests where and how researchers might observe it in the field. Like Sloan, Biesta and Tedder suggest that agency may be difficult to observe and easier to infer based on

observed actions and on a subject's descriptions of actions taken in the past, present, and future.

Priestley et al. (2012) make a second important contribution to the field in terms of research methodology when they ask questions about what aspects of agency are actually observable through research. They describe their study of school teachers in Scotland this way: "While the contexts did not always involve changes to policy and practice, they relate directly to how teachers make sense of externally initiated policy, and the multifarious factors that influence this process" (p. 194). Here, Priestley et al. suggest that agency occurs within a policy context through a process of sense making. They continue, "This analysis, therefore allows us to make a number of inferences about teachers' capacity to act as agents of innovation and change. The differing approaches to enactment allow us to investigate teacher agency as a response to or a reaction against educational policy" (p. 194).

Combined, these statements suggest that agency may not be directly observable, only inferred. What can be observed are situations—factors and conditions—where teachers are making sense of policy and other factors impacting their work. From these situations, researchers can draw inferences about teachers' sense of agency or capacity to act. The collective summation of these definitions and conditions can be seen in this study's working definition of teacher agency—a teacher's sense or belief that they have the ability and the authority to make changes in their practice based on their own judgment. In the case of assessment-driven reform, this would mean the sense that teachers have the ability and authority to use assessment data to make instructional decisions and change their curriculum.

In their study, Priestley et al. (2012) gathered data both from direct classroom observations and interviews. This approach is similar to the approach recommended by data use scholars, such as Spillane and Little, who are also critical of studies that rely entirely on "ex situ" (Little, 2009) interviews and surveys. From the data that Priestley et al. (2012) gathered from multiple sources over the course of one year, two case studies were

created and analyzed through the lens of ecological agency. Priestley et al. found that the teachers they observed achieved agency in their work unevenly and that this unevenness was not completely attributable to the single factor of external policy demands. They summarized their conclusions as follows:

First, the analysis does not explain why one teacher with rich prior experience and strongly held views about education was able to translate this strongly into her teaching, whereas the other was less successful.... Second, we conclude that the success or otherwise of externally initiated education change ... is highly problematic given the ecological factors that potentially impact on such translations.... Third, a consequence of this is that educational policy, especially when it requires changes to the social practices of teaching, needs to be designed to be more flexible, taking more account of teacher agency, and especially teachers' proactive and projective engagement with the policy in question. (pp. 210-211)

These findings reinforce Biesta and Tedder's (2006) observations about the ecological nature of agency. They also support Sloan's concept of "agency orientation." Researchers concerned with agency seem to hold three ideas as significant to our understanding of accountability, assessment, and reform. First, they have an agency orientation in that they want to understand the processes of assessment and reform at the micro level of practice in classrooms. Second, they understand that agency is not a simple mechanical quality that an individual has. Instead, teachers' actions are impacted by how they make sense of the relationships that exist between policy and the many factors and conditions that impact their classrooms and schools. Third, researchers assume that they must observe teacher practice in order to draw inferences about the degree to which teachers achieve agency in specific situations. All three of these ideas about agency can be seen in the recent research conducted in the field.

Similar to Priestly et al., Birdwell-Mitchell (2015) describes a methodology that includes both non-participant observation and interviews in one urban school over a period of one year. Birdwell-Mitchell defines three broad mechanisms that he describes as a generalizable theory for further research into the relationship between teacher agency and

reform. They are “peer learning, patterns of social interaction, and shared understandings aims and practices” (p. 148). This study is significant because, like Priestley et al. (2012), it points out the very real potential to oversimplify the concept of teacher agency as the key ingredient to reform.

At this point, it is important to point out that there are some potential philosophical and practical limitations with regard to the concept of “agency” as way to describe and observe teacher practice. Understanding these limitations acknowledges that it is possible to overstate the significance of any one theory of action over the other. It is also possible to overstate the potential of agency as it relates to school reform. First, at the more philosophical level, Popkewitz’s (2008) critique of school reform through the “double gestures of cosmopolitanism” reveals that both assessment camps are connected. While accountability assessment may be driven by fear—the fear that certain teachers and students present a danger to the goal of cosmopolitanism, and formative assessment processes may be driven by hope, hope in the potential of all people to learn—both camps attempt to reconstruct “the school, sometimes the society, but always the teacher and the student” (p. 4). Popkewitz’s critique suggests that if the goal of both camps is to shape or control students by shaping and controlling teachers, neither camp can claim to be an absolute source of or a barrier to agency. A second philosophical critique of human agency comes from Deleuze (1992). In the case of Deleuze, the discourse of accountability and its emphasis on the examination has connections to the “discipline society,” and yet the discourse of classroom formative assessment is not without fault. With its focus on standards and “best practices” working toward the specific goal of improved student outcomes, classroom formative assessment is still very much part of a “culture of control.” Deleuze talks of “molds” versus “modulation” (p. 4). Where the examination, as a technology of discipline, encloses or molds the teacher and the student, the perpetual assessment in the classroom is a modulation—a technology that controls teachers and students in ever-changing forms. Again, agency is not achieved. Again, Like Popkewitz, Deleuze would

question if, in the purest philosophical sense of the term *agency*, any reform can claim to foster it.

At a more practical level, Birdwell-Mitchell (2015) offers what could be described as a third critique of the concept of teacher agency as a driver of reform. He posits that teacher agency can go in two directions and therefore should not be seen as a driver of reform. Because teacher agency is influenced by social and environmental contexts, “teacher agency can both change and maintain institutional instructional practices in schools” (p. 140). Despite these potential limitations, Birdwell-Mitchell also argues that, for researchers, the concept of agency still has an important role to play in helping us observe and understand the relationship among assessment, accountability and reform. He suggests that an understanding of the conditions that create teacher agency can, in turn, provide a “generalizable framework for the activities that help teachers drive outcomes in their schools” (p. 140). So, even if agency, as Popkowitz and Delueuze describe it, is never truly achievable, the phenomenon of teacher agency as Birdwell-Mitchell describes it can be observed empirically—both the beliefs teachers have in their ability and authority to act as well as the actions they take as a result of creating and using assessments in their classrooms. As such, teacher agency can be a useful lens through which researchers can describe and observe the conditions in classrooms that can either inhibit or create meaningful outcomes for teachers and students.

Other studies, including 22 recent dissertations into the nature of teacher agency, can be placed in three broad categories. The first category is studies that look at agency specifically in the context of reform and high-stakes testing. Of the 7 dissertations, 5 were qualitative studies based solely on interviews and surveys. As such, their findings tended to be one-dimensional along the researchers' line of questioning. For example, Good (2011) found that teachers exhibited agency in the form of “push back” against policy when they felt they had a voice to do so. G. Mason (2013) found that evaluation and pay procedures did not impact teachers' ability to raise student outcomes. Gantt (2012) connected lack of a

sense of agency with “a) increased stress, b) increased frustration, c) growing distrust ... and d) loss of respect. For their professional judgment” (p. 1). J. Mason (2010) and Benson (2011) both found that external accountability played a negative role in collaboration within learning communities. By Sloan’s description, these 5 studies are indicative of research that views agency as a mechanistic quality.

The second category of dissertations contains studies that seek to find agency in specific contexts and settings, including teacher education. Sonu (2012), for example, conducted a year-long study of teachers' desire to teach for and about social justice. She found that, on the surface, teachers felt a loss of agency in their daily practice when faced with accountability and testing demands. At the same time, she found that teachers and students did achieve a sense of agency in what she described as “the hidden corners of the institution.”

The third category includes studies of attempts to produce agency in others. In most cases, the “others” are students, but in the first study, the others are teacher researchers. In all cases, these studies explore more deeply the nature and role of agency in teaching. The significance of this final set of studies is that they have the strongest sense of Sloan’s notion of an agency orientation. Unlike the first two sets of studies, which try to determine to what extent agency exists, or why it is or is not achieved in certain settings, these studies all take the stance in and around the importance of the phenomenon of agency in understanding the change process in schools.

Reviewing and synthesizing the recent research into teacher agency can draw several broad conclusions that inform future research. First, methods and approaches matter. While almost all of the research reviewed incorporates case study methods, those cases that tended to look at one specific factor tended to have narrower findings that also tended to describe agency in the fairly mechanistic terms of who has agency and who does not. Second, studies that relied solely on interviews and surveys and did not incorporate some direct observation also tended to have limited findings. Third, in all cases but a few, the

phenomenon of agency was the focus of the research. The impact of the presence or absence of agency was secondary. Therefore, with the exception of Birdwell-Mitchell's (2015) study conducted in one urban school, a gap in research seems to remain where the concept of agency is used to help researchers understand the conditions that influence teacher decision making and teacher practice. There is a need for more studies like Birdwell-Mitchell's, where the phenomenon of agency is used as a means of understanding something else, such as the conditions surrounding teachers as they attempt to implement policy or reform practice.

Data Use Research and Methodologies

A second area of research that also provides theoretical framing and methodological considerations for this study is work in the field of data use. This line of research is significant to this study because recent scholarship suggests that when, how, and even if teachers use data from assessments to inform their practice is complex and under-conceptualized (Coburn & Turner, 2011; Little, 2012; Spillane, 2002, 2012). Furthermore, this complexity includes the three factors examined in this study—formative assessment processes that are (1) locally produced and strive to align with local and state learning standards; (2) produced in facilitated and collaborative efforts where teacher teams work together to inform their practice; and (3) created within the context of state and local accountability structures.

At first glance, one could easily overlook the concept of data use as a mere subset of assessment theory and practice. Just as the phrase “agents of change” shows how the concept of agency can be oversimplified, the concept of being “data driven” has dominated policy language without being well researched. It seems like common sense that educators use data derived from assessments to improve schools. This line of thinking is “rooted in the conviction that if the right data are collected and analyzed, they will provide answers to key

educational questions and inform actors' decisions, and better educational outcomes will follow" (Coburn & Turner, 2012, p. 2). But this approach to data use is similar to the "black box" approach to assessment in that both assume action and results without looking more closely at the processes that may or may not be occurring around teachers within classrooms. Coburn and Turner point out that upon closer inspection of current research into data use, "few studies, or even combinations of studies, help understand how, why, and with what consequences data use interventions make a difference in schools" (p. 2). Researchers into data use have helped bring attention to this unexamined and under-conceptualized aspect of assessment practice in two recent publications, *Teachers College Record* and the *American Journal of Education*, both of which explore "the relationship between data use intervention and change in teacher practice and student learning" (p. 2). Both publications are significant for this study for two reasons. First, they call attention to assessment's role in reform at the level of practice. Like the literature on teacher agency, data use scholarship looks at the actions of teachers situated in the context of the social interactions within schools, local norms, and external policy demands. Put another way, current data use scholarship represents "a compelling case for what can be seen when data use is studied through lenses that link the micro and macro and point to how these learnings might be used to understand and improve educational practice" (Moss, 2012, p. 223). Second, this line of research offers methodological insights as to how the phenomenon of data use can be observed empirically in order to shed light into its role in understanding the relationship among assessment, accountability and reform. In particular, this review will focus on several studies that use sense making theory to understand the situated nature of data use and its influence on practice.

Data Use as a Situated Phenomenon

Taken together, the work of six authors found in *Teachers College Record* does a great deal to point out theoretical and conceptual gaps in data use research, frame important

definitions to help close those gaps, and show how local assessments can fit into the larger field of data use research. As early as 2006, Cynthia Coburn called for research that focused on data use in practice, including the mechanisms of practice, the activities, and the group dynamics involved. By 2012, Coburn continued to make the same case along with co-researcher Erica Turner. Commenting on the situated nature of data use, the two suggest, “There has been an uneven and a theoretical approach to studying the relationship between data use interventions and the contexts in which they unfold” (Coburn & Turner, 2012, p. 2). They continue, “Absent this evidence, it is difficult to assess how these interventions are used in practice, how they play out differently in different contexts, or even the impact of these interventions on various outcomes of interest” (p. 3). Within this and other papers (Coburn, 2006; Coburn & Turner, 2011), the authors point out that, while data use primarily consists of data derived from large-scale testing, it can also include demographic data, local assessments, and classroom observations, but usually defines the use of data that already exists. To date, few studies look at the data derived from local assessments or look at teacher control over the generation of data as a factor in how it is used (Wallenstein & Hatch, 2014).

In terms of definitions, Marsh (2012) also includes local assessments in her definition of data when she describes data generally as all “actionable knowledge.” She also clarifies what “outcomes of interest” are. One important contribution to the conceptualization of data use is the fact that Marsh suggests a process that must occur for data (raw facts) to be interpreted by teachers and administrators to become information, then actionable knowledge. These processes are situated in schools and are therefore impacted by context. Each step described by Marsh is also a potential point where agency may be achieved or not.

An intervention may support users in accessing or collecting data (1), filtering, organizing, and analyzing data into information (2), combining information with expertise and understanding to build knowledge (3), knowing how to respond and taking action or adjusting practice (4), and

assessing the effectiveness of these actions or outcomes that result (5). (p. 4, Table 1)

This comprehensive list of steps suggests that there are many potential places for the data use process to succeed or fail. It also supports the position that a close examination of practice is necessary for researchers to understand how data can influence outcomes. Furthermore, it shows how the field seems to leave out the creation of assessments and the nature of the assessments being created as an additional step (Wallenstein & Hatch, 2014).

Jennings (2012) suggests a different frame for understanding data use in practice by describing no fewer than five distinct uses for data by practitioners.

Together, these types of data use capture how teachers view their schools, students, and themselves (*lens*); how they determine what's working, what's wrong, and why (*diagnosis*); what they should do in response (*compass*); how they establish whether it worked (*monitoring*); and how they justify decisions to themselves and to others (*legitimizing*). (p. 4)

She goes on to clarify the differences between “productive data use” as “practices that improve student learning and do not invalidate inferences about student and school level performance that policy makers, educators and parents hope to make” (p. 4) and “distortive data use” as any practice that leads to invalid inferences about schools and teachers.

Just as Jennings suggested that data use can be productive or distortive, other scholars have picked up on the situated nature of data use by focusing on its relationship to power. Honig and Venkateswaran (2012) point out that only certain data are collected, and those data often serve one group better than another. Weiss (2012) seems to take an agency orientation when she talks of the need for data systems to help practitioners. “If data systems are to improve performance, the data systems should be designed to help school staff (particularly teachers and principals) do a better job of instruction in order to enhance learning” (p. 9). She goes on to discuss two important elements of agency within the practice of data use—trust and control over practice. She says of trust that data must be safe for teachers. “Trust in relationships within schools led teachers to believe that data would not be used against them” (p. 2). She even describes situations of low trust where teachers

manipulate data to “conceal performance so as to protect themselves” (p. 5). Weiss says of control that, without control over practice, teachers cannot use data to improve practice, implying that agency in and around data use is a necessary part of school improvement. This focus on the issues of trust and control also suggests that how teachers are supported could be a factor in how they use data to inform their practice. Weiss concludes by suggesting that a research agenda into data use that focuses on the crucible of the classroom or on the improvement of learning outcomes “should explore the ways in which teachers understand and respond to the data collected and reported, and the ways in which the consequences flow to teachers themselves from data” (p. 7).

Situating the use of data even further in the direction of the teacher’s impact on students, Supovitz (2012), in the same journal, suggests three specific roles for assessment data to assist the teacher in improving student learning, all of which have to do with feedback: “(1) feedback from the assessment event directly to the learner, (2) feedback from the assessment event to the teacher, and (3) feedback from the teacher to the learner” (p. 4). He goes on to suggest three types of information that assessments ought to give: “(1) information in the assessment system about students’ developmental path toward a learning goal; (2) information in the assessment about students’ thinking processes; (3) information in the assessment about the students’ misconceptions” (p. 4).

The situated nature of data use as defined and framed by the authors in *Teachers College Record* is seen in some recent dissertations into data use. Also, many of the findings connect data use with certain qualities of agency described in the previous section. However, with a few exceptions, the methodological approaches of these qualitative studies lack actual observations of practice. For example, five studies since 2005 (Butler, 2009; Grantz, 2005; Groh, 2013; Johnston, 2009; Van Maele, 2007) consisting of surveys, interviews, or both only had limited findings, including variability in data use among teachers, the significance of structural support for data use in terms of time and training, and some uneven findings about data use’s impact on teaching practice as reported by the

teachers in the studies. In terms of agency, Groh (2013) did specifically look at issues of trust between leaders and teachers, and Johnston (2009) looked specifically at time for collaboration. But all of these studies were dependent solely on what the subjects reported doing with data without any observations about what they were actually doing with it. Spillane (2012) warns against a methodology that looks only at the “ostensive” and not the “performative” aspects of data use. Moss (2012) suggests that insightful research connecting policy to practice (macro to micro) should bring “attention not only to the formal structures, designed interventions, emergent processes, and indirect, secondary, and unintended effects” (p. 226). Similar to the agency studies in the previous section, research that does not include direct observations of practice seems to yield fewer significant findings. Or, as Moss warns, “We miss an important part of the picture when we focus, primarily, on the intended, the designed, and the formal” (p. 226).

A smaller number of studies (Pham, 2011; Rhude-Faust, 2011; Young, 2005) did incorporate direct observation into their methodology and also revealed various elements of agency in their findings. Significant to the current study methodologically, some studies that focus on the situated nature of data use employ sense making theory to their research. The best justification for the use of sense making begins with a question that suggests the linkages between policies and practice: “How do policies, social structures, artifacts, and attendant classifications that span organizational contexts get taken up in local practice, and how in turn are they generated in and from local practice?” (Moss, 2012, p. 223). One possible answer to such a question is this: “Drawing on theories of sense making, researchers have argued that the way teachers come to understand and enact instructional policy is influenced by prior knowledge, the social context within which they work, and the nature of their connection to the policy or reform message” (Coburn & Talbert, 2006, p. 478). In a second study that uses sense making theory, the same author offers this working definition of sense making about a year later: “Sense making theorists suggest that action is based on how people notice or select information, and then act on these

interpretations, developing culture, social structures, and routines over time” (Coburn & Talbert, 2006, p. 345). Coburn's (2005) work with sense making also suggests the importance of specific actors, such as school principals and collective actors.

In sum, data use literature in general and those studies that draw on sense making theory in particular raise important considerations for this study, which sits at the intersection of assessment, accountability, teacher agency, and the reform of practice. First, this research points out the tension between the two theories of action described in Chapter I: “The tension between the use of information for external monitoring control and accountability, on the one hand and the locally relevant inquiry, decision making, and professional or organizational learning, on the other” (Moss, 2012, p. 229). Second, the research into data use, like the research on teacher agency, refers to the multiple interacting factors and conditions, or what Moss calls the “complex mechanisms” (p. 224) at play in understanding how assessments and assessment policies have the potential to impact practice and outcomes in the classroom. This includes the three factors highlighted in this study: the teacher's role in creating and using local performance assessments as part of a team, the role of facilitators and other people in the process, and the role of accountability structures in the process. Finally, data use research underscores the methodological need to look at practice (the performative) directly and not just interviews and surveys being done in conventional data use literature.

Taken together, the conclusions from the first two sections of this review can be used to focus an examination of the contested nature of assessment reform, both historically and in current research, as seen through the two approaches of accountability assessment and formative assessment processes. In particular, this lens can examine the conditions created for teachers by both approaches. Turning now to these stories will help reveal gaps in our current understanding and reveal the need for specific case studies into local assessment practice in general and specifically the creation and use of local performance assessments by teachers to inform and improve their practice.

Two Paths Emerge

Looking at the competing assessment approaches and their underlying theories of action through a historical lens reveals that the connection between assessment and reform, as well as the tension between the competing purposes of assessment, can trace its roots well back into the 20th century. Furthermore, these same tensions are very much alive in the current debate surrounding the Common Core Learning Standards (Smith et al., 2014). Historically, questions surrounding assessment and reform have been related to the needs of the learner as well as the needs of society (Kliebard, 2003). Is society best served with one public school curriculum or many? Can all students learn one curriculum, or should they have access to one that best suits their needs? How will the outcomes of public education serve the learner and society? However, these questions were answered for most of the last century, there has been a commonly held belief that schools must respond to the changing structure and makeup of American society. From scholars such as Franklin Bobbit in the social efficiency movement to progressive educators led by John Dewey, all were concerned that the traditional approach to schooling was not meeting the needs of many children (Tyack, 1974).

While the measurement story traces its roots well back into the 20th century, for many researchers, including education historian Diane Ravitch (2011), the current policy environment traces its origins to the 1980s following the publication of *A Nation at Risk* as well as the publication of America 2000 by the first president Bush in 1991 (Darling-Hammond, 1995), which attempted to respond to the call for national standards and national testing by the year 2000. Linn (2000) and O'Day (2002) also describe this most recent reform effort as different from the others. Critics of neoliberalism would also point out that this shift in policy coincided with a time period when the role of the state was shifting. The welfare state was in question and being dismantled in the name of privatization and personal responsibility (Biesta, 2009). In what is described as the "new accountability,"

three elements make it markedly different from other assessment-based reform efforts and at the same time very similar to the theory of action found in accountability assessment.

They are:

a) the emphasis on setting demanding performance standards as the basis of assessment and accountability, b) the dual emphasis on setting demanding performance standards and on the inclusion of all students, c) the attachment of high stakes accountability mechanisms for schools, teachers, and sometimes students. (Linn, 2000, p. 8)

Most recently, in their study of the Common Core Learning Standards, Smith et al. (2014) have introduced another term to help describe the “new accountability.” That is the concept that assessments are primarily “outsourced” or produced outside of schools with little or no input from teachers in classrooms, as opposed to locally produced assessments of classroom performance.

One essay that helps place the work of this study into this larger theoretical and policy context was written by Haney (1991), “We Must take Care: Fitting Assessments to Functions.” As the title implies, Haney makes the case that there are many forms of assessment and that they each serve a function. He is concerned that national testing will favor one form of assessment—namely, large-scale, outsourced, standardized tests for the purpose of accountability—and that it will displace all others. Haney suggests that there are, in fact, three types of assessment that meet three discrete purposes: Ensuring School Accountability (p. 144); Improving Instruction (p. 148); and Helping Students Learn (p. 152). Haney's conclusion is clear:

Accountability seems to me the least important of the three forms of assessment discussed here. If the primary motivation behind the America 2000 proposal is indeed educational, in that what we need—more than bigger accountability schemes and greater sanctions attached to test results—are better ways to help children learn and better ways to help teachers and parents to help them do so, then we ought to focus time and energy devoted to assessment directly on those ends. (p. 159)

This sentiment is echoed in Darling-Hammond's (1995) collection of scholarly essays on authentic assessment four years later. In essence, Haney is describing the assessment

reform narrative at a crossroads in 1991. He points down two paths that policymakers, researchers, and practitioners could follow—two paths similar to the bureaucratic accountability and professional accountability described in O'Day's study a decade later. At the time, both paths shared the same goal of reforming public education to improve student performance and to increase equity but did so with very different theories of action.

And so, a look at recent history reveals that two paths represent two very different views of reform.

One view seeks to induce change through extrinsic rewards and sanctions for both schools and students on the assumptions that the fundamental problem is a lack of will to change on the part of educators. The other view seeks to induce change by building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches. (Ancess & Darling-Hammond, 1996, p. 57)

In my initial proposal, my literature review explored two paths of research. I have moved the literature review of one path—Accountability Assessment—to Appendix B, as my study revealed more about the second path specifically.

Classroom Based Formative and Performance Assessment: “To Use Their Minds Well”

This section will review classroom-based performance assessment through historical and contemporary research, both of which argue that locally created performance assessments could play a significant role in reform. It will also show that the phenomenon of teacher agency is useful in understanding how performance assessments can influence teacher practice as well as how policy can influence teachers' ability to create and enact performance assessments. Finally, it will continue to underscore the significance of research methodology in understanding the potential gaps in current research that attempts to link assessment, accountability, and reform.

In keeping with Haney's assertion in 1991 that there are multiple purposes for assessments and that large-scale, “outsourced” standardized assessments may be the least

useful as a reform mechanism, a line of assessment theory, practice, and research based on student performances designed and implemented by classroom teachers emerged. Often named classroom “performance assessments” and “authentic assessments,” this approach to reform, which this study calls “formative assessment processes,” was also connected to the sense of urgency that came out of *A Nation at Risk*, as described in several reviews at the time (Glaser & Silver, 1994; Grant, 1991; Wolf et al., 1991). In one review of research contemporary with Haney’s essay, the concept of classroom-based performance assessment is defined as follows: “In these evaluations, students are asked to write, to read, and to solve problems in genuine rather than artificial ways” (Wolf et al., 1991, p. 55). Linn and Baker (1984, in Wolf et al., 1991), describe six qualities of classroom-based formative assessment processes: “open-ended tasks higher order, complex skills, extended periods of time for performance, group performance, student and teacher choice of tasks ... judgmental scoring” (pp. 87-88).

As a group, proponents of classroom-based formative assessment processes as a vehicle for reform would support O’Day’s (2002) three-part definition of “professional accountability.”

First it is centered on the process of instruction.... Second, much of the focus of professional accountability concerns ensuring that educators acquire and apply the knowledge and skills needed for effective practice.... Third, professional accountability involves the norms of professional interchange.
(p. 20)

In his introduction to a series of reviews on assessment and reform in the *Review of Educational Research*, Grant (1991) describes the way researchers criticize the sorting nature of standardized testing as well as how they ignore the complexity of teaching and learning and how they do little to improve either. Glaser and Silver (1994) trace the historical roots of classroom-based performance assessment back to the Progressive Movement and look to ways scholars like Grant have tried to understand assessment

beyond the “selective model,” where the primary goal of assessment is sorting to a formative, “adaptive model” that

is suggested by the findings of several decades of cognitive research that has pointed to the constructive nature of human learning, the complex nature of human expertise related to specific subject areas, the power of intuitive conceptions, and the limitations of school knowledge to non-school settings. (p. 407)

Glaser and Silver suggest that formative assessment processes take into account “the kind of educational settings and social values to be served” (p. 406) and, in so doing, suggest two conditions that are necessary for assessment to promote learning. “First, the outcomes being tested must be recognized and accepted as important objectives of the instructional program.... Second, achievement assessments must be planned and implemented as an integral part of the curriculum and the program of instruction” (p. 411). In this way, Glaser and Silver echo Messick’s (1989) idea that the validity of an assessment must look beyond simple measurement and consider the socio-cultural context of the assessment. They also reinforce Weiss’s (2012) idea that teachers need to trust that the assessments are being used toward the productive end of student learning—to improve, not prove, performance (Vidovich, 2009). Finally, they make the case for assessments created in the classroom.

Perhaps the most direct connection to classroom-based formative assessment processes and the reform vision outlined in America 2000 is the review by Wolf et al. (1991), in which they draw on a phrase from America 2000’s set of national educational goals: “And every school in America will ensure that all students learn to use their minds well, so that they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy” (p. 32). Wolf et al., in fact, use the phrase “to use their minds well” in the title as well as throughout their article as a way of suggesting where many test-driven policies are off the mark as well as describing a vision for classroom-based formative assessment processes as the drivers of reform. The authors position themselves in support of performance assessment—what they call an “assessment

culture”—by describing standardized multiple choice testing or “testing culture” as “technically elegant but distorting” (p. 33).

Wolf et al. (1991) argue two points about the role of assessment in reform. First, a shift toward the “assessment culture” created by performance assessments requires a shift in the epistemology of learning. Second, while performance assessment has promise, it still faces many “difficult questions that stand in the way between promise and realization” (p. 33). In terms of the epistemology of learning, Wolf et al. review the history of educational assessment in the 20th century to differentiate between an “epistemology of intelligence” where it is assumed that intelligence is fixed and measurable, which leads to the sorting or selection-focused activities of a testing culture and an “epistemology of mind,” where it is assumed that the mind and thought can be cultivated and improved through learning. It is in this second epistemology that Wolf et al. place classroom-based performance assessment as an endeavor to evaluate a student’s development over time. Citing the work of William James, the authors posit,

If we want rigorous evidence concerning educational progress, we must describe large scale accomplishments and we have to think developmentally, collecting longitudinal data that follow the growth of a student “in the long run” ... as compared to measurement, assessment is inevitably involved with questions of what is of value, rather than simple correctness. (p. 51)

They continue to credit William James when they suggest an agency orientation with regard to assessment: “In this light, assessment is not a matter for outside experts to design; rather it is an episode in which students and teachers might learn, through reflection and debate, about standards of good work and the rules of evidence” (p. 52). As such, the authors appear to argue for locally produced and administered assessments that involve students “in an effort to break free of the artificial knowledge evaluated in many standardized tests” (p. 55). This point is echoed 20 years later by contemporary scholarship that argues against “outsourced” assessments driving practice in and around the Common Core Learning Standards (Smith et al., 2014).

In a second move toward an agency orientation, Wolf et al. (1991) connect classroom-based formative assessment processes to the need for public accountability.

It is in part holding educators accountable to an outside community for good work done well. However, it is also designed to promote intense discussions of standards and evidence among all of the parties who are affected. This involves a move away from presumptions of pure measurement to a model of clinical judgment. (p. 59)

The move to classroom-based performance assessments is not always easy. Wolf et al. (1991) close by framing three challenges for the work of educators: efficiency, equity, and evidence. In 1996, Wolf helped publish a volume exploring initial research entitled, *Performance Based Student Assessment: Challenges and Possibilities* (Wolf & Baron, 1996). In the first chapter of this volume, Demi Palmer Wolf and Sean Reardon highlight two “problematic legacies of multiple choice testing.” The first is “meaninglessness. Because, tests have so long been curriculum independent, we have destroyed both students’ and teachers’ sense that any prior effort or specific learning really matters” (p. 32). The second, relating directly to Popkewitz’s and O’Day’s notions of agency and Ball’s notion of the “post-professional,” is “the constriction of teacher’s professional judgment” (p. 32). In response to these concerns, they suggest, “We clearly need a system of assessment that is curriculum dependent. Such assessments reconnect effort, teaching, assessments and results” (p. 19).

Jacqueline Anness and Linda Darling-Hammond (1996) build on the sense of need for new performance assessment systems in what they describe as a “current movement” toward “authentic assessment.” Anness and Darling-Hammond draw on Grant Wiggins’s (1989) definition of authentic assessment processes that engage students in “‘real world’ tasks rather than multiple choice tests, and evaluate them according to criteria that are important for actual performance in the world outside the school” (p. 54). The one study the authors cite is based on Project PACE (performance assessment collaborative for education), which took place in urban middle schools. The program looked at the efficacy of common performance assessments produced in schools that had significant professional

development for the teachers in the project. The results were not consistent among schools but could point to examples where the work on assessment influenced teaching practice and curricular changes, the most significant of which was the relationship between curriculum and assessment. The shift from “curriculum independent to curriculum embedded assessment” (Palmer Wolf & Reardon, 1996, p. 19), where “teachers have turned their attention from checking or correcting to responding to student work” (p. 21), is similar to the work of data use scholars, who suggest that assessments should be used to impact practice and outcomes (Moss, 2012).

Ancess and Darling-Hammond (1996) saw the development of classroom-based formative assessment processes very much as part of the reform agenda following *A Nation at Risk*. In contrast to standardized tests, they posit that performance assessments play a positive role in school reform for three interrelated reasons that all point to Popkewitz’s and O’Day’s concept of teacher agency. First, like Luke’s claim that assessments as “texts” “make teachers,” Ancess and Darling-Hammond recognized that assessments influence practice. “Assessment, especially when it is used for decision making purposes, exerts powerful influences on curriculum and instruction” (p. 54). In fact, they suggest that teachers and schools engaging with the right kind of assessments and assessment data can be agents of change for schools at the level of practice. “Assessment can exert powerful influences on behavior, it can be used to change school organizational behavior as well as classroom work” (p. 55). In contrast to earlier research on the influences of testing on teacher behavior, which tended to push instruction toward “lower order cognitive skills” (p. 55), Ancess and Darling-Hammond argue that authentic assessments can drive instruction toward higher-order skills.

Second, Ancess and Darling-Hammond (1996) argue that assessment development and use redefine the nature and role of accountability to include an agency orientation. Rather than accountability to a score or a number, authentic assessment is “learner centered” (p. 58). Almost anticipating Ball’s notion of the “authentic-professional,” they

describe learner-centered assessment in terms of professional knowledge. “Assessment should help schools provide education that is both *responsible*, that is informed by professional knowledge of good practice and *responsive*, that is appropriate to individual student needs” [their emphasis] (p. 58).

Ancess and Darling-Hammond (1996) do not claim that classroom-based formative assessment processes are an easy solution or that this approach does not present real challenges to teachers and to schools. In fact, in their third point related to teacher agency, Ancess and Darling-Hammond point out that a number of factors and conditions need to be addressed, which include “teacher knowledge, school capacity for improvement and problem solving, flexibility in meeting the actual needs of real people, shared ethical commitments among staff, and appropriate policy structures that encourage rather than punish inclusive education” (p. 58). Ancess and Darling-Hammond also cite Peter Senge’s work on learning organizations by explaining Senge’s assertion that learning begins with “intrinsic motivation—people’s innate curiosity and desire to improve their work, which is encouraged when they have the opportunity to discover, experiment, observe the results of their actions, and refine their approach” (p. 59). Put simply, schools and teachers need the space to refine their practice through experimentation, struggle, observations of results, and refinement, all of which require that teachers have the ability to make judgments that comes from having responsibility at the local level to make decisions about learning goals, standards, and approaches. Ancess and Darling-Hammond point out that the need for teachers to have the conditions whereby they can achieve agency is not new. In fact, they describe the dangers of removing local responsibility by citing research from the 1930s:

The study found that the more successful schools were characterized not by the particular innovation they had adopted but by their willingness to search and struggle for valid objectives, for new strategies, for new forms of assessment.... The removal of local responsibility for thinking things through then deprives schools and community members of the opportunity to engage in the empowering and enlivening dialogue needed for making change. (p. 60)

Ancess and Darling-Hammond summarize the third argument for teacher agency:

Thus, when a safe environment for innovation has been created, responsible risk taking is strengthened by opportunities for evaluating results. Engaging teachers in assessment is a critical aspect of the evaluation process that can help transform schools from procedural bureaucracies to learning organizations. (p. 60)

In this three-part explanation, Ancess and Darling-Hammond (1996) touch on three factors that are examined in this study: local control over assessment creation and use; facilitation and training to support the work; and a purpose for assessment beyond accountability to the state. In their research, Ancess and Darling-Hammond cite their research conducted largely in New York City schools; among their findings are observations of increased examples of teacher collaboration in and around assessment practice. For example, “At Brooklyn’s P.S. 261, use of the Primary Language Record transformed professional development from the workshop model of information transmission to a dialogue and inquiry model of collegial knowledge building” (p. 66). The research describes similar changes in several New York high schools.

The work on formative assessment processes including performance assessments continued through the 1990s and into the first decade of the 21st century parallel to the testing movement that gave rise to NCLB, RTT, and NYS 3012-C, as evidenced by Darling-Hammond’s 2008 publication, *Powerful Learning*, which traced recent practice and research into assessment both in the United States and internationally. Like the work almost two decades earlier, Darling-Hammond continued to present the need for performance assessments and to maintain an optimistic stance that reform in teachers’ practice and student outcomes was possible. She also added some refinements to her earlier work that fall into the area she describes as “the rub between theory and practice” (p. 204). First, she recognizes the need for curriculum to balance direct instruction and “inquiry” projects. Keeping pace with the “new accountability” and its emphasis on rigorous standards and success for all, Darling-Hammond’s approach also implies the need for a balance between

more traditional testing alongside performance assessments that emphasize the process of inquiry, knowledge construction, and presentation. To achieve this balance, Darling-Hammond offers two other refinements to her earlier writing that also reflect the notion of teacher agency. First, she emphasizes the need for teacher skills in the area of assessment creation and use. Second, she emphasizes the need for the right conditions for teacher agency: “Teachers need time—and a community—to support their capacity to organize sustained project work. It takes significant pedagogical sophistication to manage extended projects” (p. 70). Again, Darling-Hammond suggests key factors that are explored in this study—facilitation and training—as well as conditions that foster collaboration and trust based on a sense of purpose beyond accountability to the state.

In the same year that Darling-Hammond published her work on assessment and learning, Giselle Martin-Kniep (2008) published the first of two works that would form the basis of the PADI project. Her work also echoes the theory, research, and practice described as classroom-based performance assessments and authentic assessments two decades earlier. Like Darling-Hammond, Martin-Kniep’s work has several refinements in what she describes as “methods of practice” for authentic assessment. In particular, her refinements include the direct use of standards (both external and local), rubrics as well as training for teachers on assessment development and use. In addition, Martin-Kniep never backs away from the need to experiment, observe results, and refine work. Like Darling-Hammond, Martin-Kniep calls for a balance between traditional tests and performance tasks that serve both formative and summative purposes. Traditional instruction and assessment in and around rigorous standards used in concert with performance tasks help students learn new content and then make meaning of their work, while also giving feedback to teachers who can support the learning (Martin-Kniep, 2009). In this way, Martin-Kniep describes the idea of assessment as a process, not an event, in the following comparison: “Assessment as an event is a narrow and myopic enterprise, pinning all impressions and evaluations on a

single isolated experience. Assessment as a system provides a rich montage of students learning, created from a collection of multiple pieces of evidence” (p. 80).

The idea that assessment is a process, not an event, clearly links the recent scholarship on performance assessments and authentic assessment practices with the work done in the early 1990s. In fact, Martin-Kniep’s words seem to echo those of Aness and Darling-Hammond, who said in 1996, “Assessment systems should support rather than just measure learning” (p. 58). As a process, theorists, researchers, and practitioners recognize that reform is needed—teachers, schools, and curriculum need to change. But they also take the stance that this change will come only from “the bottom up.” Just as Wolf, Aness and Darling-Hammond, and Martin-Kniep point out the challenges in performance assessment, so does O’Day (2002), who describes two limitations in professional accountability that relate to potential challenges for performance assessment. First, O’Day describes the traditional structure of schools as highly “atomized,” where the judgment of the individual is favored over the collective. This, when combined with the variability of teachers’ knowledge and skills around curriculum and assessment, makes schools as systems difficult to move forward toward more rigorous outcomes for students—a collective agency in and around assessment use for reform. Linda Darling-Hammond in 2004, and again in 2009, reaches a similar conclusion when she posits that the potential improvements in outcomes for students cannot be achieved without technical support for teachers as well as resources that will encourage teachers to collaborate toward shared standards and shared goals for students. Shavelson (1992) reaches a similar conclusion with the following summary: “The measurement reality, judging by our research, is more sobering than the political rhetoric” (p. 26). Referencing America 2000, Shavelson is merely cautioning educators and policymakers that the technical questions surrounding performance assessments may prove daunting despite their promise. His specific examples fall into two of the three categories—efficiency and evidence—described by Wolf et al. (1994). O’Day’s second point, also echoed by Darling-Hammond and Wolf et al., is that of equity.

Many research studies spanning the past two decades into performance assessments as a subset of formative assessment processes follow a similar narrative structure in terms of describing an initiative and attempting to understand the impact either quantitatively by looking at student outcomes on standard assessment measures, or qualitatively through teacher interviews and observations. Thus, a second gap in research comes into focus. In almost every case, the practice being observed is the performance assessment itself, as opposed to practices that are observed or changed as a result of creating or using performance assessments. The majority of studies describe the challenges associated with implementing performance assessments. On one level, these challenges touch on one aspect of teacher agency, specifically the skills required to create assessments and the institutional support in terms of resources and training required to do so. To use Warren Little's (2011) language, these studies are still normative in nature in that they describe the how and what of classroom-based performance assessments but not their potential to impact teacher practice or student outcomes. In other words, they discuss the agency achieved (or not) by teachers to create performance assessments, but they do not study the potential agency achieved by creating and using data derived from performance assessments. They do not attempt to observe the transformative potential of these assessments as described by Aneess and Darling-Hammond.

In sum, 20 years of research into classroom-based formative assessment processes bring many possibilities but not as many results in the form of large-scale and sustained changes in practice. Perhaps the words of Wolf et al. (1991) still ring true 20 years on when they suggest that there still are many "difficult questions that stand in the way between promise and realization" (p. 33). In addition to drawing connections between the literatures of teacher agency and data use and the role that formative assessment processes can play in reform, several gaps in research become apparent. Specific to this study, two significant gaps in the research have been revealed. First, while the factor of training is indicated in several studies, there still remains a gap in our understanding as to what aspects of data

creation and use are actually impacted by training or professional development. Second, the practice being observed is the performance assessment itself as opposed to practices that are observed or changed as a result of creating or using formative assessment processes.

Despite these gaps, there is still a potential for optimism in pursuing the path of classroom-based formative assessment processes, especially when compared to the path of accountability assessment. As several scholars have already pointed out, “in an effort to break free of the artificial knowledge evaluated in many standardized tests” (Ancess & Darling-Hammond, 1996, p. 55), other assessment processes are necessary if we truly want to evaluate and support students “to use their minds well.” The work will not be easy, but if teachers are given agency—belief that they have the ability and the authority to make changes in their practice based on their own judgment—then these changes are possible. In fact, that note of optimism runs through the scholarship of the late 1990s and is even seen in the work of Darling-Hammond and Martin-Kniep, whose work was published less than 12 months before the Race to the Top legislation in 2010 and the subsequent legislation in New York about one year later. So until quite recently, some theorists, researchers, and practitioners seemed to be heeding Haney’s warning about the need to follow the right assessment path. Sadly, these recommendations have consistently fallen on deaf ears in Washington and Albany. As history has played out, the last 10 years have seen an increase in state and federal accountability testing accompanied by a research environment that seems to devalue investigations into classroom-based performance assessment (Feuer, Towne, & Shavelson, 2002) or, as Sanger (2012) describes it, our ability or inability to take a stance of “avoiding the strong temptation to misuse the seductive tools of the reform movement, paying attention to the complexity of teaching and schooling and their many sources of value, speaking a language that appropriately reflects that complexity as we work” (p. 303).

It seems now that the work on formative assessment processes in classrooms is often overshadowed by the first path taken by policymakers, where teachers are treated not as

learners but as a problem to be fixed (Apple, 2004). This, in turn, underscores the concern that the current policy environment may be an additional factor in creating the conditions that negatively impact teacher agency—their belief that they have the ability and the authority to make changes in their practice based on their own judgment. In the case of assessment-driven reform, this would mean the sense that teachers have the ability and authority to use assessment data to make instructional decisions.

Synthesis

By examining selected assessment literature over the past two decades, we see that the two paths that Haney laid out in 1991 are not simple opposites, but do suggest a set of consistent observations over time where one path of testing, high-stakes accountability and a loss of teacher agency, is juxtaposed with a second path of formative assessment processes where agency is somehow a given. We have also seen the need to develop “a language that appropriately reflects that complexity as we work” (Sanger, 2012, p. 303). Central to this study is the need to understand under what conditions agency can be achieved in local contexts, thereby allowing teachers to create formative assessment processes and, in turn, use them to reform their practice. The body of research covered in this review should remind us of the words of Baker et al. (2010), who warn us that policy and practice can allow us to be “seduced by elegant simplicity” (p. 20) as we work to find answers to “difficult questions that stand in the way between promise and realization” (Wolf et al., 1991, p. 33). Historical and current research into accountability assessment has shown how policy can undermine agency, collaboration, and trust. The theory of action behind these policies influences practice in multiple and often unintended ways that seem to prevent reform initiatives involving “outsourced” assessments from actually changing practice. This point is countered by research into performance assessments that emphasize locally created and controlled assessment initiatives. Finally, scholarship into classroom-

based formative assessment processes has shown promise but still leaves gaps in our understanding of the “everyday practice that makes up schooling” (Little, 2012, p. 143).

The major sections of this review have also offered some suggestions for future research in terms of what to look at, where to look, and how to structure the research in order to better understand the relationships among assessment, agency, practice, and reform. Scholarship on teacher agency has given us a powerful lens to understand the complexity of how policies enter schools and become enacted in classroom practice. This body of research has also demonstrated that agency can be a collective phenomenon suggesting that social groups within schools impact whether or not agency is achieved. Data use scholarship reinforces the complex and situated nature of teacher decision making. That research exposes some of the links between the macro world of policy and the micro worlds of classroom practice, but only if researchers ask the right questions in the right methodological contexts. Sense making as a theory and as a methodology is one productive way for researchers to access examples of assessment data use at the level of practice. Both agency and data use research point to specific methodological considerations involving qualitative case studies in general and, in particular, studies that draw on multiple data points, including direct observations of practice.

At this point, gaps still remain in the research. These gaps include the need for studies in which assessments are created locally, studies that focus on the degree to which assessment work is facilitated and supported with professional development, and studies that focus not just on the “events” or product of assessments themselves but also on the “process” or practices that assessments can produce when used by teachers to examine practice.

That is why this study examines further the potential impact of locally created and implemented assessment practices on data use by examining the impact of assessment data derived from formative assessment processes that are (1) locally produced and strive to align with local and state learning standards, (2) produced in facilitated and collaborative

efforts where teacher teams work together to inform their practice, and (3) created within the context of state and local accountability structures. In doing so, this study attempts to understand two problems not currently addressed in the literature. First, it is not clear how data derived from locally created performance assessments actually impact practice. While much of the research in this review claims that knowledge can be built and changes in practice can be induced through local, collaborative assessment practices, the theory has not been well explored empirically. Second, while performance tasks and performance assessments have been developed and researched for two decades, only recently is this work being attempted within the context of high-stakes accountability. Therefore, the impact of the current policy context on teachers' ability to use data derived from locally created performance assessments is not well understood.

Chapter III

METHODOLOGY

Introduction

This qualitative case study investigated the development and use of local assessments in public schools, because I wanted to find out how the act of creating local assessments and using data derived from local assessments can influence teacher practice within current state accountability policy. This study looked at three interrelated factors that impact the creation of as well as the use of data derived from locally created performance assessments: (1) the manner in which teachers design, create, and implement the performance assessments; (2) the manner in which their work is facilitated through training and administrative support; and (3) the manner in which the work is part of local and state accountability structures. My methodology and theoretical lenses were shaped by recent data use literature and sense making theory, in particular, their emphasis on understanding teacher practice and teacher agency as situated in social and institutional contexts.

This chapter will outline the study's approach in five moves: First, it will describe the role sense making theory played as both a theoretical and methodological lens for the study. Second, it will describe the qualitative case study and research questions that were used. Included will be a description of my research sample within the case. Third, it will explain how sense making theory influenced the study's data collection in general and interview questions in particular, as well as my approach to data analysis. Fourth, it will describe some of my initial findings from pilot studies and how they influenced my study design.

Finally, it will describe some of the potential subjectivities that influenced the validity and the limitations of the study.

The Role of Sense Making in My Methodology

The use of sense making methodology is a source of methodological guidance for virtually every aspect of research step-taking—conceptualizing and framing questions, observing, interviewing, listening, and analyzing. (Dervin & Clark, 1999, n.p.)

It has been said that sense making serves as both a theory and a methodology (Foreman-Willmet, 2003). Developed in the early 1970s by Brenda Dervin, the theory of sense making has been applied to a variety of fields, including education. This section will outline some basic assumptions within sense making theory that help to justify its application when attempting to observe and understand teacher agency in the creation and use of local assessments. It will also point out some specific methodological considerations around data collection in general and interview questions in particular.

Dervin and other scholars who have used and studied sense making as a methodology point out at least three important assumptions about the situated nature of knowledge and learning. These assumptions, in turn, lend themselves well to a study of teacher agency, which was described in Chapter I as follows: “All this implies that agency is not something that people *have*. It is, as we suggest, something that people can *achieve*, and they can only achieve it in transaction with a particular situation” (Biesta & Tedder, 2006, p. 19). The first assumption that sense making theory brings to this study is simple but important. Since people learn in particular times and locations, the situated nature of their learning must be taken into consideration in any study. “For sense making, the answer to attending to people as potentially changeable across time is to re-conceptualize the unit of attention in the research and system design from the person to the person-in-situation” (Dervin, 1998, p. 40).

The second assumption is that “knowledge is a verb, always an activity, embedded in time as space” (Dervin, 1998, p. 36). This means what a person knows about a situation is the result of a process of constructing that knowledge as opposed to simply receiving knowledge as a discrete or defined thing. Dervin goes on to reinforce the first two assumptions by saying that this process of making knowledge occurs “at the juncture between self, culture, society, and organization” (p. 36). This suggests the third assumption of sense making, which is that learning is situated. Learning occurs in social contexts and is therefore influenced by the people and organizations around us. This point is expanded upon: “From the perspective of sense making, who we think we are (identity) as organizational actors shapes what we enact and how we interpret, which effects what outsiders think we are (image) and how they treat us, which stabilizes or destabilizes our identity” (Weick, Sutcliff, & Obstfeld, 2005, p. 416). Combined, these assumptions suggest that teacher agency can take several forms or can exist toward several ends—agency for the teacher to change practice; agency for the students to improve learning; and agency for a collective such as a team of teachers or a school to create organizational change.

On the other hand, this third assumption from sense making theory about how people create knowledge within organizations can take a different turn when seen through the lens of institutional theory. Sense making scholars make an important connection to institutional theory to help explain these limitations; according to institutional theory,

Organizational members are socialized (indoctrinated) into expected sense making activities ... behavior is shaped by broad cognitive, normative, and regulatory forces that derive from and are enforced by powerful actors such as mass media, governmental agencies, professions, and interest groups. (Weick et al., 2005, p. 417)

This observation has important implications with regard to agency. It suggests that there exists the danger of over-emphasizing the potential of agency in this study or any study of teachers and assessment. Popkewitz (2008) and Birdwell-Mitchell (2015) also point out the potential limitations of agency by suggesting that an individual’s agency is

constantly checked by the norms of the group. Birdwell-Mitchell goes on to say that “teacher agency can both change and maintain institutionalized instructional practices in schools” (p. 140). There is, therefore, an inherent challenge in this study to be aware of and perhaps attempt to control for the institutional forces when observing teachers developing assessments in a school environment. In the case of this study, the fourth assumption played a role in the selection of my case, my subjects, and my data samples.

PADI: The Case Defined

In this study, local assessments are defined as assessments designed collaboratively by teams of teachers with the support of the Performance Assessment Development Initiative (PADI) with the intended outcomes of measuring defined curricular goals, providing feedback to students about their learning, and providing feedback to teachers about the nature of their instruction. Meaningful outcomes are defined in the data use literature by Coburn and Turner (2012) as “(a) outcomes related to student learning; (b) those related to changes in teacher and administrative practice; and, (c) those related to organizational or systemic change” (p. 177). One of the two unifying goals of all the PADI schools in the consortium is an interest in promoting creative and critical thinking as measured by performance assessments (Martin-Kniep & Wilson, 2015). The end product of each PADI project is a curriculum defined by state and local standards. Each project utilizes pre-assessments and through-assessments to help guide teachers and students toward a culminating performance assessment that promotes creative and critical thinking. Teachers involved in PADI come from suburban districts in the Tri-States Consortium—a collection of approximately 40 school districts in New York, Connecticut, and New Jersey.

A Summary of the PADI Process

The PADI process has within it a set of assumptions that outline a linear progression of planning, which includes the creation of a standards-based goal, a pre-assessment activity, rubrics for evaluating the students' progress, curricular activities in the forms of lessons and traditional tests to drive the instruction, and a culminating performance assessment from which growth is evaluated. The process assumes that teachers use data generated from the pre-assessment and during the lessons to adjust their instruction as they progress toward the final performance assessment, which is designed to produce data that not only informs student growth, but also informs teachers' future instruction.

The PADI training for teachers consists of three training sessions. The first two occur before the school year begins, and the third occurs in the fall. Each training session follows a similar pattern. There is direct instruction, followed by time for groups to work in their teams. The day usually finishes with some peer review and reflections as well as a session for facilitators after the teacher teams have left.

In the first session, teams spent a day learning more about the PADI year and began to think about how to select the topic or focus of their project. The opening presentation to the teams focused on defining assessments as a shared set of expectations that are evaluated before, during, and after instruction has occurred.

In the second session, teams worked for three days on the design process. This included re-emphasis on the connection between standards and tasks as well as an introduction of the blueprint as an organizing structure for the assessments (see Table 1). Martin-Kniep emphasizes that the blueprint begins with an overarching task aligned to specific standards. These standards can focus on content, skills, or dispositions like the six mentioned in the District's Shared Value Outcomes (Collaboration, Communication, Problem Solving, Thinking, Innovation, and Commitment to Growth). It is here that Martin-Kniep suggests a broad definition of assessment and assessment data by suggesting that the

defined standards and their corresponding rubrics will be refined as the project develops, and especially after the initial pre-assessments are administered in the fall. The blueprint moves on to listing specific student activities as well as corresponding teacher behaviors that support the activities. In sum, Martin-Kniep stresses that the blueprint aims to help the teachers align what they value with what they will ultimately measure in the form of student growth.

The second session also covered assessment topics, such as the nature of feedback and rubrics and the role of rigor in assessments. Below is a brief summary of the blueprints from each team.

Table 1. Team Goals and Plans Summarized

	Shore School	Hill School
Goals as described by each team	Develop a social studies assessment that has a focus on student's ability to read and interpret multiple points of view. Expand the use of rubrics from ELA to social studies to measure growth in specific skills and dispositions. Increase teacher and student use of rubrics to direct learning and measure growth.	Develop a social studies assessment that includes interdisciplinary work with Art that focuses on the relationship between beliefs and world view/perspective taking.
Essential Question	Where does the Truth lie? "How do I determine my own version of a story that has multiple perspectives?"	How do beliefs impact your view of the world?
Standards Planned	5 Standards: 2 reading, 1 writing (CCLS) and 2 Social Studies	8 Standards: 3 Art and 4 Literacy, 1 Social Studies
Rubrics Designed	Single Rubric: Covers the dispositions associated with taking multiple perspectives.	Two Rubrics: The Social Studies Rubric reflects the skills found in Teachers College Learning Continuum for Persuasive Writing and has one Social Studies category. The Art Rubric has categories but is mostly incomplete.

The third PADI session took place in November 2015, timed to occur after the pre-assessment. The goal of this session was to help teachers review their pre-assessments and make adjustments to the instructional lessons in the planned unit as well as to the standards, the rubrics, and the final performance assessments.

It is an explicit assumption of PADI that the work will spill over into other aspects of the curriculum. “They are also able to leverage their learning by making incremental but important changes to their units and lesson that provide students with spaces for performance and even authentic learning experiences” (Martin-Kniep & Wilson, 2015, p. 5). To that end, teachers involved in PADI are trained in the development and implementation of performance assessments that align to existing state and local standards that focus on content knowledge and skills. Many projects, including the two projects at Suburban Woods, also attempt to define dispositional standards like the ones defined by the District’s Shared Value Outcomes. Over time, the second unifying goal of PADI is for teachers not only to reform their own practice, but for them to influence the practice of colleagues in their districts.

As stated in Chapter I, PADI is one example of a case where practice designed to improve inputs and process is being played out in the current context of accountability policy designed to prove outcomes. To that end, it is also a case where policy forces and local district forces, as well as the social contexts within schools, are all impacting the agency of teachers who are attempting to reform their practice as well as the practice of their peers. Finally, PADI is an example of assessment practice being enacted in the “crucible” of the classroom. PADI represents assessment practices at the micro level akin to the data use scholarship described above, with the added distinction that these are teachers both making assessments and interpreting the data that those assessments produce to inform their practice. As such, teachers in PADI are in the process of sense making in and around assessment.

Research Questions

How do creating and using local performance assessments influence teacher practice?

- How do teachers who participate in the work of creating local performance assessments both as individuals and as members of a group use the data produced from these assessments to inform their practice or change their curriculum?
- How does this process influence their understanding of the role(s) that assessments play in their work? What changes do they make in their uses of assessment and in their curriculum?
- What factors support or limit teachers' participation in the creation of local assessments?
- What factors support or limit the changes that teachers make in their instructional practices or their curriculum?

Research Sample

For my dissertation study, I selected one school district to be named for this study as Suburban Woods School District. Suburban Woods is located in an affluent suburb of New York City. The district serves just over 4,000 students K-12 across five school buildings. Historically, Suburban Woods has a strong reputation for academic excellence and regularly places more than 95% of its graduates into four-year colleges. At the elementary level, two district initiatives impacted assessment and curriculum in the past year. Both initiatives were influenced by the district policy surrounding testing and assessment. First, the elementary schools had just adopted a standards-based report card that reflected the movement toward common units of curriculum in Math and Language Arts across the three elementary schools. Second, the District was also looking to add “Shared Value Outcomes”—non-standard measures of student work. Some of these shared values include Collaboration, Communication, Problem Solving, Thinking, Innovation, and Commitment to Growth. These

shared value outcomes, or SVOs, are part of the building goals for all principals in the district, but they are not yet part of the new report card.

This study followed several PADI teams from Suburban Woods through one complete cycle of design, implementation, and refinement over the course of one school year. As such, this case study was bounded by time and location. Within this case, the unit of study—“from the person to the person-in-situation” (Dervin, 1998, p. 40)—was individual teachers working in PADI teams. PADI teams are made up of classroom teachers and, on occasion, some special area teachers designing and implementing a PADI unit and performance assessment along with their supporting process facilitators. These process facilitators can shed light on the context of the PADI unit being developed and provide perspective on the teachers and on the goals of the school. The total sample size of teachers was six, including the process facilitators. Perhaps the best way to introduce the schools is to introduce the principals. Both teams in the study consisted of three members made up of one classroom teacher and one special area teacher supported by one process facilitator. The teams were from two different elementary schools in the district and, in both schools, the facilitator was the building principal.

The Shore School

The Shore School is a K-5 elementary school in Suburban Woods with a student population of 350 and 3 to 4 class sections per grade level. One distinguishing feature of the Shore School is that it has been running PADI cohorts the longest of any elementary school in the district. The team in this study was the fourth cohort to go through the training. At the time of the study, Charles was in his third year as principal of the Shore School. Prior to that, he was the Suburban Woods District Coordinator for Social Studies. He also had 10 years teaching experience. Charles described his background with PADI as one of his primary motivations for joining/creating this 4th grade PADI team. This project marked the third year of PADI projects in Suburban Woods, and he has supported at least one project in his

building each year. Charles creates PADI teams by asking for volunteers. He sees the current 4th grade project as an extension of existing social studies curriculum, picking up on an existing initiative in 4th grade to bring interdisciplinary work to Social Studies and English Language Arts. He believes that this group is interested in seeing themselves as designers of curriculum. He also sees this work as a district shift toward performance assessments away from the skills-based assessments currently in place. Outside of the literacy curriculum, where the teams have experience using rubrics and learning progressions with the Teachers College Units of Study, Charles would like teachers to make more use of rubrics and assessments as ways to engage students in their learning. Of the new 4th grade PADI team, he said, "So to be honest, this is an area they're still working on, there isn't a lot of dynamic rubric use and actually what we're doing next year, so the building focus for next year here is how to engage students through assessment, so it's the connection between assessment and engagement" (Charles Int 1: 4968-5439). He defines authenticity in assessment as "having a connection to the real world and to real experiences" (Charles Int 1: 8187-8539).

Charles sees a few trends shaping his team's work. The District, and his school in particular, are very outspoken against standardized testing. Teachers at the Shore School embrace the Suburban Woods Board of Education policy that explicitly values measures of performance and achievement other than tests. As such, he is proud of his teachers, who have taken on PADI projects over the past three years, and sees the work as impacting the entire faculty, which has a reputation in the District of being collaborative and innovative. With that said, as a principal, he understands that State Annual Professional Performance Review measures are still on teachers' minds even though most receive high ratings. He also sees pressure to cover district prescribed curriculum units as a factor in teacher work. At the conclusion of the first interview in the summer of 2015, Charles was optimistic that the 4th grade team had a strong plan, and he looked forward to supporting them as they work to develop a project that will teach and assess the standards and dispositions they value.

The Hill School

The Hill School is also a K-5 elementary school with a population of about 425 students or four class sections per grade level. As a school, they are newer to the PADI process, having only completed one project the year prior to the study. Lynne, the principal at the Hill School, was the facilitator of her 4th grade PADI team. She had been principal for 14 years, was assistant principal for 4 years, and had another 14 years as a classroom teacher prior to becoming an administrator. In the round one interview, Lynne described her interest in PADI in two ways. She had taken over as facilitator for a 3rd grade PADI team the year before and was impressed by the kind of work those teachers were doing with performance assessments. She was also inspired by a presentation of student work based on other PADI projects given to her administrative team in the fall of 2014. In addition to the focus on interdisciplinary work, Lynne is interested in bringing an assessment focus to her team's curriculum work. She saw this 4th grade project as a way to support the District's move toward interdisciplinary performance assessments as it teamed up a classroom teacher with the school's art teacher. Lynne defines "authenticity" as projects that have real-world connections, and she was proud of the 3rd grade project. "I think the goal for the team is to take a look at what we're teaching, bring it to a higher level through the authenticity piece, getting kids more engaged, but also focus a little bit more on different ways that you can assess students who perform with varied assessment opportunities" (Lynne Interview 1: 2754-3036). On several occasions in the interview, Lynne described authentic tasks as ones that move students to a "higher level." In terms of assessment experience, she said,

In the literacy, we have formal, formal and informal assessments. In science, we have teacher-made assessments. And then social studies, I think that's the area where it's a little bit weaker. I don't know that we have anything formal and more teacher-made or it's more project-based. (Lynne Int 1: 7748-8032)

Like the team in the Shore School, Lynne saw the project as an opportunity to bring assessment into the social studies curriculum at the Hill School. Since her school had

significantly less experience with the PADI process, she was happy that one of the team members on the new 4th grade team had worked on the 3rd grade project and was moving to the 4th grade this year. She also tied the work to the district initiative of SVO's shared value outcomes.

Well, we always have initiatives from the district and where we're looking at now is that we're working with Gisele [Martin-Kniep] on shared value outcomes. Our focus is going to be more on what we believe and what we value, and that would be incorporated a lot more into instructional goals and assessments. (Interview 1: 8818-9118)

She acknowledged that by the end of the summer session, the team had not fully defined the project, but that each teacher knew her individual role in their initiative to support interdisciplinary learning. In addition, she was hopeful that new assessments would drive the work.

Selection Considerations

The selection of subjects in this case was considered both convenience and purposeful selection (Marshall & Rossman, 2010). My reasons for selecting this district for study were three-fold. First, I made a conscious decision not to study assessment processes in my home school district. While I did have an interest in how local assessments are developed there, I realized that my role as an administrator in the district would compromise my ability to observe teachers and collect data (Agee, 2009).

Second, there are several features of the Suburban Woods district's approach to PADI that made it a potentially rich source of variability in teacher experiences in and around the PADI process. For three years, this district had consistently supported multiple PADI teams each year. Both the longevity and the scale of PADI in the district allowed the study to observe teachers from a variety of grade levels and disciplines. The following specific demographics were tracked for each subject: grade(s) taught, subject or specialization, teaching experience, and years in the district, including tenure status. In addition to potential variations among teachers, the selection of a location with multiple PADI teams

also allowed for greater fidelity in completing the study. Through watching PADI teams from multiple districts over two years, I had observed that not all teams complete the process by completing and implementing a performance assessment—a problem more common when a district enrolls a team for the first time. The most common reason for a team not completing PADI is a lack of district-level support for the project. For example, teachers have been known to work on the PADI assessment over a summer only to learn that their teaching assignment was changed in the fall. In other cases, the district sends a team to PADI to try to leverage the training to expand or legitimize existing traditional test-based assessments.

The Suburban Woods district had supported teams of teachers through the PADI process for all of the first three cohorts. This fourth cohort included six teams from four different schools as well as one team that was returning to the PADI process after completing the training in cohort 3. Another unique feature of this district was that it used the PADI process with the intention of developing assessments for an entire grade level. Therefore, they expected that the PADI teams would work with each other as well as teachers on their grade level who had not been trained. Teams in cohort 4 ranged from primary grades to middle school. Projects ranged from subject-specific topics, to special education and elective classes. While there were six teams from the District participating in Cohort 4, this study focused on two fourth grade teams from two separate elementary schools new to PADI as well as one third grade team who had completed the training in cohort 3 but was continuing the work this year. These three teams both had homeroom teachers who were impacted by state assessments directly in their curriculum as well as teachers in subject areas who were not tested. The homeroom teacher was the “teacher of record” for their students and was therefore seen as responsible for each students’ ELA and Math test results according to the State’s existing APPR plan. Their projects touched on a combination of state and local standards. Some of these standards were part of state assessments in ELA and Math and some were not. By limiting the sample to these three

cases, this study hoped to get a better understanding of the factors and conditions defined by the three research questions.

The third factor about this district that was significant to its selection for this study was its clear position on local assessments in relation to outsourced standardized tests and State accountability measures. The district has a clearly stated position on assessment that is part of board policy on high-stakes testing adopted in November of 2013. Specifically, the resolution emphasizes first that the district measures authentic student progress over outsourced standardized tests, and second that state testing will only be used for state mandated functions and no local decisions about students or staff.

By selecting a case where the “local” position on assessment is clearly defined, I was more able to observe the process of teacher sense making in and around that position. In this way, this district has taken a position that seems to acknowledge the tensions between the “prove” and “improve” (Vidovich, 2009) function of assessments and therefore served as a rare, if not unique, location for this study, where some of the institutional forces acting on teachers have been somewhat neutralized or at least normalized. Put another way, the “expected sense making activities ... shaped by broad cognitive, normative, and regulatory forces that derive from and are enforced by powerful actors” (Weick et al., 2005, p. 417) are kept more or less in check in Suburban Woods, thereby giving the researcher an opportunity to observe the issue and to describe the sense teachers are making as opposed to the sense they think they are supposed to be making. The potential limitation of this choice will be addressed in a later section of this chapter.

Introducing the Participants

In this section, I introduce the participants in each of the two school-based teams. In addition, I also describe each participant’s initial views about assessment and assessment practices in their work.

Table 2. Team Members and Goals Compared

	Team Goals	Role	Years Teaching	Years in Role
The Shore School				
Lily	Develop a social studies assessment that has a focus on student's ability to read and interpret multiple points of view. Expand the use of rubrics from ELA to social studies to measure growth in specific skills and dispositions. Increase teacher and student use of rubrics to direct learning and measure growth.	Classroom Teacher 4 th Grade	21	1 on 4 th Grade
Paula		Resource Teacher	20	1 on 4 th grade
Charles		Facilitator	10 Teaching 5 Administrator	3 as Principal
The Hill School				
Amy	Develop a social studies assessment that includes interdisciplinary work with Art that focuses on the relationship between beliefs and world view/perspective taking.	Classroom Teacher	28	1 on 4 th Grade
Tina		Art Teacher	12	2
Lynne		Facilitator	14	14

Site #1: The Shore School

Lily. Lily is a veteran teacher with 21 years at the Shore School, but she is new to the 4th grade. She has known Martin-Kniep, the designer and leader of the PADI project, for many years. She is the classroom teacher on the Shore School team and was looking forward to developing a project with a social studies focus. She sees the PADI project work as similar to curriculum development work done in the 1990s before the annual testing was implemented by the State. She said this when asked to talk about the connection between the district's new Shared Value Outcomes and the new PADI assessment: "They do, and they

definitely connect. Have we drawn the line between the dots? Not necessarily. But it's in the planning and in providing kids the opportunity to apply the thinking" (Interview #2: 8917-9096). How students think and how they apply what they have learned through thinking are actions that Lily values in good curriculum and assessment.

Lily described the Shore School PADI Project they developed in the June training:

Our objective was for kids to think about- in all aspects of life, reading watching TV, listening to stories to think about the fact that what they are, that the story that is before them is not the story in its entirety. It's just one perspective. So we used multiple resources- pictures, simulations, picture books to expose them to that kind of thinking and after they practiced it, they the generated questions to keep in mind as they come to any new source of information. Questions like: Who is telling the story? Whose voice is not heard? What gain could be got by the storyteller? How do they benefit? When does the story begin? Is it the whole—you know—start to finish? (SHORE SCHOOL Team Interview 2: 303-1040)

The project was designed to have several lessons that provided students with multiple points of view of the same event. Some examples were historical and focused on the Native American Unit; others were not based on curricular content, but were designed to help students understand that the same story could be told from multiple perspectives. One activity involved students reading multiple descriptions of a cafeteria food fight, for example. The final performance project for the students that was mentioned in the blueprint was described by the facilitator:

They're coming up with a guide that's a little looser than what I thought, but then they're going to actually have kids perform an analysis in front of parents and an authentic audience. They're going to actually show them using the guide. So the performance is going to be less about sharing a product and more about sharing a process. (SHORE SCHOOL Team Int 2: 6465-6801)

The goal would be to complete this later in the year and share it with an audience either of parents or other students in the school.

Paula. Paula, a resource support teacher on the grade level, was the non-classroom teacher on the Shore School team. Her motivation for joining the PADI project was her sense that it seemed like a way to bring back the best parts of her teaching before the state tests. A

veteran of 20 years in the school, she wanted to see more interdisciplinary work. She was, however, new to the 4th grade team but worked on a PADI project with 3rd grade the year before. Paula values teacher observation above all other assessments: “Mmm, so real true assessments, I’d say it’s lots of teacher observation, lots of you’d be sitting down one to one in a lesson ... I mean, how do I say it? It would be the answering of questions, labels, short things” (SHORE SCHOOL Team Interview 1: 17588-22012). She said she was very familiar with pre- and post-assessments as well as benchmarks in writing through the school’s use of TC Units of Study. And she said this of authenticity: “It means sitting up, pulling a chair up next to a kid and see what they do on a day-to-day basis, not on a like a memory test or talking to a kid “ (SHORE SCHOOL Team Interview 1: 6090-6236). She does not see current social studies assessments being used by many people. She feels supported by the district’s stance against standardized testing and does not have a strong opinion about the new report card. In the end, she concluded that the PADI project would be a way for her to work in ways that she and her colleagues used to do before the state testing and new curriculum calendars pushed it out.

Site #2: The Hill School

Amy. A veteran classroom teacher with 28 years of experience in the Suburban Woods district, Amy had been on a 3rd grade PADI team the previous year. She welcomed the opportunity to do another project when she moved to 4th grade this year. She saw the PADI projects as a return to the way she used to work. “And it also kind of was a hope for me to go back to way of teaching that I really embraced and I felt that we were getting away from that in education, so now I feel like it was a movement to getting back to something that I really believed in” (Amy Interview 1: 1528-1771). She saw, especially in areas like social studies, that content had taken a back seat to teaching skills. In addition, she described how teachers used to be able to blend or “layer” content and activities in multiple participants to enhance instruction.

Well, you know, content sometimes goes by the wayside now with the focus and emphasis on the reading and writing and math parts because of the assessments. So I ____ some fourth graders also assessed, so my fear was that social studies would kind of fall by the wayside if I, if we didn't bring it to the forefront. So part of my goal is to make sure that we cover the content in social studies and with the art component, it would be a layering, a layering activity so that the kids, you know, it really sticks with them. (Amy Interview 1: 2319-2841)

The "new curriculum" is more top-down, and the assessments are about skills. Amy described herself as having experience with performance assessment, "not like Giselle's," and that in recent years she felt as though she had lost her way in terms of its values in teaching. She reported to be assessing "everything all the time." She described her assessment of and with children as conversations, goal setting, and coaching.

I don't think they get a lot out of assessments that have a numeric value on it, except for maybe a ranking in the class, which (pause) I think that's probably all they get out of that kind of system at this age level.... And maybe that's what they get out of it at every grade level, you know, I don't know. But I think the more informal assessments I do allow me to have conversation opportunities with kids, where we talk about goal setting and what they, what I would like to see them accomplish and what they would like to accomplish, and then those conversations lead to little plans of action. Did you meet your goal? You know, whether it's academic or behavioral. And what did you do? What are the steps that you took and how can I help you? It might lead to extra help activities or enrichment activities. And so those small conversations where I actually coach children I feel are more beneficial. But that comes a lot out of, probably more out of the informal assessments that I do. (Amy Int 1: 11026-12029)

Amy mentioned the terms "pre-assessment and formative assessment" in the context of literacy units of study. But above, she also described the formative potential of assessment activities that involve students and teachers planning learning together through conversations and goal setting. She often used "we" language to describe performance but said that she did not have her own definition of "authentic."

I'm not sure, I'm not sure about that. I, I feel that when kids do projects that have a real outcome and a rubric is involved and they kind of know like it's, it kind of presents the goal for them and what they need to do to get there, and me being the person who coaches them to get wherever they need to go to or wherever they want to go, I think that's, I think that's more authentic than giving them an assessment, getting a score, putting it away, and then never

revisiting what they need to do. So and I think that's also what drives kids. I think they like that, that feeling of ownership and automaticity with PADI. So I don't know. I'll have to get back to you on that. (laughs) (Amy Int 1 12427-13114)

Amy stated with confidence that the curriculum is "solid." She talked about the new report card but claimed that she would not let it drive what she does. A standards-based report card helps her like a checklist, but she uses her cumulative sense of the child's work. The report card "makes reporting less subjective." At this point, she tends to look at student work on her own.

At this point in the project, Amy felt that the plan was clear and that they would start and end it early in the year. The PADI project in the Hill School was slightly different from the Shore School because it consisted of two parts united by a common essential question. First was an outdoor art activity and reflection where students created art using resources from the natural environment. The second element of the assessment would take place as a series of social studies lessons followed by and an "on demand" writing prompt given in class. At this point, the team was not certain about the final interdisciplinary performance project. In terms of the rest of the 4th grade team, Amy was hoping to invite them to see what the PADI team was doing. "I'm hoping to invite them to do it because it would be a great way to get everybody else involved in this. And I think that was our hope, that they would want to do it with us. If they don't, that's okay, we'll do it this year and we'll lock it, and then next year they'll want to do it" (Amy Int 1: 8127-8409)

Tina. As the art teacher at Hill School, Tina was relatively new to the district. In her second year as the K-5 art teacher, she had 12 years experience in her previous district. She pointed out the real differences between her old district and Suburban Woods in terms of the kind of work that is valued and the kind of curriculum that is created. In this description, she highlighted two types of assessment.

I came from a district that was a very skills-based district, that we were, our assessments were very much based on (pause) basic skills, that thinking was, you never really (pause) assessed.... [WE assessed] their end product, but

we (pause) in the end, all we've assessed skill over anything else. That was just part of that district's values assessment. [Suburban Woods] is very different from that. [Suburban Woods] is, I would almost say opposite of that, where the understanding comes first, and that aligns much better with like my own personal way of learning, my own personal way of teaching. (pause) ... PADI has affected how I assess, but I think that the district has as well because of just my environment and I mean, PADI comes along in my mind, and that's all connected (pause) where I was before would have never been a PADI school because we would [say that] we don't have time for this, we have to worry about testing and that kind of thing. (Tina Int 2: 5222-6331)

Her perception of the Hill School is that it is less innovative than other schools in Suburban Woods. But she emphasized her enthusiasm for creating a project with an interdisciplinary focus in her school, where this kind of innovation is less common.

I know [the other elementary schools] have had many PADI projects, they have culmination projects all the time, and I think this is our first go at it and only two people were interested in doing it. So I think the culture of that building, I don't know if Lynne could even expect, but it really is not a (pause) a passion, a respect, an understanding of interdisciplinary work among teachers in a way that I see in the other elementary school and the middle school. (Tina Int 2: 15426-15893)

But on the whole, the students are asked to think more and perform more than her old district, which was very skill-driven. She still felt new and not ready to share with all the teachers. She was excited to join the project and design assessments.

Data Collection

Like most case studies, data collection took many forms (Creswell, 2013; Merriam, 2009), including: site descriptions and district documents, assessment development materials, two sets of direct observations, two sets of participant interviews, and one round of focus group interviews with each team (see Table 3). The study's set of data sources and the order of their collection, as well as the decision to add the direct observation of teacher work sessions, came directly from the data use literature. Spillain (2011) and Warren-Little (2012) both warn against studies that rely simply on "ex situ accounts of practice," such as

surveys and interviews, to gain insight into teachers' use of data and assessments to inform practice.

Table 3. Overview of Information Needed

Contextual: To provide context and background	Specific information about the district(s) and the background to its goals around assessments. History of current local assessment process	Documents, administrative interviews, survey
Demographic	Descriptive information regarding participants: Gender, years experience, subjects, position in organization, role in PADI team.	Survey/Questionnaire as part of first interview.
Perceptual	Participants' general impressions about the initiative: its purpose, its supports, and its barriers.	Interviews
Research Question 1	How does the degree to which teachers design, create, and implement the performance assessments impact their sense of agency?	Documents, Participant Observations, interviews focus groups
Research Question 2	How does the degree to which their work is facilitated through training and administrative support impact their sense of agency? Collaboration, and communication.	Participant Observations, interviews focus groups
Research Question 3	How does the degree to which the work is part of local and state accountability structures impact their sense of agency?	Documents, Participant Observations, interviews focus groups

Although interviews, surveys, and self-report logs and diaries supply ex situ accounts of practice and point to salient dimensions of interaction and context, it seems unlikely that a robust understanding of practice can be achieved absent the strategic use of methods that capture the detail, nuance and patterning of social interaction. (Warren-Little, 2012, p. 146)

This lends credence to the idea that agency must be examined and understood both in individuals and in groups of individuals.

Data collection began during the summer of 2015 and followed the work cycle of approximately 6-8 months of local assessment development following the PADI phases that feature a timeline of training and support, which emphasizes the design, implementation, and refinement of an assessment cycle culminating in a performance assessment (see

Table 4). Documents were collected over the summer, and work sessions were observed into the fall and winter. In particular, the PADI training material and corresponding assessment materials produced by the teams were examined for background information and to inform subsequent interviews and observations. These included the “blueprint” documents and rubrics designed by each team that outlined the goals of the PADI unit and the standards by which the student work was assessed.

The non-participant observations followed the PADI timeline beginning with a formal training session in June/July of 2015 as well as a second session in November of 2015. The June session is where the teams first outlined their project and began to design pre-assessments and lesson plans that aligned with State and local standards. At this session, the team produced the first draft of the curriculum “blueprint.” The October session occurred after the team had given a pre-assessment and consisted of training to help the team refine their “blueprint” to meet the needs of the students based on what they had learned from the pre-assessments. These observations were conducted by creating field notes that followed the outline of the training session, which was usually organized in a PowerPoint handout. These observations gave some insight into the decisions teams were being asked to make based on their training.

Table 4. Timeline of Data Gathering

	June/July	August	Sept	Oct	Nov	Dec	Jan	Mar	April	May/June
Event	Team work with PADI trainers and facilitators	Teams work in their own districts		Team work with PADI trainers and facilitators	Teams work in their own districts				Performance assessments given in District Final PADI Showcase	
Doc. Data	Blue Print Draft			Blue Print Revised Pre-assessments and student samples					Performance Assessments completed Student Work Samples	
Observation Interview/ Focus Group	Observation #1 Interview #1 after the training	Interview 1 con't		Observation #2	Interview #2					Focus Groups

Ideas to help narrow the size of the study:

1. Select individuals from each team (6-9) based on criteria to create variability.
2. Limit to two observations.
3. Include entire group in each of 3 focus groups.

When examined along side the “blueprint,” patterns of decision making emerged. Again, these observations served to inform the subsequent interviews.

Following the pattern of data collection allowed for several member check opportunities (Cho & Trent, 2006; Marshall & Rossman, 2010) during the design, implementation, and refinement process. Following the assumption from sense making theory that sense making occurs both individually and in groups (Weick, 1995), it was important for the purposes of triangulation (Marshall & Rossman, 2010; Merriam, 2009) to see subjects at several points in the process and in multiple settings. For example, the documents and the observations were referenced during the interviews and focus groups. These individual interviews were followed by one round of individual member interviews and follow-up interviews as needed. Focus groups occurred toward the end of data collection.

Interview Protocols

The sense making interview “is designed such that the respondent is able to circle, or repeatedly engage with, the given phenomenon or situation” (Foreman-Wernet, 2003, p. 8, in Agarwal, 2012, p. 10). For this reason, I chose to conduct more frequent short rounds of interviews followed by a focus group session that would allow for the observation of changes over time in terms of how teachers were making sense of the assessment design, implementation, and refinement process (see Table 5). The interview questions for both interviews as well as the focus groups were drawn from a single set of semi-structured interview questions and follow-up questions that were also informed by my observation notes and the documents produced during the training sessions. The first set of interviews (10-15 minutes) took place after the June observation, and the second set took place after the October observations. The focus groups occurred after the final performance assessments in the winter/spring of 2016. For the semi-structured interviews and the focus

Table 5. Interview Questions

Demographics and Engagement		
Your Role (Verbs)		Your Role (Nouns)
Did you: Attend Training Sessions Design/Implement/administer/Refine		What is your position in the team (teacher, specialist, facilitator)
Research Questions	Interview Question Structure: Timeline Questions	Sense Making and Agency Follow-up Sequence
<p>What is the relationship between the teachers' engagement with the process of designing and implementing performance assessments and their sense of agency?</p> <p>What role does facilitation, support and training during the process of assessment design and implementation play in teachers sense of agency?</p> <p>What role do local and state accountability demands play in teachers' engagement in the assessment process and the development of their sense of agency?</p>	<p>Describe the work you have done since the summer Describe the work since the pre-assessment and October work session.</p> <p>How has the "Blueprint" structure supported your work? What role do rubrics play in this work? What role do standards play in this work?</p> <p>How do you plan to use the information that the assessments give you? With whom do you plan to use it?: Students/Colleagues/Administrators/Parents</p> <p>What is about this process of designing and implementing your own assessments is working for you? What is not? What has been easy? What has been difficult?</p> <p>What have you learned by making and implementing the pre-assessment? Have you had a chance to share the student work with: Students/Colleagues/Administrators/Parents What was their reaction?</p> <p>What other concerns or issues has this assessment process raised in you mind? How does this work connect (or not) to your districts work with state assessment requirements?</p> <p>How does this assessment fit in with other assessments at school? Where do the other assessments come from? (you, school/department, state, other)</p> <p>What changes have you noticed in the students through watching them take on this assessment? How has the experience of designing and implementing your own assessments influenced your work?</p> <p>What are the reactions of others in the school: Other Teachers Administrators Parent and community.</p>	<p><i>These questions may follow any of the questions as needed:</i></p> <p>Is this the work that you feel you should be doing?</p> <p>Is PADI shaping your thinking about what you should be doing? (How? Challenging/Changing/ supporting your thinking?)</p> <p>Do you feel you can do this work?</p> <p>Do you feel your team/grade/school can do this work?</p> <p>What is helping you or not helping you?</p>

group questions, I followed the methodology laid out by sense making scholars. In particular, sense making “mandates the framing of research questions such that the respondent is free to name his or her own world rather than the researcher to define the phenomenon in question” (Foreman-Willmet, 2003, p. 8, in Agarwal, 2012, p. 10). Two specific approaches achieved this. Both were intended to focus on the interview describing their actions in the past, present, and future and, in so doing, naming their experiences. For example, to ask directly about “agency” would risk missing the point of the study, which was to understand how teachers make sense of their agency. The second specific approach also involved time, often referred to as “micro-moment time line” (Foreman-Willmet, 2003, p. 241), where interviewees were asked to describe a situation in timeline steps. Through this, the subject revealed not only what they did, what they were doing, and what they planned to do next, but also their sense of agency—what they were able to do, what they are able to do, and what they are able to do next. These questions are reflected in the “Timeline” column of Table 5. Beyond factual timeline questions, interpretation questions should be open-ended enough to allow the subject to name their experience. These follow-up questions can be in the form of: “What helped? What hindered? What are the barriers? What do you conclude? What emotions/feelings relate? What would help? What things need to be discussed here that aren’t being discussed? Whose voice needs to be heard that is not being heard?” (Dervin, 1998, p. 44). These questions are reflected in the “Sense making and Agency Follow-up” column of Table 5. In constructing questions in the manner described as “micro-moment timeline,” a sense making interview reflects the “chordial triad” described by agency theorists Biesta and Tedder (2006). Biesta and Tedder place the concept of agency into some methodological considerations. As such, the interviews attempt to document agency as it is perceived and enacted in particular situations. The final round of interviews created another member check opportunity that emphasized the study’s desire

to understand examples of collective agency in order to see if there were developments in the collective agency of various PADI teams and their schools.

Data Analysis

Merriam (2009) suggests that data analysis begins immediately and continues throughout data collection. My sequential analysis of data followed the general processes described in Bogdan and Biklen (2007), as well as Creswell (2013). I read through the initial materials and first interviews that took place just after the PADI training session in June/July. From there, I developed initial descriptive codes to help me see the structure of the observations and interviews. This process is itself one of sense making. “During data analysis, there is a search for patterns in terms of processes or verbs rather than things or nouns” (Agarwal, 2012, p. 11). As such, my coding reflects descriptions in the form of actions and processes as well as nouns. For example, instead of looking for “change” in a team’s plan, an observer would record discussions of teachers “changing” or “modifying” their work. The agency or sense of agency lies in the action—the ability to take action or not—not in the event itself.

The process of initial read-through of interviews and subsequent coding occurred throughout the case study time period. In addition, ongoing analysis took the form of journaling and memo writing. I wrote memos pertaining to the code counts surrounding my research questions, an analysis of themes outlined by the literature, and an individual running record of themes and quotations from each subject on each team. For Research Questions 1 and 2, the lens of Data Use and professional accountability came to the fore. In particular, I was influenced by the three elements of Coburn and Turner’s (2012) “meaningful outcomes”: “(a) outcomes related to student learning; (b) those related to changes in teacher and administrative practice; and, (c) those related to organizational or

systemic change” (p. 47). In addition, I was influenced by Aneesh and Darling-Hammond (1996) and their idea of professional accountability that strives

to induce change by building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches. (p. 57)

For Research Questions 3 and 4, the lens of Agency and “agency orientation” played an important role. In particular, my working definition from Chapter I:

a teacher’s sense or belief that they have the ability and the authority to make changes in their practice based on their own judgment. In the case of assessment driven reform, this would mean the sense that teachers have the ability and authority to use assessment data to make instructional decisions and change their curriculum.

In addition, I focused on Birdwell-Mitchell’s (2015) framework, which examines ways that agency can lead to teachers both supporting and resisting change. Finally, Sense Making Theory also steered the analysis.

I wrote an extended memo—a mid-case review—after the second round of interviews. At that point, I had made three site observations of training PADI sessions as well as two sets of interviews of both my case teams. This mid-case review provided an opportunity for reflection before the teams would embark on their final projects (followed in interview 3) and then a final member check reflection (in interview 4). I agree with Miles and Huberman (1994) that some degree of perspective taking was important at the mid-point of a case. The review memo attempted to summarize my data collection to date and had three goals. First, it presented three data sources in narrative and chronological form for the first time. This rich description helped to illustrate the issues revealed by each source as well as begin to reflect some of the initial coding from my notes. Second, it labeled examples of resonance and examples of dissonance between the data sources. Marshall and Rossman (2010) and Merriam (2009) suggest that triangulation of data is a key step in understanding a case and in developing validity. Since this work was still “mid-case,” I

hesitated to draw conclusions, but I tried to outline patterns. These patterns marked the start of more sophisticated analytical coding, were a way ensure that my data collection was aligned with my questions (see Table 3), as well as helped me raise questions for subsequent interviews, member checks, and data collection.

The mid-case review helped me set out the pattern of analysis described by Maltbia (2013) to assist the researcher to move through a 4-stage process from “data to information to knowledge to insight.” In the pattern of analysis, the researcher passes through the data as many as seven times, each time narrowing and defining codes, code clusters, patterns, and themes.

It was during this process that I modified my original coding system to help analyze documents and interviews. The two major strands of initial codes are from the Data Use literature (Coburn & Turner, 2012; Jennings, 2012; Marsh, 2012; Supovitz, 2012; Weiss, 2012), as well as from Sense making theory (Coburn, 2006; Coburn & Talbert, 2006; Spillain, 2012; Weick, 1995) (see Table 6).

I made these changes because the initial codes did not reflect the instructional choices and students learning outcome I was hearing in the interviews. The new code reflected both language from the PADI training sessions as well as the kinds of talk I was hearing during the interviews. Keeping in mind the sense making notion that subjects should name their experiences, the new codes reflect the language of instruction that the teachers were using in the interviews.

Table 6. Initial Codes Compared to Revised Codes and Sub-codes

General Codes	
Initial Codes	Revised Codes
Agency Power and Politics Pressure Source of Pressure Feeling Pressure Trust (Low to High)	Agency <ul style="list-style-type: none"> • Choice/Options • Decision to Act • Making Changes to Curriculum • Making Changes to Practice Outcomes <ul style="list-style-type: none"> • Instructional Change • Organizational Change • Student Learning • Teacher Practice Barriers to Action Support/Facilitation Pressure Trust Collaboration on Team Collaboration with Organization
Assessment Codes	
Initial Codes	Revised Codes
Feedback Purpose <ul style="list-style-type: none"> ▪ Grading ▪ Formative ▪ Feedback Cycle (L to H) Assessment Purpose <ul style="list-style-type: none"> ▪ Lens ▪ diagnosis ▪ compass ▪ monitoring ▪ legitimizer ▪ Rubric Use 	Assessment/Data <ul style="list-style-type: none"> • Making/Designing • Rubrics • Standards • Using to inform Students • Using to inform Teacher
Sense making Codes	
Initial Codes	Revised Codes
Ambiguity Retrospection Interactive Endeavor Signaling Negotiating "Who is my audience?" "Good Enough" Thinking	Sense Making <ul style="list-style-type: none"> • "Good Enough" Thinking • Negotiating • "I Think Statement" • "I Know Statement" • Ambiguity • Decision Making • Retrospection

Coding was being entered, organized, sorted, and analyzed using Dedoose software. The new codes were used for the purpose of deeper analysis and pattern construction (Creswell, 2013; Maltbia, 2013).

In addition, I made certain to quantify certain code applications and co-occurrences that related directly to my four research questions (see Table 7).

Table 7. Code Use and Code Co-occurrence

Research Question	Quantified Codes (# applications)	Co-Occurrence Codes (# co-occurrences)
How do teachers who participate in the work of creating local performance assessments both as individuals and as members of a group use the data produced from these assessments to inform their practice or change their curriculum?	Assessment: to inform Teachers/Students(42)	Outcomes: (18) Teacher Practice Instructional Change Agency: (4) Changes to Curriculum Changes to Practice
How does this process influence their understanding of the role(s) that assessments play in their work? What changes do they make in their uses of assessment and in their curriculum?	Assessment: Making/Designing (32)	Outcomes: (15) Student Learning Teacher Practice Instructional Change Agency: (8) Changes to Curriculum Changes to Practice
What factors support or limit teachers' participation in the creation of local assessments?	Barriers (77) Support (42)	Assessment: Making/Designing (9)
What factors support or limit the changes that teachers make in their instructional practices or their curriculum?	Barriers (77) Support (42)	Outcomes: (8) Student Learning Teacher Practice Instructional Change Agency: (13) Changes to Curriculum Changes to Practice

The quantification process revealed patterns that lined up significantly with all four research questions. In particular, it helped to reveal how and when “data” was being used as well as how and when it was not being used. This further demonstrated patterns in design and instruction decisions. Finally, it shed light onto areas where teachers discussed their

sense of agency and led me to understand where and when a sense of agency was achieved and where and when it was not.

In all, the following themes emerged. “Data Use” for teachers is not a linear progression, but follows a more organic process related to curriculum, practice, outcomes, and design. The process of making and designing is also closely linked to outcomes and to agency. Barriers and support are closely intertwined with time, colleagues, and culture. Finding these themes prompted me to organize my notes around these themes and gather quotations and excerpts that illustrate the patterns in themes in chronological and therefore narrative form. My analysis process concluded with some initial insights organized around each research question, which, in turn, helped shape the structure of the Findings chapter.

Pilot Work

My research into the development and use of local assessments began several years ago and has been the focus of most of my research methods classes. So in the broadest sense, I have conducted a variety of informal pilot observations, interviews, and focus groups.

One important lesson from this early pilot research was my decision to seek a case study outside of my own district. As an administrator, it would be difficult to interview teachers who work in my building.

In terms of pilot studies with the PADI districts in particular, I have a great deal of observation experience, having followed PADI for close to two years. I am familiar and have helped facilitate the training sessions, including the development of blueprints and rubrics. I have many samples of PADI materials, observation notes, and blueprints to draw from and have included one pilot study in Appendix E.

Validity and Limitations

One goal during my pilot studies was to be very open about this interconnectedness of assessment and accountability with all participants. By conducting the pilot openly, I hoped both to build trust among my participants and to encourage them to want to ask and answer more questions about their use of assessment as they work to improve their practice.

Giving subjects the opportunity to reflect on their sense making or how they are making sense of their work is one way this study may benefit the districts being studied. This kind of transformational validity (Cho & Trent, 2006; Marshall & Rossman, 2010) was one of my findings in all pilot interviews and an important aspect of my decision to embark on a qualitative study. One way I want to judge the validity of this work is whether or not the teachers in the study perceive a benefit in participating in the study. It is my belief that the act of reflection on practice can be somewhat transformational (Cho & Trent, 2006; Marshall & Rossman, 2010) for teachers.

In terms of my limitations based on my positionality or subjectivity, as per the work of Peshkin (1988), I believe the two most important would be my subjectivity as an experienced classroom teacher in conjunction with my role as an administrator. I need to recognize that in conducting research, I am myself acting within the process of sense making. Therefore, I must acknowledge that my own history and my own context would influence that work. As a teacher, I understand the challenges and frustrations that high-stakes testing brings to a teacher's professional life. I do recall feeling my time, my choices—my professionalism—"constrained" by state testing requirements where I felt forced between choosing to prepare students for tests and choosing to teach and assess the local curriculum as I understood it. At the same time, as an administrator, I believe high expectations in the form of standards can both challenge and empower teachers to improve their practice. I also believe teachers should be held accountable for their performance in

helping children learn. Beyond these first two subjectivities, I must acknowledge subjectivity as a researcher who is also a school principal. Agee (2009) warns of this relationship challenge, as I attempt to conduct research into the nature of accountability when I play a role in that accountability structure. This is an important consideration in my decision to look outside my own district my study.

Even though my study occurred outside my home district, I was still aware that I could not completely eliminate my subjectivity as a school leader. My subjects may still have perceived me as a school leader in the Tri-States Consortium. Some, but not all, of my subjects knew me from other Tri-States Consortium functions. Some, but not all, knew me from my work observing previous PADI cohorts. It was important with my subjects in PADI cohort 4 that I was clear about defining my role in PADI as an observer, not as a trainer and not as an evaluator of their work. During my pilot interviews, I was clear to mention these distinctions before each interview.

Limitations that potentially stem from the nature of the case study and methodology are two-fold. First, there are potential limitations due to the nature of the sample selected. By selecting only one school district for this study, its generalizability to schools and districts in different circumstances is limited. In particular, the district studied is able to position itself with regard to standardized testing and State accountability mandates because it has the economic means to do so. As an affluent district, it, along with many other PADI districts in suburban New York, is less dependent on State funding and has therefore been able to maintain a certain degree of independence from the current accountability structures in New York.

Second, this study may be limited based on the inherent limitations of sense making itself as an observable phenomenon. The fact should never be forgotten that a teacher's sense making may be shaped by "expected sense making activities ... behavior is shaped by broad cognitive, normative, and regulatory forces that derive from and are enforced by

powerful actors” (Weick et al., 2005. p. 417). Especially with regard to creating and interpreting assessments, it may be difficult to separate the sense making decisions being made in the best interests of the student and the sense making being made in the best interest of the teacher. And yet, it is perhaps this very confusion that this study is trying to better understand for practitioners, policymakers, and scholars alike.

Conclusions

It was my intention that the methodology I have chosen follow in the footsteps of other data-use researchers. As such, my approach helped address the potential limitations of the study. Triangulating multiple data points, member checks, and an awareness of my own subjectivities as a person engaged in an act of sense making, combined with my desire to give voice to practitioners engaged in a genuine effort to reform their practice, contributed to the validity and significance of my findings. The notion of “reciprocity” was also central to the validity of my research. It is my hope that this research project will help the teachers and the district understand how the implementation of local assessments is impacting teachers and students. The work must inform several audiences: first, teachers in schools; second, administrative colleagues; and third, scholarship. If successful, this study will open the door to future research in the districts I am working with. In districts that are “talking the talk” about assessment for learning, I hope this study will go beyond the rhetoric of assessment in order to examine and support the values and behaviors at the level of practice.

Chapter IV

FINDINGS

Observing data use by teachers and its impact on outcomes is a complex task, especially when understood in the context of the current accountability environment. This case study explored the lived experiences of two teams of elementary school teachers as they were trained, as they developed, and as they administered performance assessments locally over the course of one school year.

In this chapter, I present the key findings from my observations of the Performance Assessment Design Initiative (PADI) training sessions as well as my four sets of interviews conducted prior to, during, and after the school year where the assessments were developed and administered. The chapter is structured to follow findings as they relate to my four original research questions. For each question, the primary data source is the participant interviews. These are triangulated by direct observations, some documents, and the facilitator interviews. I will discuss the ways in which the teams did and did not use the data from the assessments they produced locally throughout the yearlong project. From there, I describe how the work did and did not influence their views on assessment as well as changes they made to their curriculum plans. This will include instances of variation among teams and individual teachers on teams. Finally, I will describe factors that influenced their choices and are seen by the participants as both barriers and supports for their work.

Research Questions Introduction: Research Focus Changes from the Original Plan

It is important to note at this point in the study that my observations and data collection did not follow the linear approach predicted. The original structure of the four research questions and research plan, as well as the plan laid out by the PADI training described in Chapter III, predicted a linear pattern of assessment creation over the summer of 2015, implementation of a pre-assessment in November, followed by data analysis and adjustments to the final performance assessment. The final or post-assessment would be given later in the spring of 2016, followed by more data analysis. After completing the training, designing, and implementing of pre-assessments in mid-November 2015, however, neither team ever actually created the final performance assessment as they had planned during the summer training and described in their blueprints. When this occurred, I consulted with Giselle Martin-Kniep. She said that each year in the four years of PADI projects, a small number (less than 20%) do not create fully complete or entirely original performance assessments. The majority of PADI teams do complete well-articulated performance assessments. Martin-Kniep also said of teams that do not create an original assessment task, while this complicates the teachers' task of assessing growth in the areas defined by their blueprints, as long as the students have the opportunity to "perform" or apply the skills in a task, the work of assessing growth with a rubric can move forward.

In place of creating new performance assessments, both teams extended the classroom instruction and activities beyond the initial social studies units and into subsequent units and then either adapted existing projects or repeated activities from the fall to stand in as the final performance assessment. The table below will assist in painting a picture of the difference between the intended goals of each team and what they actually did, I refer the reader back to Table 1 on page 68.

Table 8. Team Original Plans and Modifications Summarized

	Shore School	Hill School
Goals as described by each team	Develop a social studies assessment that has a focus on student's ability to read and interpret multiple points of view. Expand the use of rubrics from ELA to social studies to measure growth in specific skills and dispositions. Increase teacher and student use of rubrics to direct learning and measure growth.	Develop a social studies assessment that includes interdisciplinary work with Art that focuses on the relationship between beliefs and world view/perspective taking.
Essential Question	Where does the Truth lie? "How do I determine my own version of a story that has multiple perspectives?"	How do beliefs impact your view of the world?
New Plan	Overlay the task of evaluating multiple points of view and evaluating source materials onto an existing Ellis Island Role-play activity. Students will adopt the persona of an immigrant then research that person's experience coming to America through Ellis Island. Along the way, they will be asked to reflect on the multiple perspectives of the immigrant experience and well as evaluate the perspective of multiple authors.	Teams of students will repeat the activity of creating art from natural resources. Then they will be asked to write a reflective essay describing how this experience relates to what they have learned about Native Americans beliefs about their environment. Students will also reflect their thinking through video interviews as they prepare to write.

The Shore School team did the following: They decided that the work done in the Native American Unit was very engaging for the students, so they planned to continue to teach the concepts—namely, that all stories in history can be told from multiple points of view—throughout the year with an eye to creating the final performance assessment in the spring. By spring, they decided not to create the original performance assessment described in the blueprint—the media guide. Instead, they combined the assessment rubric developed in the fall with an existing performance activity on immigration. In effect, they adapted an existing performance project to stand in place of the performance assessment planned the previous summer and outlined by the blueprint (Appendix C).

The Hill School team followed a slightly different path. If you recall from the blueprint, this team planned to do the entire PADI unit in the fall and planned the performance assessment after the November Training. On this team, both teachers had

given both a pre-assessment and post-assessment activity in their individual classes before designing a final performance assessment. In other words, their original plan in June had them completing most of the PADI unit separately (pre- and post-assessments) prior to November. Recall that they had never developed their final interdisciplinary performance assessment during the summer training session, only the post-assessment for the art and the social studies individually. After the November training session, they made two decisions. First, Amy planned to continue to teach the concepts of the social studies unit—namely, that one’s environment impacts one’s beliefs and culture—throughout the year. Tina concluded that the first art post-assessment activity—creating art from objects found in the natural environment—was such a formative experience for her class that she should repeat the activity in the spring to see how the children had grown over the course of the year. In her explanation, repeating the activity would in effect treat that fall post-assessment as the a pre-assessment to be compared with a new post assessment in the spring. Tina and Amy, with Lynne’s support, decided to give this second spring post-assessment together and treat it as the final interdisciplinary performance for the PADI project that would assess both art and social studies standards outlined in the original blueprint (Appendix D).

In sum, while both teams did not implement the post-assessments as initially designed described in Table 1, by the spring of 2016, both teams did create or adapt final performances for the students that they could use as post-assessments to measure growth. In turn, both teams did collect data based on those final performances that aligned with standards described in their blueprints. So, while the work of both teams did not follow a simple linear pattern from assessment creation, implementation, and analysis, the two teams did each of these actions in different ways throughout the year. As a consequence, the ways the teams approached, then adapted and at times rejected the initial PADI plans became part of the processes I observed, and this became part of an altered research plan.

Research Question #1

How do teachers use the data produced from these assessments to inform their practice or change their curriculum?

When talking about data use through the code “assessment to inform teaching” (over 40 occurrences), less than half of those occurrences mention the data from either the fall pre-assessments or spring post-assessments as outlined by PADI, and a smaller fraction still (fewer than 10) specifically mention how they went about using the data from the same planned PADI assessments. This section explores three ways that teachers answered questions designed with sensemaking methodology. In those answers, they talked about using data in their interviews that are also reinforced by other data including observations by both facilitators. First, the teachers on both teams used data derived from observations made during the fall classroom activities combined with the fall pre-assessment to make adjustments to the unit during the year and made plans to change the unit in the future. Second, teams tended to view and discuss observation data from assessments and classroom work in order to describe student understanding, student dispositions, and student growth both during the process and at the end of the final project. Third, the teams were not observed scoring or using data in ways assumed by the PADI process and rarely discussed data in this way.

Finding One

Both teams used the data they collected through classroom observations and through the pre-assessment process to make adjustments to the PADI units during the year and made plans to change their units and assessments in the future.

In the Shore School, through administering the pre-assessment and through class observations in the fall, the team concluded that they had to teach into the task more in order for the student work to reach the goal they set out in the blueprint. They also concluded at the November PADI training session that they wanted to push the final performance assessment back until later in the year. By looking at “where does the truth lie”

in more social studies units, they hoped to deepen student understanding and skill and then enable the students to actually create the final performance. As Lily put it as she reflected on the November Training session,

No, in the summer we thought that we would have concluded the summer at the end of the Native American Unit. The truth is that we decided that they would have more—if they were actually going to do the project—which is to create almost a user's guide for media or text that the more experiences they had the better it would be. So we decided to extend the project through if not the end of Am Rev right through all the way until the end of the year and make the project the sum total of all they have learned. (SHORE SCHOOL Team Int 2, 2766-3278)

The team was excited to continue the lessons and activities on multiple perspectives beyond one social studies unit. As the year unfolded, the team added the work from the originally planned Native American unit to the Colonial America and Immigration units as well. This work brought the team to the end of May 2016, with no plan to create the media guide. They eventually decided to modify an existing Immigration activity to create the final performance assessment. The team described their final change to the final assessment project in the spring as follows:

The kids had to take on a persona. They had to become an immigrant, so they chose a country of origin at the beginning of the study, research the push and pull factors for those immigrants, and then actually created an immigrant persona that they embodied on Immigration Day ... so they planned, they talked, they wrote letters to their family's home, they tried to engage um a photo that would capture without words their perceptions, and they had to write an oral history.... They then came to Ellis Island where we had parents and faculty members recreate the experience as much as we could of an immigrant passing through Ellis Island. (SHORE SCHOOL Team Int 3: 292-1011)

In short, the team did not create an assessment described in the blueprint document as a media guide. In its place they combined the final reflection post-assessment writing activity with a pre-existing culminating activity on immigration. The facilitator observed the following about the process and the team's use of data:

I think there was deviation from the plan and I think that closing of the assessment loop is always the challenge with PADI and I think that's in this first year what they struggled with the ___, you know, understanding how to connect the performance to an actual post assessment that drew a comparison to a pre-assessment. (F Int 3: 2029-2358)

Here, Charles was suggesting that he had seen this kind of deviation in other PADI projects. The team seemed to struggle with the final assessment more than the activities and the lessons that led up to it. It was as though Charles knew the kind of trouble a team could get into, but did not use this knowledge to change the team's approach to the work—a decision he would explain later in the process as wanting the team to discover what was needed instead of telling them how to approach the work.

In terms of the performance, there was a dramatic piece to it, you know. They were really playing the role, and on the day um, you know, some of the stations that were set up that day, they needed to have a reflection um and an assignment based on their roles, so I think that, that idea of perspective and multiple narrative was designed into the experience in a way that was different from before. (3484-3883)

Here, Charles seemed to be explaining the team's decision to fold the theme of the PADI unit into the existing Immigration activity. He did emphasize that this new element did change the immigration activity.

But I think the post, really seeing, okay, what was the impact of exploring this whole year and then of the actual performance. I guess the issue is it didn't—because it was a thread throughout the year, which my teams tend to do and I, I don't discourage that, but it makes it harder to measure the impact of the performance piece. (Shills Facilitator Int 3: 3895-4228)

But here, he also admitted that their decision to extend the project combined with their decision to connect the post-assessment to an existing project was problematic from the perspective of measuring student growth over time. The facilitator's summary of his team's data use as it connected to assessment is three-fold. First, he had a clear understanding of what they did to create assessments and collect data from those assessments to measure student growth in specific standards. Second, he understood how and perhaps why they struggled with this work in the first year—something he called “struggling to close the

assessment loop.” Third, he had a sense of what they would need to do the following year to apply what they had learned to the work going forward. Through these three, it is evident that Charles saw the Shore School work as a process that extended beyond a single school year. This understanding is a difference between the two schools that will be seen when Charles’s thinking is compared to that of the facilitator at the Hill School.

The team from Hill School also used their observations to adjust their projects. When asked about the steps they took, like the Shore School Team, the members of the team described what they understood and what they decided to do. Similar to Shore School, they realized that they wanted to extend the projects. First, Amy planned to add activities and lessons on belief to subsequent social studies units. Second, the post assessment art project in the fall would actually serve as a pre-assessment if they repeated the activity in the spring. Combining the extension of both projects would add additional “layers” of practice and thinking to help the students take the work further. The art teacher described the process as follows:

But what we discovered after running the pre and post was that really what they needed to do was to go back again, that there were too many questions, ... I guess the pre didn’t (pause) teach them enough so that their post really felt like a pre. It felt like they were kind of experiencing things for the first time so it ended up that the pre was maybe just like a (pause) ordinary activity, and what we thought was the post became the pre and that we’re going to go back, and after they reflect with Amy and they do this writing and they look at this from the lens of social studies, they’re going to come back and we’re going to do it again with a whole, I guess, more expanded mindset. (Tina Int 2: 1558-2268)

Lynne, the facilitator, said it this way:

And we talked about how things went with students. And it was decided that we would use that as the pre-assessment and then Tina would incorporate another component where the students would do something very similar in the spring....

And we’re hoping now that these same children who were involved in the environmental art project would look at the resources a little bit differently here in our backyard,... But now we’re looking to see if there’s anything that

the children learned about in the Native American unit that they're going to go to capture and use in their own philosophy about how where we live affects how we live. (GH Fac 2: 1010-2692)

In sum, the Hill School team also delayed and extended the work. In the case of Tina, the art teacher, she repeated the activity to delay the final assessment. In the case of Amy, the social studies teacher, she extended the work into other units. Lynne's plan in the spring was to facilitate the teachers working together with the art class. Compared to the Shore School team, the Hill team was primarily focused on simply making the interdisciplinary project work rather than assessing the growth of the students' learning.

Finding Two

While teachers did not use data in the ways expected, both teams discussed observation data from assessments and classroom work in order to describe student understanding, student dispositions, and student growth both during the process and at the end of the final project.

The “data” they created and acted on were based on observations from their lessons and closely tied to language of teaching and planning and not to assessment rubrics. For example, the teachers from the Shore School described the work in the fall this way:

But in any case, they then had to have those questions in their minds as they entered the Native American's study for a social studies unit in the fall. And they applied those questions throughout which was really, really powerful. Even when we had a speaker come in to the school. Teepee Ted had lived with Native Americans out west, and as he was telling the stories to the children, they came back to the classroom talking about his angle and what he wanted us to take away from the stories. So it was a different way of thinking. (SHORE SCHOOL Team Int 2: 1040-1581)

In this case, the “data” observed were the student behaviors—their responses and their talk. The team concluded in November 2015 that some students were able to understand and apply the concept of “angle” and point of view in non-fiction writing. The students were beginning to understand that writers and presenters have an “angle” and an intention when telling a story. And they saw that some students could ask questions in novel settings. The team said things like, “So they are in it thick right now. They are really

into who is telling the story. We use books like *Split History*. So it is the two versions of the Boston Massacre or the Boston Tea Party, etc.” (SHORE SCHOOL Team Int 2: 1838-2031).

These materials were gathered with the help of the facilitator, Charles, after the pre-assessments in combination with their observations of the students doing class activities.

And Paula added,

I do see them thinking in terms of perspective more than I would have ever thought possible. So when they hear a story, they do not just hear a story. They think of, they think about the perspective. And they think about what is not told or what has been left out. (Team Int. 2: 6356-6623)

Here, Paula was making a general observation based on her observations and conversation with students in the classroom, during activities as much as during an assessment. Again, this relates to her notion that assessment happens when you “sit close” to a student and talk and listen. At the end of the year, Lily shared what she saw in the following ways:

Well, you know what was interesting? I think the biggest Aha! that we got was that the study that we were sort of undertaking, where does truth lie, is more difficult, it’s more of an anecdotal kind of study than it is a quantitative study, so their responses showed depth and growth, but it wasn’t something—and we tried to plug them into the rubric, but it wasn’t clear they internalized it completely and applied it. So we saw sporadic applications, which was definitely growth from the fall where there were none, but we kind of concluded that it was—it almost goes against their nature as children because they want to believe what they’re told, they’re taught to believe what they’re told, so that sort of cynic or cynicism is, is, goes against the grain. Definitely the higher-level thinking kids were more apt to do it, but we saw definite glimmers in, in the stronger students and, and growth just in general, but it was, it was interesting, definitely interesting to see. (SHORE SCHOOL Team Int 4: 1491-2473)

Here, the team’s observations can be summed up as being focused on student actions in class and some samples of student writing. The team’s data use is largely “anecdotal,” not “quantitative,” but it is guided by predetermined standards and rubrics. Their comments represent their evolving view of their project’s initial goals and their understanding of the standards they selected to measure those goals. They were describing what the growth of their students looks like in those standards. This approach is similar to what the team

described about their assessment practice in the first interviews. Words like “anecdotal,” “sporadic,” and “glimmers” are important here. They mentioned student growth in the actions of the students, but had not placed it onto a rubric or continuum, even though the team had designed both in their PADI planning process, and that process assumed that they would be able to do so immediately.

After the spring assessment, the Shore School observed, “Yeah, last year they didn’t reflect on it too much, but this year they were reflecting along the way.... But they were, they were much more mindful of what they were doing” (Team Int 3: 4895-5201). When I asked specifically what they had observed and how they had made the observations during and after the performance, they said,

In the writing, they incorporated things that ____ crafting(?), you know, like the experiences in the homeland that would, that would make them leave more. When they were acting like the immigrants, they came in (pause) dirty and tired and haggard, whereas last year they came in like they were playing dress-up. I don’t know, there was a little bit authenticity, I think, to what they—to the person they were on Immigration Day. (SHORE SCHOOL Team Int 3: 2925-3355)

In sum, the team from Shore School was able to point to multiple examples where they observed changes in student thinking, writing, and talk as they related to their essential questions throughout the year. They interpreted these observations as growth based on the year’s work. They also noted some increase in the students’ use of rubrics later in the year after the final performance, which was one of Charles’s long-term goals for the school. The key finding here is that most of the observations they reported, including the increased student use of rubrics, occurred during lessons and assessment activities not based on their evaluation of the final assessments such as student reflective writing. The teachers at Shore School spent their first year focused mostly on classroom activities, not on the summative assessments.

Hill School also observed the students in ways similar to the Shore School team based on classroom observations combined with pre-assessment work. For example, when asked about her part of the project, Amy, the classroom teacher, said,

... So when I gave my pre-assessment, we did have quite a conversation on what we meant by beliefs.... I think the kids (pause) what they, what we see and do, I think they, they saw the scene wasn't a perspective as much as a literal thing, so it—the question itself required some teaching into.... So kids were able to make those connections and they did really a great job of it, I think. They did not end up writing a persuasive essay, which is what the intent was, but it became more of an opinion piece. (Amy 2: 982-3134)

Here, Amy was making several conclusions about how her students were thinking in class about the pre-assessment activity. She observed her students thinking about beliefs and concluded that the students were still confusing perspectives with the literal. These conclusions caused her to change some of her teaching during the year. In turn, she described seeing a different outcome (“They did a great job of it”) after the post-assessment writing in June 2016. Like the team in Shore School, Amy was assessing her students in the context of classroom activities prior to written assessments. She went on to describe her overall assessment of their understanding:

But I feel like they, I think they got, many of them really got it. What I'm thinking about and what I was considering now is asking those same questions now after we've studied the colonial period, and then how did what, how did what the colonists see and how did their beliefs affect what they saw and what they did? (Amy 2: 3258-3508)

Amy's assessment and corresponding rubric were heavily focused on the persuasive writing process, but the dispositional goal of students 'understanding and expressing of beliefs was not covered by the assessment. She seemed, instead, to rely on her observations of students in her lessons. Here, Amy concluded that her students understood the relationship between beliefs and perspective in November, so she began to think of doing the same kind of lessons and assessment in the next social studies unit, hoping to capture more evidence of this emerging understanding.

Tina, the art teacher in Hill School, described the students' first assessment in November as follows. By way of context, the original plan in the summer was for Tina to do a pre-assessment activity in early September followed by a post-assessment in October.

But what we discovered after running the pre [assessment] was that really what they needed to do was to go back again.... It felt like they were kind of experiencing things for the first time so it ended up that the pre was maybe just like a (pause) ordinary activity, and what we thought was the post[assessment] became the pre[assessment]. (Tina Interview 2: 1558-2017)

Her evaluation of the students was based more on what she observed in class rather than her written assessment or her rubric. Like Tina, Lynne, the facilitator in Hill School, concluded that the teachers had not observed enough to conclude that the students had made the interdisciplinary connection between the art project and the social studies unit as the team had hoped they would. The hope was that the work on the art project, which involved gathering resources in nature to make art, would impact the students' thinking about how Native Americans used resources. This connection between the two participants harkens back to Amy's idea of "layering" curriculum to help the students push deeper into a concept. This is something that neither teacher seemed to observe in the fall, but informed what they were looking for in the spring.

Like the team in Shore School, the Hill School team did not create a final performance project as described by the PADI process the previous June. In its place, the team did work together to repeat the art teacher's original post-assessment activity tried before November. In other words, the art teacher followed through on her plans from the November training sessions, where she repeated the original project from the fall in the spring, and she worked with Amy to adapt this art experience into an interdisciplinary experience that assessed both the art and the social studies standards outlined in the team's original blueprint. In the final art experience, the art teacher observed, "They were really reflective of what they were doing and be able to make connections between um, you know, historical groups and the things that they were do—it was, it was exactly what we had

hoped that we'd see and felt like we weren't getting in the fall" (Tina 3: 569-826). When asked specifically how she saw this, she said,

... the language they were using. They were calling things resources and, we actually were going with this kind of videotape interviewing them about things that were happening, and there were all these really interesting things that they were coming up against. Like "I don't have enough, you know, this or that resource," and we would say, "Okay, well, what are you doing? You don't have resources." You know, they were able to problem solve, like, "Well, I can, I can move, I can do this, I can do that. I can, you know, find something else." Um and (pause) that kind of, I guess, narrative on their part, we could really see that they were making connections. (Tina 3: 2476-3207)

Like the teacher observations in the Shore School, Tina described what the students said and did during the activity in the terms of the art and social studies standards set out in the original blueprint. The students used the language of resources, scarcity, problem solving, and collaboration. In terms of standards, Tina observed,

I think, I think so very much because one, um one of our standards is, is, our overarching standard is talking about um, given the flexibility of the materials, and, and using a material um, and the kind of understanding the difference between one material and another and what the limitations are. Um so that's exactly, I think this might be one of the most cleanly to see that because they don't have a test experience using all these art materials. (Tina 3: 5160-5611)

The description of student dispositions is incomplete, reflecting Tina's developing understanding of what the learning standard looks like, not in writing or on a rubric, but in action in her classroom. And Amy, the classroom teacher, observed,

Well, I want to say that they were better, you know, in the beginning of the year they weren't essay writers per se, you know, with test prep, all that, you know, we had to race and, you know, it was—they, they're definitely more proficient and that's what they did, they wrote down the side of the page (laughs).... I think like they said, they know (pause) the word "belief" really tricked them up in the fall, so it didn't really relate to anything except of their own religious beliefs. (Amy 3: 7254-8465)

Here, Amy is noticing changes not only in their writing, but also in the content of their writing. She has a sense that their understanding of beliefs is changing. This seems to be a second example of how the teachers at the Hill School are uncovering what the standards

they planned to assess in their blueprints actually mean and look like in the student work. In other words, they are in the process of learning the differences between what the written standards look like and what they look like enacted by their students in their classrooms.

Amy continued,

I'm going to say that the assessment piece that we denote was I think not as strong as it could be this time.... I think, you know, I know how to create a rubric. I, you know, I can—we know how to do that, but (pause) I feel like we—that kind of went by the wayside... The informal was, you know, what I documented while they were actually doing the project.... But the final formal assessment, which was the written piece, (pause) you know, I'm not, I don't know. I'm not sure how well that went.

Here, Amy appears to be self-critical of the assessment part of the work. She observed that she was confident with the student work and with their understanding in classroom activities, but was less so with their writing. In her words, the “informal” seemed stronger than the “formal” assessment. When asked specifically how they observed and what they observed, they said that they observed the students directly, they videotaped them doing the work and observed writing following the final art project. Again, this echoes Amy's approach. She prefers the informal assessment that happens during lessons to more formal measurements. Amy's descriptions of her data use overall are very similar to how she described assessment in her first interview. She is holistic and descriptive in her approach more than linear and analytic. Standards and rubrics serve as a guide, but they do not drive or organize the process.

For both school teams, their data took the form of observations made in the classroom during lessons and during the assessment activities. Both teams drew upon this classroom data to assess their students after both the pre-assessments and the post-assessment projects. They observed student talk and student actions in class as well as written responses to draw these conclusions. In addition, the Shore School team could point to examples of where they were reaching their stated goal of using assessment tools differently, but only at the end of the year. Here the team was increasing their use of

standards-based rubrics to assess growth in social studies skills and dispositions. In addition, they began to increase the student use of those tools in the final project. The Hill School team was not as far along in its creation and use of assessment tools such as standards-based rubrics; they had achieved the goal of creating an interdisciplinary project, but they were just beginning to define what standards and dispositions they wanted to assess.

Finding Three

The teams were not observed scoring or using data in ways assumed by the PADI process and rarely discussed data in this way.

Data use specifically in the form of assessments scored on a rubric that is set to predetermined standards, when discussed at all, was mentioned mostly at the very end of the project, not during the instruction phase and rarely during the pre-assessment. Put another way, while they mentioned assessments 40 times, they never discussed nor were they observed using the results of the pre-assessments to inform their instruction. In total, data use in the form of rubrics and scores was mentioned fewer than 10 times. In addition, recall that during the year, the Shore School team did have a completed rubrics draft reflecting skills and dispositions. In contrast, the Hill School had one rubric that emphasized persuasive writing skills with little or no mention of social studies dispositions and an incomplete art rubric. During the year, I did not observe teachers scoring assessments against a rubric and recording or reporting on results and then using the data to inform their teaching going forward. For the Shore School, the facilitator, Charles, offered both observations and context for how his team did and did not use data in ways assumed by the PADI process. For example, in November, this was his observation of the team's process that essentially did not use traditional data. Here, Charles was explaining that after the pre-assessment, the team at Shore School was not yet focused on measurement either of the whole class or of the individual students.

I think for them it was less of what they're measuring and how they're measuring it, developing the right instruments, really working through, you know, whole class measurement versus very specific individual student measurement and how that was going to happen. (SHORE SCHOOL Charles Int 2: 2055-2318)

This is similar to what I observed of the Shore School team at the PADI training in November 2015. Talk was of lessons and materials—which ones were “flat” and which ones were fruitful. In terms of standards and the rubrics, this is what Lily said in November after the pre-assessment:

That is probably our weakest link. We do have the rubric—I am planning a lesson that Charles [the facilitator] is going to observe tomorrow on the Boston Massacre. And the rubric is definitely in play in terms of what I am planning and what I hope to get from the kids at the conclusion of the lesson. But the kids have not been using a rubric to date. They have not gotten their hands dirty with those yet. (SHORE SCHOOL Team Int 2: 3965-4358)

Lily was suggesting here is that the teachers were focused on the materials of the lessons more than on the outcomes being produced by assessments. They understood the role that rubrics can play in instruction, but were only just beginning to introduce them into the instruction. It could also mean that they were not ready to trust data in the form of numbers and scores at this point, preferring the observational data from classroom interactions. This limitation of number can be part of the process of teachers using rubrics and standards as they work to align what they measure with what they value. Rubric use by students was a goal of Charles and his school, which he mentioned in the first interview. The team was looking for support, suggesting that the rubric and its use by teachers and students had not been a priority to that point.

Some of the most intense data use at Shore School actually came at the end of the project in June of 2016. The team described their work as follows:

We did sort through all the work and we looked at pre and post responses. We tried to find trends with kids like any kind of response that jumped out at us, if there was a general trend, things like that. So we did have a chance to sort of pore through. We tapped everything they did as part of the study 'cause we

didn't know what was going to be important. So it was really kind of culling through all that paper. (SHORE SCHOOL Team Int 4: 893-1309)

Language such as "jumped out" and "culling" does suggest an analytical approach, but not one that used the rubric as a reference. And after the spring assessment, the teachers observed a comparison of their current students versus last year's students: "Yeah, last year they didn't reflect on it too much, but this year they were reflecting along the way.... But they were, they were much more mindful of what they were doing." (SHORE SCHOOL Team Int 3: 4895-5201). Again, these observations of student work seem to have been done without reference to a specific measure or rubric. When I asked specifically what they had observed and how they had made the observations during and after the performance, they said,

In the writing, they incorporated things, you know, like the experiences in the homeland that would, that would make them leave more. When they were acting like the immigrants, they came in (pause) dirty and tired and haggard, whereas last year they came in like they were playing dress-up. I don't know, there was a little bit of authenticity, I think, to what they—to the person they were on Immigration Day. (SHORE SCHOOL Team Int 3: 2925-3355)

The team did analyze student writing as a data source in the spring, and they did score it against the rubric. In terms of the standards and the rubric, they also observed that the students used the rubric in their writing:

I think they were really in tune to the standards and the rubric and they got so used to scoring and so on, and (pause) you know.... I think it hit the standards, especially like the perspective writing, using the research like but not just rewriting what they read in research, but interpreting it and making a person's truth based on that, seeing the expectations, seeing where they needed to go.... They were using rubrics, they were, even as they were working, they had parts of the rubric to guide their work. (SHORE SCHOOL Team Int 3: 4096-5121)

Here, the team suggested that the students themselves were using the criteria in the rubrics to help them work. The teachers could then see evidence of growth in the students' writing. But they never presented that evidence to the students or to me in the form of scores or averages. They never discussed the number of students who reached

“proficiency.” In sum, they were not observed using data the way the PADI training or the initial parameters of this study suggest. There is some evidence that they were moving in the direction Charles had intended for his school, as outlined in his first interview. They were starting to incorporate rubrics and standards into their assessment, and they were beginning to have the students use rubrics more. But it would appear that this work was just beginning after one year. Put another way, it may have been more important that the teachers were focused on the strengths and limitations of their lessons and their assessment instruments in the first year—i.e., what proficiency looks like and how best to see it—rather than counting the number of students to achieve proficiency.

When I asked then specifically how they had used the rubrics and the blueprint developed over the summer, they said they did not follow the blueprint or the rubrics developed over the summer. At the November 2015 PADI training, the pair of Tina and Lynne worked on planning in and around the students' work in art class. They discussed the fact that the work in the art class was touching on two shared value outcomes—“cooperation” and “commitment.” In other words, the two were beginning to see a dispositional outcome take shape. They talked about how to plan the activity in the spring and how to involve the entire grade level through changes to the schedule. Like the Shore School team, the focus was on activities, lessons, and the structure of instruction more than measuring results.

Here is my exchange with Amy, late in the year, when I asked her specifically about data use:

Interviewer: In the fall when we talked, that you guys had kind of moved away from the rubric that was tied to the standards, but have you had a chance to look at the work that the kids just did through the lens of that standard ... or through that rubric?

Teacher: I think we will ... I'm not sure we'll get to it now, but I know that we will get to it because we will use this information to reframe what we're going to do in September. (Amy Interview 3: 10870-11481)

It is only in the second year of the project that Amy intends to use the rubric to analyze results and modify her planning and instruction—the steps outlined in PADI and assumed at the outset of this study. The struggles that Amy and Tina may be having in designing and using rubrics could be indicative of how they are still uncovering and operationalizing what they want the students to be and do.

In sum, the fact neither team was observed using data derived from the assessments in ways PADI process suggests is a contradiction to assumptions made in my original study and therefore could be an important finding. Observing what the teams did with the pre- and post-assessment results, how they made observations about their students, and how they talked about their work sheds light on how the teachers were influenced by the process of designing and implementing local assessments. In addition the observations of the teachers' thinking and of their decisions around what to do with their lessons and units tell an important story of data use. Both teams continued the activities started in the pre-assessments and extended the work into multiple social studies units. One team, the Shore School, adapted an existing performance to observe how the students could apply their understanding of historical sources to a new context at the end of the year. They developed a rubric and used it with their students in the final performance assessment activity to begin to measure growth. The other team, the Hill School, repeated a post-assessment activity in the spring to help them define and focus the standards and dispositions they wanted to observe. In comparison to the Shore School team, they were not as far along in the process of observing growth in terms of the students' understanding. Their talk, at least, in and around their work with the PADI assessments contains the language of data use and vocabulary from the PADI training.

Research Question #2

How did the process of making assessments influence teachers' understanding of the role(s) that assessments play in their work? What changes did they make in their uses of assessment and in their curriculum?

In terms of changes to teacher understanding and to practice, the case study points to examples where changes have happened, where they have not happened, as well as where teachers talk about making changes in the future. Like the first research questions, most of the evidence is in the form of what teachers report along with some triangulation based on observations by the facilitators. In Shore School, the team talked directly about the role of making assessments 12 times and how it impacted their understanding or their practice 6 times. Indirectly, impact and influence were seen many more times in talk of planning, struggling, and thinking. Most of those instances of talk were about their practice but in places imply their understanding as well. Both teachers on the team in Shore School talked about how the PADI process of pre-assessment and post-assessment was shaping their project and their practice, and in doing so how the work seemed to reinforce certain understandings about assessment. In Hill School, the participants talked about the role of making assessments 20 times and how it impacted their understanding or their practice 8 times. Like Shore School, impact and influence were also seen indirectly many more times in talk of planning, struggling, and thinking.

The research question breaks down into three findings. First, only the participants from the Hill School spoke directly about the PADI process's impact on their understanding and practice. One said that the PADI did influence both her understanding and practice, while the other said it did not. Second, the PADI work deepened and reinforced the participants' understanding about what constitutes good assessment. Finally, a particular way participants talked about deepening their understanding of the PADI assessment process was seen in how they learned from their mistakes. Teachers on both teams

articulated what they had missed or what they could have done differently in the PADI process to improve their assessment practice.

Finding One

Only the participants from the Hill School spoke directly about the PADI process's impact on their understanding and practice. One said that the PADI did influence both her understanding and practice, while the other said it did not.

In the Hill School, the classroom teacher, Amy, in contrast to most of the participants in the study, talked openly about her understanding of assessment. In that talk, Amy did not feel her understanding of assessments nor her practice had changed. For example, early in the process, she said,

I've been doing this for a long time, so I have seen the pendulum swing so many times, it's hard to even really remember what I believe in.... But that was really the way I was trained, that was really the way I had always taught....
(Amy Int 2: 22481-22958)

Later she added, "I don't think so. I don't see it—my, my view of assessments has changed because I use a whole, I use very many different types of assessment every day" (Amy Int 3: 15644-15795). In her final reflection, she did find some flaws in the PADI project. This critique of the process below implies that she did learn throughout the project itself, even if her overall views and practice had not changed. For example, she said this of the rubric:

I know how to create a rubric. I feel like we—that kind of went by the wayside because we were just trying to figure out the project throughout the year that in the end, the final assessment (pause) was done like in the very last week of school.... I'm not, I just feel like we let it go by the wayside. (Amy Int 4: 12646-13046)

Here, Amy seemed to understand that making and using a rubric was an important part of the PADI process. In keeping with the tendency of the teachers to extend instruction and delay summative assessment, she used the need to "figure out" the project as the reason for letting the rubrics go "by the wayside." It would seem that in addition to knowing how to make a rubric, the PADI work might have been influencing her thinking about the need for

rubrics in performance assessments. As shown below, Amy was critical of the project and used her understanding of PADI expectations to critique the team's essential question:

The essential question that we came up with Gisele is basically a, question that we can't do. She kind of crafted it for us and gave it to us, and I'm not sure if we really understood it ... and it was kind of derailed, and then we were kind of using someone else's essential question of drive the project, and I don't think either one of us owned it. And then the kids didn't really get it because it was so broad. (27954-28159)

Here, Amy was suggesting that the project lacked focus because the essential question ("How do beliefs impact your view of the world?") was too broad. She suggested also that the training caused this. The result was that the kids struggled with it. In terms of how the children were changed by the experience in her classroom, "I don't, I don't know that the kids changed.... To be honest. (pause) I don't know that there was a significant or notable change.... As a direct result of that project, I would say none" (Amy int 4: 27954-28159). Unlike the team from the Shore School, and to some extent unlike Tina, Amy's view of her students seemed unchanged. This kind of statement lends itself to the conclusion that Amy's understanding and practice were unchanged. And yet, she valued the concept that a big curriculum idea should be experienced by students in many forms to help them make connections.

I think that the art piece of it is another layer of learning, so you know, when I taught younger kids, my experiences that if kids can talk about it, read it, sing about it, you know, dance about it, you know, the more they can do sensory-wise, the more profound the learning is, for early childhood especially. So I felt that the art piece would be a different way for them to make a connection ... in a more global way, those larger concepts. So that was—that opportunity was exciting for me. (Amy Int 4: 21042-21797)

So while Amy was strong in her own beliefs and in the work done by the students in her class, she did see something different in the art project. The language of "layers," "connections," and "transfer" would suggest that the students were growing and changing because the PADI project had them working on the same curriculum ideas across disciplines. For Amy this was "exciting." Perhaps, given her background and initial views on

assessment and learning, this was in fact not new. But the expression that it was exciting is at least significant in that the PADI project as a whole was influencing, or at least reinforcing, her thinking. In addition, she brought a new kind of performance assessment to social studies, where it did not currently exist.

The art teacher at the Hill School, Tina, who was new to the school, in contrast to Amy, reported more definitive changes in both her understanding and her practice. As such, Tina was the only participant of the four to discuss how the PADI process changed her understanding and her practice around assessment. For example,

Interviewer: But is this, is this impacting how you are looking at, how you're assessing kids' work outside of this project?

Tina: Yes, you know, yes. I think this year, maybe as a result of this experience, I've kind of (pause) shifted um (long pause) my, I guess, thinking about how to assess what's going on and I've been more focused on their reflections and thinking than their product. You know, An art product is always going to matter just as evidence of the understanding, but there's things, you know, situations where there are students who, for example, just have a really hard time and their product is not coming out great because they have a really hard time. But their, I guess, reflection on what's going on and even being able to say, you know, this isn't working and I know why it's not working, um I think is taking more value than it used for me. (Tina Int 3: 3906-4840)

Here, Tina expressed two direct changes from her previous practice. First she mentioned her shift in focus from what she would assess toward reflection and thinking and away from product in art class; second, her understanding that students should be able to reflect on what was not working and why.

At the Hill School, the facilitator reinforced both of her teachers' observations. Lynne also noticed that Tina seemed to have been influenced the most by the PADI project. She noticed that Tina was starting to do things she had not done before. Lynne seemed to understand herself that this kind of work was what she wanted all her teachers to do more of. Toward the end of the case, she repeated her belief:

I think the attention to assessment was important because my teachers didn't necessarily value assessment at that much of a level, and I think that what's happening now is that seeing the pre-assessment and then the first post-assessment, particularly with art, and then going back and revising the task and then looking at the student work after that I think was an eye opener, particularly for the art teacher, because the kids really did achieve at such a higher level than they did initially. (GH Fac Int 4: 607-1103)

In part, Lynne gave credit to the learning and changes made during the process. In particular, she pointed out that Tina's focus seemed to be more on the shared value outcomes of collaboration and creativity (process) and less with the product of the art. And yet, she also saw less change in Amy's understanding and in her practice. She offered some explanation as to why. While Amy's topic and content focus shifted to Social Studies, she did so through a traditional performance—a writing prompt.

Interviewer: What about with Amy? Did you get a sense that she, that this was changing her practice or, or not?

Facilitator: I don't think as much, particularly because she had something, she had a goal in mind. She followed it and she did the writing process the way she normally teaches. I don't think that the assessment made that much of a difference for her. (GH Fac Int 4: 1679-2020)

In sum, the teachers at the Hill School diverged in their assessment of how much the PADI experience seemed to be influencing their understanding and practice around assessment. But unlike their counterparts from the Shore School, both teachers spoke directly about PADI's influences.

While the teachers from the Shore School do reflect on their understanding and practice in subsequent parts of this chapter, neither teacher spoke directly to the question of whether or not their understanding or their practice had changed. However, the Shore School facilitator offered some suggestions of ways the work might be affecting them. In addition, both the participants and the facilitator from the Shore School often talked of continuing the work into a second year, which implies, at least, a commitment to changing their curriculum to keep the PADI assessments going in year two. In one closing remark

toward the end of his interviews, Charles, the facilitator for school one, offered this observation:

Yeah, I think both because it's just not the way they were trained, I was trained (laughs), you know. It's also new. It's, you know, it's kind like if you think about universal design and some of the newer models of instruction, the idea of metacognition being a newer concept. It's the idea as the assessment as learning and as the child, the student, is a key part of the assessment process. (SHORE SCHOOL Fac Int 4: 3510-3760)

His recognition that the work of making and implementing local assessments is difficult, because it is different from what many teachers do, is an important observation. The work is not simple and rarely linear. In many ways, the question of whether or not the teachers in Shore School changed remains open and a focus of the next chapter's analysis. Unlike Amy, both Lisa and Paula did talk of learning from the work. They did plan to continue the project with several plans for how to modify it. Charles, the facilitator, saw the beginnings of changes in how they were applying their knowledge of rubrics to the social studies units. He saw them beginning to "close the assessment loop," and therefore remained optimistic as the work moved forward into a second year.

Finding Two

The talk about the work deepened and reinforced the teachers' understanding about what constitutes good assessment.

Both teachers on the Shore School team talked about how the PADI process of pre-assessment and post-assessment was shaping their social studies units, and in doing so, how the work seems to reinforce certain understandings about assessment. In particular, the teachers described, then Charles reinforced, the idea that assessment is a process. For example, their description of using PADI structure of defined standards and rubric use in pre- and post-assessments led to this exchange,

Interviewer: Did you find that move [pre and post assessment], that teaching move would be something that you would want to apply to other projects or other work in your curriculum?

Classroom Teacher: Definitely, definitely. I think we walked away believing in that a thousand percent. (SHORE SCHOOL Team Int 4: 8110-8341)

Lily also said,

So I think the PADI experience, um, made me more thoughtful in my decision-making. It made me more aware of the outcomes that I was seeking as I was working through the unit. It made me look at things through a different lens so I was always trying to find resources or questions that would enable the kids to think differently... and own their, you know, to validate their thinking and to, to put them in a position where they wanted to share it. (Shore School Team Int 4: 12488-12972)

Throughout their interviews, the team from the Shore School talked about the future of the PADI unit and assessments beyond the first year as an inevitability. In talking about the future, the facilitator also said the following as a reinforcement of the team's view:

And instead of scoring, it's about identifying where and locating where I am and where I need to go, and I can only know, and so this is definitely something that Gisele has taught us, really knowing where we are in order to know where we need to go. Not just being familiar with what I'm not doing, (laughs) you know, or where I didn't reach.... (SHORE SCHOOL Fac 4: 17569-17921)

The team echoed the importance of pre-assessment along clear standards, which was a central teaching point in the PADI training. Assessment is not about scoring or calculating a grade. It is about locating where a student is and what they need next to grow. The team was describing a process in the classroom that the teacher and the student were engaged in. It is interesting here that Lily was talking in the first person, implying that both the teacher and the student need this insight from assessment. Throughout the interviews, the Shore School team talked about assessment as a process that crosses an entire unit, even an entire school year, not as an event focus on one or two tests.

In addition to these reflections on their technical understanding of assessment, the teachers in Shore School did talk about how the project they developed reinforced the kind of performances that they value. For example, they said of the work itself, "I do see them thinking in terms of perspective more than I would have ever thought possible.... It's

inspiring. It makes we want to do more of it" (Shore School Team 2: 6356-6856). About the students in particular, the special educator on the team said, as she compared the students doing the PADI work in Lily's class to others on the grade, "I'm in all the different classrooms, there was a striking difference in the level of conversation in some classrooms in fourth grade, and in fifth grade" (SHORE SCHOOL Team Int 4: 2254-2698). Again, the teachers in Shore School were seeing a kind of thinking on the part of students they had never seen before. In addition, they were describing this thinking in the assessment language of growth over time.

Similar to the teachers, Charles, the facilitator at Shore School, saw intrinsic motivation and value in the work the teachers were doing. He viewed his teachers this way:

Well, they're huge proponents of it. I mean, I think um the work that they're starting to see has really made them feel more positive about their work and, you know, and more, you know, that's, that was there, but at first it was, okay, it's because it's an alternative to state testing and test prep. (Shore School Fac Int 3: 9776-10076)

Here, Charles saw that the teachers valued this kind of work as compared to mandated test-driven curriculum. And in the final interview, he said,

I also think the big change is, even though it wasn't perfect last year, is the idea of the presentation of assessment and reflection on that, that it's not just a one-time event, that it's a process. (SHORE SCHOOL Fac Int 4: 9441-9641)

Charles was echoing the key idea of the PADI process, one that teachers seemed to have embraced, that assessment should be ongoing. It should be as Paula described it in her first interview: "It means sitting up, pulling a chair up next to a kid and see what they do on a day-to-day basis, not on a like a memory test or talking to a kid" (SHORE SCHOOL Team Interview 1: 6090-6236).

Both the team and Charles touched upon several values that they had expressed in their initial interviews. First, the team valued the idea that assessment is not simply a test, but rather a process that happens regularly in classrooms with teachers and students working together on activities of value. Then, the idea of teacher being designers of

curriculum was also appealing to the team—designers not only of activities, but also the processes for evaluating and supporting learning over time. Finally, the team’s observations suggest the iterative quality of curriculum where teachers are expected to experiment and improve their practice over time, unlike the planning blueprint, which suggests that the entire process can be created at once. Charles understood what the teachers were in fact doing—designing lessons and experiences first, then following that with a performance assessment. All of this would take more time than the PADI planning would suggest.

In Hill School, all members of the team also talked about their work in terms of developing their understanding of good teaching and good assessment, but Amy and Tina did so in different ways. Amy’s comments were more general, and Tina’s more focused on aspects of the work that were newer to her practice. For example, Tina described a shift from product to process thinking when talking about how the PADI work had deepened her understanding of interdisciplinary work. Again, she described the differences in terms of student thinking, valuing work that goes beyond the “cute tie in.”

I guess the PADI mindset, is that you have this collaborative interdisciplinary project that has a depth to it, and I’ve always done what was called interdisciplinary work.... But it’s not true interdisciplinary work, it’s kind of like just a cute tie-in. And being able to experience that you can take two different or even three different areas of thought and how to really make deeper connections, not just kind of like thematic connections, I think is something that is going to impact me going forward. (Tina Int 4: 7521-8369)

Here, Tina’s understanding of the kind of thinking she was asking of her students was shifting. Interdisciplinary art is not just a simple add-on activity. But like Amy’s idea of ‘layering,’ it is looking for more meaningful connections that develop thinking. She also discussed the way students could talk about the process of making art was changing in her mind based on what she saw. In addition, she recognized an additional skill that she valued in the process: “I’m seeing that they’re working on collaboration. They don’t have it yet, and collaboration was a big piece of their project and a big piece of their assessment in their rubric” (Tina Interview 2: 20296-20617). Lynne, the Hill School facilitator, also picked up

on this and reinforced the teacher's point of view: "... what she [art teacher] discovered in the first two attempts in the fall is that she really had to teach into collaboration" (GH Fac Int 3: 4073-4191).

Since "collaboration" is one of the Shared Value Outcomes at the school, this was of particular interest to the team. Here the influence the PADI process seemed to be having on Tina's practice came through her recognition that they changed the focus of the project and the structure of the project mid-year. As such, we see that Tina's experience was unique to the four teachers in the study in that she reported not only a deepening understanding that reinforced her values but also actual changes in her practice. In particular, she was seeing the power of the students layering the art experience onto their understanding of social studies and vice versa. Limited resources were experienced in a real way in the art project, and then understood in the abstraction of other cultures such as Native Americans and American Colonists. In addition, the Hill School team did talk in the assessment in terms of measuring change or growth over time. But unlike the Shore School team, who actually began to do the work of measurement with rubrics, the Hill School team was at the beginning stages of defining what they wanted to measure.

Yeah, we're set for next year. We learned a lot ... if the students are reflecting with her and then doing one art experience with me and then going back to reflecting with her, the thinking, we can measure that thinking and change. (Tina Int 3: 9525-10556)

"Measuring thinking and change" was an emerging focus of the work beyond the activities themselves. Amy, classroom teacher at Hill School, also talked about what she values,

I think what I appreciate about the pre- and post-assessment is the evolution of an idea. And in our, especially with our question because the initial question was, you know, how did beliefs impact what we see and do, and the kids really made it about themselves.... And then to see them move outside of themselves to a different group of people, (pause) which was interesting to me and really to see that evolution is really what prompted me to think that now I'd like to do this again.... I don't know, it just seemed to me that it's an

evolution of an idea and that the kids are really starting to get their feet wet, and I just don't feel right abandoning it here. (Amy Interview 2: 9510-10486)

Here, Amy used the expression “an evolution of an idea” twice to describe a process that she valued in her classroom. Amy saw her students' thinking evolving from a place where they could understand something about themselves to one where they could understand the resource needs and beliefs of others. She also seemed to believe intuitively that this kind of thinking takes time and that it therefore deserves more attention in her lessons.

At Hill School, the facilitator, Lynne, reported some deepening understanding on the part of her teachers. In fact, she went farther to suggesting that their understanding was changing, particularly in Tina's case. Like Charles, the facilitator at Shore School, Lynne reinforced the fact that her teachers valued having the ability to design and teach outside of the prescribed units of study in Language Arts and mathematics.

I think it's changing their thinking.... Now we never used to have programs. Everything was teacher-planned and it was all individualized. But now we have TC writing, we have math of focus, we have TC reading. The units are just about scripted, even though you have leeway to do different things. But having the opportunity to do a project that is totally up to you is something that they're going back to and they haven't done in a while. (GH Fac Int 2: 8077-8679)

Lynne also saw this work as a return to an older way of teaching, whereas Charles saw the freedom to design as an opportunity to embrace new assessment and teaching practices. Lynne closed her final interview with the observation that she saw intrinsic value in this kind of practice, which seemed to reinforce her belief about how best to teach and assess. She explained that teachers traditionally did not seem to value assessment and were benefitting from exposure to this kind of work.

We're struggling because assessment is not something that the teachers value or have ever valued here at [Suburban Woods]. I don't know if you know the philosophy here, but it's more about the whole child and assessment should be driving everything we do, and our teachers don't always use it to drive it ... they don't always take the time to reflect on it and go into that

deeper level, so having the opportunity to incorporate assessment more than we have is really important.... (GH Fac Int 3: 5652-6846)

Lynne saw the PADI focus on the process of assessment as potentially changing practice in terms of how assessment can “drive” instruction and push curriculum to a “deeper level,” even though neither teacher expressed this sentiment directly.

So while they had no intention of creating a new performance assessment at the Hill School, the two teachers and their facilitator, Lynne, saw some value in the pre- and post-process. For them, the main value of the PADI process lay primarily in the addition two art experiences combined or layered with one or two social studies units that culminate in opinion writing. But the impact of this accomplishment seemed different for Amy and Tina. For Amy, the only actual change may have been that she added the pre- and post-process to a social studies unit where it did not exist before. But for Tina, she changed what she was assessing and how she was assessing it.

Finding Three

A particular way teachers talked about deepening their understanding of the PADI assessment process was seen in how they learned from their mistakes.

Teachers on both teams articulated what they had missed or what they could have done differently in the PADI to improve their assessment practice. In particular, both teams saw the challenges of measuring growth over time when assessing complex tasks and skills. The most significant evidence for this can be seen in changes in their understanding of the PADI structure and its assumptions about the pre-assessments standards, and rubrics, as they drive instruction toward growth.

The Shore School team did not articulate changes in their understanding specifically, but the influence of PADI on their understanding of assessment’s role in practice can be seen in the way they talked about mistakes they made or realizations they made as the project moved forward. For example, they concluded that the choice they made in November to extend the project to other units had become a problem by the end of the year.

"I mean this, if I'm being perfectly honest, this was a freight train out of control by the end.... It started as a little project that we were going to quit after Native Americans, but because it fit with everything we were doing, we just kept it rolling" (SHORE SCHOOL Team 3: 6760-7028). This was perhaps a reference to their decision to extend or expand the initial project into two or three social studies units. Specifically, they talked about the challenges of the final assessment without a solid rubric in place.

I think the ending was very challenging for us. How to wrap it up, what, what we were actually using to measure their growth, that was really, that required a full brainstorming session with [the Facilitator] himself. (SHORE SCHOOL Team #3 8587-8798)....

Yeah, right from the beginning because it was like we lacked that target at the end that we were shooting for, so we were kind of this way and that way and this way and that way, and then we would hone in on the target. But as we were going through the process, we kind of waffled. (9271-9552)

By the end of the year, they saw that to simply extend the activities based on a general observation that what the kids were learning did not result in continued growth over time. Without a clear rubric from the outset, and pre-assessments tied to that rubric, their goals for growth were not clear. In other words, realizing the mistakes they made throughout the year—moving away from the prescribed PADI steps in particular—reinforced their understanding of the need for clear learning targets and clear rubrics in order to measure growth over time. The team understood this after the first year. Again this is why Lily concluded,

I think my advice to somebody going through it next year is to keep it small. Don't let it get bigger. You know, I think if we had clipped it and really forced ourselves to end it after the first unit, we could have taken what we learned and used that going forward instead of just kind of muddling through the whole. (Shore School Int 3: 9491-11618).

The realization here seems to be their connection between the size of the project and its lack of a focused outcome. Put another way, Lily had made sense of the problem—keeping

the project contained in one social studies unit would make it easier to focus the intended standards and measure growth.

The facilitator, Charles, reinforced the teachers' position when he reported progress in the teachers' understanding and practice around assessment. He also saw challenges and shortcomings in their work that were similar to those experienced by teams in other PADI cohorts, and still feels that they need help "closing the assessment loop."

I think the hardest thing for people to understand, myself included, is that closing of the assessment loop (pause) and really, how do you show growth? How do you show growth and specify growth in an area and tie that to the pre- and post-assessment? (SHORE SCHOOL Int 4: 3510-3760)

In addition to describing their assessment understanding as "developing," Charles provided an additional detail specific to his three years of experience with PADI unit and assessment development. For example, he said of how the teachers were extending their work, "But the other thing that I'm noticing with that group, because they started so early embedding it and thinking about it, is they're seeing it everywhere" (SHORE SCHOOL Fac Int 2: 5073-5226). Embedding the work everywhere may make for exciting and compelling activities, but it does make the growth harder to assess. The suggestion here is not that the teachers were wrong to extend the work, but rather that the way they approached the task made it difficult to assess. And specifically what the team learned about the rubric use after they had deviated from the PADI blueprint:

And I think at that point, they realized even though they really did make use of the rubric, just going into next year how they could use the rubric in a more developmental way so that it can be used as a source of reflection for the kids throughout the whole process, starting from day one. (Shore School Fac Int 3: 2746-3040)

Here, Charles returned to his original goal for the school that year. He believed that activities and assessments and therefore growth would improve when the rubrics were in the hands of the students and they were more aware of the learning targets.

In Hill School, the team also revealed changes in understanding as they talked about mistakes they made and what they learned. But in some ways, the two teachers on this team seemed to take away different understanding about the PADI process—again with Tina seeing more change in her understanding and practice. Tina saw mistakes in her understanding and practice, while Amy tended to find fault with the PADI process. Amy blamed the students' initial struggles not on the activity but on the essential question the PADI blueprint required. She said at one point that the Hill School project “was kind of derailed, and then we were kind of using someone else’s essential question (pause) to kind of drive the project, and I don’t think either one of us owned it. And then the kids didn’t really get it because it was so broad” (27995-28159).

In contrast, Tina, the art teacher, reflected this way:

I’ve been trying to actually, trying to adjust, I just like ___ reflections with the kids, and I’ve been trying to kind of over the summer tweak how I’m doing that to kind of get a little bit more out of it in that way in looking at my general teaching. (pause) You know, because honestly, not all kids, even if they are in art eventually going to make great art, you know, in that, in that sense, they don’t necessarily have the ___ skill that they’ll ___ get and that’s not something (pause) that should be (pause) I guess penalized or, you know, being able to look past what (pause) unimportant limitations are to seeing like the real true (pause) I guess understanding and capability and appreciation is, I guess it’s harder but it has more value. (Tina Int 4: 9625-10380)

Here she was describing how the PADI project was causing her to adjust her approach to other teaching to assess and to focus on what she truly valued in addition to the shared value outcome of collaboration mentioned earlier. Her values seemed to be shifting from the art produced toward the students’ ability to “reflect,” to “understand,” and to “appreciate.”

Tina, like the other teachers on the Shore School team, revealed some of her changing understandings through her reflections on what went right and what went wrong with the project. For example, Tina observed that in the first year the experiences or activities were not logical or cohesive. This reveals some increased understanding of the pre-/post-

assessment structure. One specific change to the project and to Tina's thinking can be seen in her talk around the use of the rubric.

What we ended up doing was making a rubric that was basic because when the students made the rubric, they wanted to explain a lot of things and then it became so wordy.... It was all, it was a cumbersome activity for them and it took a long time.... They weren't able to just refer to it and refer to their work and say, "I'm here, I need to get here." It was, it was kind of too much. (Tina Int 4: 7521-8369)

In this observation, one can also see that her understanding of the role of rubrics was shifting. Like the team in Shore School, Tina observed that the rubric can be used not just to guide the teacher in the assessment, but also guide the students in their learning.. In their year-end reflection, the team seemed to envision a better order and structure for the project where they timed the written reflection for after the art experience. Like Shore School, the team learned from their mistakes and now envisioned a more streamlined and less drawn out set of experiences with targeted rubrics and learning goals.

In sum, teachers from both teams planned to take what they had learned and build on it. In their interviews, the team members could point to PADI's influence on their understanding of the nature of assessment. In many cases, these influences may in fact reflect a deepening in understanding. On the surface, they described the excitement that comes with successful lessons and students' work as well as the frustrations of mistakes made and final outcomes that may have fallen short of goals. Their commitment to continuing the work in the future and to continuing to work on the new units based on what they learned in the first year points to new understandings and to potential changes in practice, but not in a way that demonstrates a conclusive causal relationship between the PADI work and changes to practice.

Research Questions 3 and 4: Barriers and Support

Research question 3 asks about barriers and support for teachers while creating local assessments, whereas question 4 asks about the barriers and supports for teachers in making changes to their instructional practice or to their curriculum. This final findings section will look at all occurrences of barriers relating to questions 3 and 4; then it will look at all occurrences of support.

Barriers

In total, participants mentioned barriers 77 times in their interviews and supports 42 times, which is significantly more than they talked about data use. Many of these occurrences also overlapped with concepts of trust, pressure, making, and design, as well as agency. In terms of barriers to creating local assessments, the participants described barriers that fell into four areas—mandated curriculum requirements (sometimes referred to as the “pacing calendar”), time, state assessments, and the difficulty of the PADI task. The findings, however, focus on just two: mandated curriculum and state assessments. Time, or the lack of it, is a universal concern and did not yield new insights to my questions. And the difficulty of the PADI tasks were already explored in the section of “learning from mistakes.”

Finding #1. Mandated curriculum requirements (sometimes referred to as the “pacing calendar”) was described as a barrier. Mandated curriculum was cited seven times directly by all teachers as a barrier to the creation and use of local performance assessments. The classroom teachers on both teams had 20 plus years of experience. They observed that the new mandates of common curriculum and pacing calendars had crowded out other work. For example, the team from school one said,

Interviewer: Can you say a little more about “so much else going on”?

Teacher: That is coming from 7 reading units, 7 writing units, a new math program (Laughing). We have varied challenges, what we are supposed to be on lesson by lesson. So our toes are to the fire on any day.

Interviewer: Is that district or school level?

Teacher: It's district. It's district and school. [Facilitator] is the most flexible of the principals, but we do have challenges in terms of the math curriculum. (SHORE SCHOOL Team 2: 4902-5372)

And Amy from Hill School described the challenge as the “constraints of the day.” She also said it this way:

I feel like we have a lot to do and in some ways it's, it's just neither, it's just neither doing it the way that we've been doing it. You know, you had your objectives, you worked to meet them, and (pause) you know, you kind of just work through it to get to, to do the next thing. Does that make sense? (Amy Int 4: 10165-10469)

In both schools, the experienced teachers saw mandated units as supported or defined by pacing calendars as new features in the last 5-10 years. In particular, they mandated units in literacy, in math, and to some extent in science. Both teams referred to the “old way” of creating curriculum, where individual teachers and grade level teams had much more control over how units were created around New York State Standards. And they also mentioned social studies in particular. It is interesting that both teams selected social studies as the focal point of their projects. PADI teams in the previous year in both schools also selected social studies-related projects. In particular, Amy from Hill School said it this way:

Well, you know, content sometimes goes by the wayside now with the focus and emphasis on the reading and writing and math parts because of the assessments. So I ___ some fourth graders also assessed, so my fear was that social studies would kind of fall by the wayside if I, if we didn't bring it to the forefront. (Amy Int 1: 2319-2634)

Amy was explaining here that the desire to focus on the content of social studies was one of her motivations for taking on the PADI project both because literacy, math, and science units are dominated by new curricular demands and because she feared that social studies content was being ignored. Both facilitators also mentioned the changing curricular demands of mandated units defined by pacing calendars as a challenge for the teachers to

plan the PADI work and to fit in the final performances. For example, Charles from Shore School said,

The first part was taking a realistic look at how it integrates into the existing curriculum and the calendar, and if there's a pacing calendar, how the project would fit into the pacing calendar. So for example, with the fourth grade project, there is kind of a strict pacing calendar with the Teachers College Readers and Writers Workshop. We do multiple units and so trying to make it work in a way that is true to the project, but that also doesn't require, you know, so many changes with the pacing calendar. (Charles Int 2: 345-858)

And, "so it's hard to differentiate what is something, what needs more attention, what needs less attention, 'cause the way it's presented, everything it seems like it's equal" (Charles Int 4: 11454-11624). Here, he was implying that teachers struggle to prioritize among the demands placed on them by mandated curriculum as defined by the pacing calendar. Lynne also recognized the scripted nature of new curriculum and, like Amy, saw the idea of teacher-created lessons and units as a move "back" to curriculum before the current accountability environment. In terms of adding new content, she even tipped her hand to playing a role in the "additive" nature of innovation at the Hill School: "It's just a matter of finding the time and the school year where they're willing to try one more new thing and in addition to all the things they want to do" (Lynne Int 4: 11995-12101).

Within these statements is the suggestion that even if the teachers wanted to change their practice to do more work like this, there simply was not room to do so. Amy spoke almost with a sense of loss:

Amy: I used to be able to do that much more.

Interviewer: Yeah.

Amy: And that has really been taken out of my hands, as we've become more of a, a top-down kind of system. (pause) So for me, you know, having that essential question and kind of being the engineer of that kind of puts it back in my hands, and it's, for me as an educator, it gets back to what I think is important is, you know, the flexibility of thought, the

evolution of thought. You know, and that it isn't lock-step for every kid.... (Amy Int 2: 11784-12269)

This statement relates back to Amy's statements about the kind of curriculum she values and the kind of student thinking she believes goes along with it. She longs for experiences for herself and for her students where ideas are "flexible" and can "evolve."

One closing observation about mandated curriculum that is important comes in the notable silence of Paula, the special educator in Shore School and of Tina, the art teacher in Hill School. Only the classroom teachers and the facilitators raised the issue of mandated curriculum and pacing calendars as barriers.

Finding #2: State assessments were described as a barrier. State assessments were only mentioned three times directly. But the frequency with which they are mentioned may not be the best measure of their significance to planning. At both schools, the tests combined with the mandated curriculum had a significant impact on determining the timing of final performance assessments. The phrase "after testing" and the freedom of June were raised multiple times. Both teams chose to delay their final performance assessment activities until June for this reason. For example, Shore School suggested a somewhat mixed message about testing:

You know, we're so fortunate that state testing is not something that is—it's obviously on everybody's radar, but we are not crazed by it. It's not, we're not in a test prep kind of place, so we have the luxury of doing what we needed to do when we needed to do it. (SHORE SCHOOL Team Int 4: 8514-8962)

But they also said, "June is a great time for kids to synthesize all that have learned and done. And there is not as much curriculum demand in June then in the months leading up to the tests, So they can really get into it" (SHORE SCHOOL Team Int 2: 5975-6178). This statement actually captures both the demands of mandated curriculum alongside state testing. The idea that June is an open month speaks to the pressures in the other months to comply with external demands.

Then Lynne from Hill School captured the assumptions behind choosing June as follows: “We’re going to wait till after the tests are done so the teachers can really get keyed(?) into this project. We’ve wanted to do an interdisciplinary unit and get away from teaching to the test for a while” (Lynne Int 4: 5864-6069).

In all, the way the teams mentioned the “obviousness” of June to do the kind of performances they wanted to try shows how much the position of state testing and the mandated units has shaped their thinking and perhaps limited their choices for designing new performance assessments.

Support

The concept of support, like barriers, is found in both research questions 3 and 4. As with barriers, most of the 42 instances where support was mentioned show an overlap between support for the work of making local assessments and support for changing practice.

Finding #1: Both teams expressed feelings of support from the district culture and from leadership that helped make the challenges of the work achievable and worthwhile. Both teams of teachers expressed feeling supported by their school culture four times and by the facilitators five times. For example,

You know, we’re so fortunate that state testing is not something that is—it’s obviously on everybody’s radar, but we are not crazed by it. It’s not, we’re not in a test prep kind of place, so we have the luxury of doing what we needed to do when we needed to do it. (SHORE SCHOOL Team Int 4: 8514-8962).

And Amy from the Hill School offered about Lynne, “I think, I think she would support anything that we needed to do. I think she would be completely supportive and give us planning time. (pause) Yeah, I think she, she’s always been very supportive of whatever we choose to do” (Amy Int 4: 24074-24299). Likewise, when it came to the facilitator’s support for their team, the teachers from the Shore School also expressed feeling supported.

Finding #2: Both teams saw the facilitator as taking actions to provide support, but in different ways. In terms of the specific steps taken for support, the two teams had less overlap except for references to time. Both were grateful for the time to do the work. One team was much more detailed in how the facilitator supported them. The teachers from Shore School described several ways that the facilitator supported specific aspects of the work. For example, "He gives us the autonomy to do what we know we need to do. He is encouraging but he gives us the reins, so he does not have to be in to watch every step" (SHORE SCHOOL Team Int 3: 8428-8602).

In terms of the planning process, he has been integral. He has been our secretary, our coordinator. He has helped bring additional resources to what we are doing. In terms of the classroom, he hasn't really played a part. You know he will see it tomorrow. He is definitely our biggest cheer leader. We have Flex time, so one out of the cycle, we have 30 minutes to meet to talk about the PADI project, so he has been working with the grade level to create a rubric. So he is trying, he is definitely trying. (SHORE SCHOOL Team Int 2: 7110-8225)

The teachers from the Shore School also described Charles analyzing the data with them both during the pre-assessment and the post-assessment. One teacher had him in to observe lessons. He also brought Giselle Martin Kniep to observe and work with the team in their classrooms. In all of these steps, Charles's actions were aligned to the Shore School's goal around creating and using rubrics to support the measurement of growth. Knowing that the work takes place in the classroom, he literally stepped into that space to facilitate the work where it was occurring.

In the Hill School, the teachers cited one moment when Lynne provided specific support for their scheduling problem: "... she ___ was very good about giving me and Annie collaborative planning time, giving us release time so we could push into each other's classroom" (Tina Int 4: 24895-25012). Amy also mentioned that Lynne got her a sub so the team could co-teach the final project. But beyond that, it is remarkable how much Amy talked of not needing Lynne and of her non-involvement. She said,

Yeah, I just don't think we've needed her and (laughs) maybe we did. I don't know. Maybe we would have done better if we asked her for more help. But really the only thing that we really needed was time to meet and talk and craft(?) and plan and those were things that she really—other ____ outside of that, and giving us that time to make that happen and ____ what she could have done for us. (Amy Int 2: 16595-16987)

In general, the description of support on both teams seemed to be at its strongest when it addressed a corresponding barrier. This connection is similar in the ways that both facilitators talked about how they saw themselves supporting or facilitating the work. While the facilitators said similar things in general terms, Charles was much more specific in what he thinks the teachers need and what to expect of them in the future. It seems, again, that Charles was drawing on his three years of experience with PADI projects. First and foremost, he described how he had built collaboration into the fabric of his school's schedule.

We have, and I might have mentioned this in the summer, but we have collaboration planning times built into the day.... That are a byproduct of our schedule. So we have the specials block and then we have what we call an academic specials block, which is where they have flex or library or enrichment. And the whole grade has it at that time, so that's a block of time, you know, a few times a week, half hour, where we can get together or they can get together on their own to keep working on this, tweaking it, planning it. (Charles Int 2: 7139-7674)

Charles seemed comfortable offering substantive support in the form of social studies content, assessment knowledge, as well as his experience as a PADI facilitator. At the same time, as seen below, he struggled with how he helped his team.

I'd rather them, it not be perfect that first year and let(?) them eventually get to where they are now, than to try and manage it maybe too much and then they don't really—they're not invested in it and they don't understand it, you know. So it's like constructivist learning, even though it wasn't perfect last year, if they eventually get to that place. (Charles Int 4: 14255-15234)

The idea that he would rather his team not be perfect and make some mistakes in the first year is an interesting stance for a facilitator to take.

Lynne at the Hill School had a clear vision for her teachers. She was hopeful that this work would engage her teachers and, as a result of that engagement, create great experiences for the students. She saw her support in two ways: first, as a cheerleader for the work, and second, in the way she manipulated the schedule so that the team could work together in June. She then described a third intention as facilitator:

But the next step for me is really reflecting on the process of the two teachers because they do want time to debrief and they want time to plan for next year because they see the, they see the pitfalls, they understand what's not working, so together we need to figure out how to go back to the essential question, how to start with the standards and make sure it's all there, but how to guide it more and then utilize the assessment piece. (Lynne Int 3: 8921-9171)

Here, Lynne's intention to help the teachers analyze results is interesting. It suggests that she saw a need. What is interesting to note, however, is that this reflection work with Lynne did not seem to take place either during the post-assessment activity or by the time of the teams' final reflection.

Finding #3: In contrast to the teachers, both facilitators had intentions to share the PADI work with other teachers in their schools. The facilitators talked about the work being shared differently than the teachers did. Both teams expressed barriers to sharing the work with other colleagues or growing the work to include the entire grade. For example, the team from School One said, "My grade level is a little tricky this year. It's not really functioning as a cohesive group. So they don't really—I don't know that it is something that they would be interested in adopting or adopting (pause and laugh) with integrity" (SHORE SCHOOL Team 2: 7830-7618). And Tina from School #2 said,

I think this is our first go at it and only two people were interested in doing it. So I think the culture of that building, I don't know if [Lynne] could even expect, but it really is not a (pause) a passion, a respect, an understanding of interdisciplinary work among teachers in a way that I see in the other elementary school and the middle school.... (Tina Int 2: 15544-15893)

In both cases, it is clear that the teachers themselves did not see a way forward to how the work could be or should be shared and spread across the grade levels at their schools.

Both facilitators, however, took the opposite view. They very much wanted to share the work with other teachers with the hope that it would help their respective schools. They saw the sharing process as a way to support the teachers and support their school's work in the area of assessment.

Charles, the veteran of several PADI projects in his building, saw the power of the work itself in motivating others. If you recall, he set the stage for sharing at the beginning of the year at his first faculty meeting.

... in the Monday faculty meetings here at the school, that's all we were doing this year was really talking about the role of designers, of designers of learning opportunities, and we really kind of used the PADI framework to structure the PD in the Monday meetings. (Charles Int 1: 4018-4374)

Charles began the year by sharing the PADI framework with his entire faculty meeting. He saw the process of designing as a way to transform teaching in his building, so he was very supportive of spreading the work. Later in the year, as the work was developing, he presented the PADI work to parents and indicated his plans to share the work in the spring.

I think that's a good question. I would like to for the first time present about PADI at a PCA meeting or PTA meeting, so that's something that I'm hoping to do this spring. And, and in the spring because then fourth grade will have finished their projects and I can talk about upcoming projects. (Charles Int 2: 25137-25434)

And here, you see that he was planning the work to continue school wide.

So we've started to have some conversations, and while we're talking about the end-of-year project that involves everybody, that those PADI teachers will share what they're doing in their project and/or make those connections. So we're specifically focused on measurement and developing an interdisciplinary rubric and really showing to students and parents the growth in their project. (Charles Int 2: 16198-17249)

Here again, you see how Charles was leveraging the PADI teams in his building toward further ends. First, he was looking to have teams design year-end performance assessments and projects, and second, he intended for these projects to have a measurement component.

Lynne also shared enthusiasm for supporting the sharing of the PADI work with other teachers. In two places throughout the year, she particularly saw the potential for Tina to influence other teachers on the grade. She sang her praises twice—first in December, then in the spring.

I think that this could, could really knock the socks off of everybody, particularly herself, and when she reflects on what she learned from all this, she'll share that with the other teachers. They may choose to do something like that, incorporate that into the unit for the following year. (Lynne Int 2: 10937-11229)

I think she's poised to do it. She's got the respect from the teachers. They know how talented she is, and it's just a matter of finding the time and the school year where they're willing to try one more new thing and in addition to all the things they want to do. We have such a set schedule, with pacing calendars for every participant. (Lynne Int 4: 11895-12103)

Lynne saw Tina's work in particular as a lever for change. This is in contrast to how Tina saw herself and her fellow teachers for much of the year. Also, Lynne's seemed less focused than Charles in her reasons for sharing the work. She hoped that Tina's work would inspire the others to join the work. This would result in every fourth grader having the experience, but it does not speak to whether or not the other classroom teachers would work the same way Amy did on parallel projects in social studies. Lynne almost caught herself at the end by pointing out that the teachers may be too busy to choose to do the work.

The facilitators really diverged in terms of how they saw the future of the projects in their schools. One, Charles, saw the project continuing into the next year and growing into the rest of the school. The other, Lynne, in what seemed to be a contradiction of her statement above, was ending the PADI work and moving the school onto new initiatives

with the fourth grade and the rest of the faculty. This difference had a significant impact on the direction of the two schools and could be a significant finding.

Charles, said this of the work moving forward:

And what I, what I really want people to understand is, you know, first of all, picking one thing. I think that, just going back to the fourth grade group, which did such great work, but it's just focusing on the one thing. What is that one key idea that will have the biggest impact on their students, but that you can clearly define and measure? (Charles Int 4: 16589-16936)

Here, we see his continued focus on measurement and its potential to impact students. He wanted to harness to potential of locally created performance assessments to transform his school.

Lynne also wanted to see change in her building. She, too, was looking to add a new project in June.

Now we're doing the SVO work, and now I'm also adding in the Invention Convention for everybody at the end of the year, and I think that we're still trying to put so much into the calendar and we're not taking anything out. So I think for her, she can do it. She just has to get the others to find the time in the year where they're willing to do that with her. (Lynne Int 4: 12103-12595)

A contrast between the two facilitators emerges here. While both intended to support their teachers, Charles planned to do so by leveraging the PADI work toward other projects in the school. He wanted to focus the time spent sharing work like PADI by looking specifically through the lens of measurement ("The assessment loop") and its ability to transform instruction. Lynne was less focused. She intended to move on from PADI to other projects that she hoped would incorporate some of the assessment work. She seemed to leave it up to the teachers to "fit it in" if they could find where the other 4th grade teachers were "willing" to do it. She viewed sharing as a chance to inspire others, but did not plan to continue to facilitate the work to expand it into the rest of the grade.

Conclusion

This Findings chapter set out to tell the story of two teams from two schools both working to develop and implement local performance assessments in social studies. Their goal was to use existing standards to identify the knowledge, skills, and dispositions they valued and then develop activities and assessments to help teachers measure what they value. The findings uncovered a story of two teams attempting that work, diverging from the original plan, but also emerging from the process describing their excitement and passion for the work. Along the way they also recognized that they made mistakes, and they expressed a commitment to improve. So, while both teams did not do exactly what this study assumed they would do, the findings nevertheless reveal the inner workings of the development, implementation, and assessment of new performance assessment activities as expressed through teachers talk along with some triangulation from the facilitators' talk, the documents, and direct observations.

Both teams revealed where real barriers to the creation of local performance assessments exist and where support is needed and welcomed. Finally, they demonstrated the roles that school leaders play in facilitating the work and its potential to impact schools.

For both teams, the primary data source for their decisions rested in classroom observations during lessons and assessment activities. They also revealed ways in which both teams talk about how their understanding and practice around assessment were influenced by the work. In the case of one teacher, that influence pointed to changes in her understanding and practice around assessment. In other cases, the influence took the form of deepening and strengthening their previous understanding. These findings did not find conclusively any direct causation where the PADI work itself could be shown as the cause of changes to understanding or practice. Nor could the “local” nature of the work be tied to changes in understanding or practice while it did contribute to talk around enthusiasm for and value of the work. With that said, however, differences in the experiences between the

two teams were noted. In particular, the impact of the more focused goals of the Shore School team around the creation and use of rubrics combined with the more hands-on approach of the facilitator did result in differences between the two teams.

Chapter V

DISCUSSION OF FINDINGS

This study has attempted to understand two problems not currently addressed in the literature. First, it is not clear how data derived from locally created performance assessments actually influence teacher practice. Second, while “formative assessment processes”—formative assessments, performance tasks, and performance assessments—have been developed and researched for two decades, only recently is this work being attempted within the context of high-stakes accountability. In the process, this study has examined how two theories of action interact and how they influence classroom practice. The first theory of action, known as “bureaucratic accountability,” which gives rise to “accountability assessment,” assumes that practice is changed through external monitoring, sanctions, and rewards. The second theory of action, known as “professional accountability,” which gives rise to “formative assessment processes,” claims that knowledge can be built and changes in practice can be induced through local, collaborative assessment practices. The study looks deeper into the potential strengths and limitations of “formative assessment processes” as a mechanism that can induce reform by “building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches” (Ancess & Darling-Hammond, 1996, p. 57).

This chapter will explore the findings from the case study in four moves. First, a discussion will explore the findings as they relate to the literature reviewed in Chapter II. These findings include challenges to the assumptions made in data use literature, understanding the role of accountability in the creation and use of local assessments, as well as the strengths and limitations of “professional accountability” to understand the impact of local assessment development on individuals and organizations. Second, this chapter will suggest limitations to the study and case findings. In particular, it explores the limits the methodology had on evidence of changes in teacher practice, teacher understanding, and organizational change around assessment. As a third step, I will return to my original dual stance as a practitioner and as a researcher in order to offer conclusions relating to different aspects of my work in schools and in scholarly research. In particular, these conclusions will explore the role of “localness,” the relationship between bureaucratic and professional accountability, as well as a deeper look into the strengths and limitations of “professional accountability” and “formative assessment processes” to understand assessment practice in classrooms and in the reform process. Finally, the concluding section of this discussion chapter will suggest how these findings have implications for practice, for future research, and for policy.

Discussion of Findings

Data Use Assumptions Challenged

Findings having to do with data use are the primary focus of research question 1, which asked, “How do teachers who participate in the work of creating local performance assessments both as individuals and as members of a group use the data produced from these assessments to inform their practice or change their curriculum?” While the findings do resonate with the literature on data use and with the assumptions behind both theories of action, they also challenge some assumptions about the nature and patterns of data use in

three ways. Assumptions are challenged in terms of where and when data are used by teachers, how data are used, as well as what constitutes the data teachers use.

Where and When Data are Used: In Classrooms

In contrast to the first theory of action's focus on the use of data from end-of-unit/year assessments to guide practice, the teachers in this study rarely used such summative data. In fact, they didn't even do the post-assessments until late in the year. However, consistent with Hatch and Wallenstein (2016) and Marsh (2012), they did show evidence of data use in other ways, both making observations while students took the assessments and collecting and using data throughout the instructional process. Teams at both the Shore School and the Hill School made adjustments to their PADI units by extending the lessons throughout the year. They made these decisions in ways consistent with Marsh's (2012) analysis, which showed that assessing instructional practice and assessing the effectiveness of adjustments in instructional practice are two significant examples of instructional data use. This was seen in Lily's observation of classwork at the Shore School as well as observations from members of the Hill School team. The fact that teams used data directly from classroom observations is also consistent with Young and Kim's (2010) observation that teachers do not always use data purely for measurement. Formative assessment processes shift the use of assessment away from performance indicators to actual performances, from measurement toward "shaping instruction as it unfolds, gauging student achievement, and evaluating curriculum" (p. 5).

The first observation made in the case that demonstrates the Suburban Woods District teachers' lack of conventional data use came when neither team actually produced the summative assessments in their original PADI plans. Put another way, the actual creation of conventional data use took a backseat to other work in their PADI projects. The second observation saw that both teams extended instruction and delayed assessment in place of developing new performance assessments. This skews the theory of performance

assessments that suggests that data are created and then instructional decisions are made from regular classroom observations produced internally to help students, as well as from summative performances of learning. While the intention of summative performances is not only to demonstrate learning but also influence what skills are learned (Darling-Hammond, 2008; Martin-Kniep, 2009), the teachers primarily look for those demonstrations of learning and skill building in classroom work. It seems that during the first year, both teams acted on classroom observation data instead of waiting for data derived from scoring assessments. This final point again reinforces Little's (2012) observation that we are just beginning to understand how "data driven decision making" actually plays out "the system of everyday practice that makes up schooling" (p. 143).

What about instances where the teachers did talk about summative assessments? Building on the idea of assessment as a process, in the rare instances when actual assessment data derived from summative, end-of-year assessment including rubrics, the data were primarily used toward the end of the case and primarily by one school. Specifically, the team from the Shore School did demonstrate work toward developing and using those rubrics, applying them toward assessing the final project.

Teachers from the Hill School, however, never really use data derived from the summative assessments, but Amy claims to aspire to do so in the future. Thus, the only data used by the Hill School team in the first year were derived from classroom observations and impressions from observing students' work. This difference between the two schools may reflect the behavior of the facilitator—a difference that will be explored further in a later section.

How Data are Used by Teachers

The act of delaying summative assessment and extending classroom activities may be a reflection of the way the teachers are making sense of the new skills and standards they are trying to teach and therefore assess. Seen in that light, the delay may not be a deliberate

act of resistance to standards-based assessment but rather part of the process along the way to understanding what the standards look like in action in the classroom. Consistent with this view, several of the teachers and one of the facilitators describe their first year as a “stage” leading toward more changes. However, it is not clear whether this is an actual stage in development or simply a retrospective rationale to explain why they veered from their original plans in the PADI blueprint.

In studies related to how teachers make sense of their classrooms, Kennedy (2005) also adds to the complexity of teacher decision-making. Kennedy found that most teachers described a pattern of thinking that started with an “intention” and was followed by a decision to act based on their understanding of a specific moment in class. Both teacher intentions and decisions to act were based on pre-existing beliefs and values as well as new ideas that might be introduced by a reform process. In the classroom, teachers make observations throughout their teaching based on goals and intentions. Those observations are the “data” with which they make decision to act. In Lily’s case, she is concerned with the unevenness of student achievement at first, where some kids seem to be “getting it” while others do not. Her goal or intention, then, is for every student to “get it” before moving on. This extension of instruction could include differentiation or simply applying the concepts into additional lessons and units to give struggling students more time. In Amy’s case, she also sees students struggle with their understanding of the “concept of beliefs” or a culture’s worldview. She chooses to give students more time to work with this abstract concept before she can move forward with the idea that beliefs are shaped by environment. In both cases, these teachers are making judgments about the pace of student learning and are making adjustments accordingly prior to moving toward summative assessment.

What Constitutes Data: Observations Trusted over Numbers

Another explanation for teacher talk and behavior in this case has to do with what information teachers trust on the surface. Recalling the observations on trust made by

Porter (1995) and Taubman (2009) in their critique of “bureaucratic accountability,” data-driven conversations often expect teachers to trust in numbers over other factors. Unlike the data-driven conversations critiqued by Porter and Taubman, the Shore Hill teachers’ behaviors suggest that they do not trust the numbers derived from tests and rubrics. On the surface, these patterns of diverging from the original PADI plan as well as delaying summative assessment could suggest an avoidance of data use. But one can conclude avoidance only if one is to define data through the lens of “bureaucratic accountability”—narrowly as information that comes from the scores and grades produced from assessments alone. But through the lens of “professional accountability” and as Marsh (2012) and Young and Kim (2010) suggest, data take many forms.

Based on what teachers and facilitators reported during the case, it may be more important that the teachers are focused on the strengths and limitations of their lessons and their assessment instruments in the first year—i.e., what proficiency looks like and how best to see it—rather than counting the number of students who achieve proficiency. Here again, not only is assessment a process that is stretched beyond a single event, but also the creation of local assessments is a process that is not complete after the training sessions are done (Martin-Kniep, 2014). That creation process extends beyond the summer, and into the cycle of instruction, pre-assessment, continued or extended instruction, summative assessment activities, and perhaps even beyond that. It is in these moves that the actions of both teams seem to challenge assumptions in the data use literature. While Marsh (2012) and Young and Kim (2010) consider classroom observations during lessons and assessment activities, they always do so in combination with summative assessments. This limited trust in numbers and scores can be part of the process of teachers as they work to align what they measure with what they value. In the process, teachers may trust in classroom observations over summative scores at first. But in this case study, the more balanced uses of classroom and summative data did not emerge.

In sum, the finding from Chapter IV that teachers tended to delay summative work and extend their classroom instruction may suggest a first step toward changing their understanding of assessment—what some of the teachers and one of the facilitators describe as a “stage in a process.” The observation made about education reform by LeMahieu and Eresh (1995) more than two decade ago seems to hold true: “The classroom is the crucible” (p. 126). However, these findings could also be an example of teachers making sense of their intentions and decisions to act based on their years of classroom experience. In other words, it could simply be the reinforcement of what they already believe about the role of classroom processes over conventional summative testing.

Influences of Accountability on Teacher Agency

The teacher descriptions of the influences of accountability on their work touch on research questions 3 and 4 and are discussed as a barrier to assessment creation and use in Chapter IV. These descriptions reflect the teachers’ sense of agency around the work of creating local assessments. These findings also speak to an important question underlying this study: To what extent does the “localness” of work on assessment affect teachers, and to what extent to do outside forces impact the creation and use of those local assessments? These influences described by both Suburban Woods District teams, in particular, resonate inconsistently with the conclusion of many scholars that the “bureaucratic accountability” and related “outsourced” curriculum leaves teacher agency severely limited and constrained (Darling-Hammond, 1995; Hamilton, 2002; Linn, 2000; O’Day, 2002; Sanger, 2012; Wills & Stanholtz, 2009). At times, teachers express feeling unaffected by accountability, while at others, they see accountability as a barrier to when and how they can work on local assessments.

The Limited Influences of High-stakes Testing

The findings about barriers to the work do not appear to align directly with the concerns raised by “bureaucratic accountability,” because the teachers state directly that their leaders and their community do not hold them accountable in the ways described in the first theory of action of “bureaucratic accountability,” “[which] seeks to induce change through extrinsic rewards and sanctions for both schools and students on the assumptions that the fundamental problem is a lack of will to change on the part of educators” (Ancess & Darling-Hammond, 1996, p. 57). If anything, teachers at Suburban Woods describe the opposite when they speak of the culture of Suburban Woods. They do not blame the tests themselves, because they believe that the District and the community do not value them.

Teachers from both schools claimed often that the District generally and the PADI experience specifically welcome them to be the designers of curriculum and assessment. In other words, the “localness” of curriculum creation seems not to be constrained by exterior structures. Gipps (1999) suggests that power and control are two important factors when considering assessments in the context of teacher agency. Her worst fear, also known as “Stage Three of Control,” where teachers are measured by the external curriculum they teach, is never realized in Suburban Woods. As such, the teachers’ description of the barrier of testing might suggest that the teachers have retained at least some power and control over curriculum and are not entirely constrained by state testing. This observation is consistent with the work of Mourished et al. (2010), who found that, globally, the highest performing school systems shift away from high-stakes accountability toward trusting and supporting teachers. In this way, the “bureaucratic accountability” found in the first theory of action may not be a direct influence at Suburban Woods.

Indirect Influences of Accountability

However, despite the claim that the accountability from state testing is not a direct barrier to the work, two other findings suggest that the current accountability structures have impacted or “constrained” (Wills & Stanholtz, 2009) the PADI work indirectly and in

ways described in the literature. Taken together, these two indirect barriers of mandated curriculum and the timing of new units around state testing resonate with Ball's (2006) notion of "post-professional."

The practitioner is left or held responsible for their performance, but not for the judgment as to whether that performance is "right" or "appropriate," but rather whether it meets audit criteria.... Within all this, teachers have lost the possibility of claims to respect except in terms of performance. (p. 669)

The first indirect influence is the timing of the tests. The phrase "after testing" and the freedom of June were raised multiple times. Both teams chose to delay their final performance assessment activities until June for this reason. The idea that June is an open month speaks to the pressures in the other months to comply with external demands. The second indirect influence mentioned by the teachers is the increase in mandated curriculum. In both schools, the experienced teachers see mandated units in literacy, in math, and to some extent in science as new features in the last 5-10 years. Both teams refer to the "old way" of creating curriculum, where individual teachers and grade-level teams had much more control over how units were created around New York State Standards. They often describe the new way as curriculum defined by "pacing calendars" and numbers of units to be covered. Amy from Hill School describes the challenge as the "constraints of the day." Charles, the Shore School facilitator, implies that teachers struggle to prioritize among the demands placed on them by mandated curriculum as defined by the pacing calendar. Lynn recognizes the scripted nature of new curriculum and, like Amy, sees the idea of teacher-created lessons and units as a move "back" to curriculum before the current accountability environment.

Within these findings is the suggestion that even if the teachers wanted to change their practice to do more work like the PADI performance assessments, there simply is not room to do so. Their work is governed by predetermined performances set out by a pacing calendar. In Gipps's (1999) words, power and control have shifted not by state testing directly, but rather by the increase in non-local externally created curriculum. This idea is

perhaps the strongest argument against concluding that the PADI work in the first year is a “stage” toward greater change. Even if the teachers wanted to change more or create curriculum that is more “local,” even if the teachers see the delay as the first stage of a process and are using it to make sense and understand how to use the assessments, the lack of time and bureaucratic pressures will continue to influence their use of the assessments.

“Professional Accountability,” which gives rise to “formative assessment processes,” suggests that teachers need to have the agency—ability and permission—to explore what student performances are “right” and “appropriate,” and that by doing so, through experimentation, they will create better learning experiences for students than outsourced standardized assessments. But, according to both teams, that is only likely to occur in untested subjects like social studies and in months like June, when the pressure of testing is off. In this way, the range of “local” work around curriculum and assessment is limited. It should also be noted here that this study has only revealed how teachers feel and react to these constraints. It has not set out to demonstrate that the mandated curriculum or the pacing calendar is producing inferior results. In this way, the barriers discussed above speak more to the teachers’ sense of agency around the work rather than what they actually do around the local assessments, which seems to be constrained by the mandates of curriculum and testing calendars. Furthermore, despite how both facilitators speak of the work going forward, there is no evidence that the work by these two teams is impacting the organization from the bottom up—a point that is explored later.

This study set out to understand the interplay between the first and second theory of action—the tension between bureaucratic and professional accountability. The findings around barriers suggest that there is not as much of the direct interplay, as I suspected based on the “all good/all bad” (Scheurich, Skrla, & Johnson, 2000) view of the anti-accountability literature would suggest. Granted, this finding may be unique to Suburban Woods and districts like it where some degree of “localness” is still welcome when it comes to curriculum and assessment design. However, even here, the indirect

influences of accountability are apparent. Even in systems where teachers are trusted and seen as a source of system improvement (Mourished et al., 2010), the impact of accountability has not been eliminated.

Formative Assessment Processes Explored: The Impact of “Localness”

Moving from the interplay between the two theories of action, this next discussion looks into the findings that push more deeply into the second theory of action in three ways. First, it looks at the positive outcomes related to “localness” assumed by “professional accountability.” Second, it looks at “professional accountability’s” relationship with agency. Finally, it looks at the shortcomings of “professional accountability” with regard to “localness,” agency, and the potential for collective or organizational change.

The Positive Impacts of “Localness”

“Using their minds well” and validity. The “professional accountability” literature comes into play in this study when the teachers talk about how the development of local assessments influences their understanding. It does so not by changing their views of assessments but by reinforcing what they value or understand about good curriculum and assessment. The PADI work appears to be reinforcing a kind of assessment and performance that the Suburban Woods teachers value. It is also the kind of assessment that is described in the literature as encouraging students to be “using their minds well” (Wolf et al., 1991, p. 32). For example, Lily from the Shore School and Amy and Tina from the Hill School describe their work in terms of two of the three elements of O’Day’s (2002) notion of “professional accountability.” “First it is centered on the process of instruction.... Second, much of the focus of professional accountability concerns ensuring that educators acquire and apply the knowledge and skills needed for effective practice” (p. 20). The second theory

of action seeks validity beyond simple test performance, or what Messick (1995) describes as the impact assessments have on teacher practice and on student learning.

Authentic tasks. The way the teachers talk about the purposes of the lessons and activities developed through the PADI work meets two important conditions for valid assessments to promote learning put forth by Glaser and Silver (1994) and Palmer-Wolf and Richardson (1996): (1) the assessments created are relevant outcomes of the program (Shared Value Outcomes in the case of Suburban Woods), and (2) the assessment is “curriculum-embedded” or integrated into, not separate from, the instructional program. Again, this resonates with the literature from the “professional accountability.” “In these evaluations, students are asked to write, to read, and to solve problems in genuine rather than artificial ways” (Wolf et al., 1991, p. 55). Lily sees a direct connection between her practice and how the students are thinking and communicating. She is making a connection between the resources she shares with students and the ways students will use them. In addition, she wants to assess student thinking and communicating. Assessment, therefore, is more than just scoring what information a child has learned. Lily sees her work in assessment very much in formative terms, where assessment is largely a reflective process-making decision based on where the students are and where they need to go. Again, the Suburban Woods teachers value assessment as a process differently from the policymakers’ assumptions of teach, test, reteach. Linn and Baker (1984, in Wolf et al., 1991) describe six qualities of classroom-based formative assessment processes: “open-ended tasks, higher order, complex skills, extended periods of time for performance, group performance, student and teacher choice of tasks ... judgmental scoring” (pp. 87-88). Similarly, Tina from the Hill describes a complex skill that she is assessing and teaching into. Amy also describes the students embracing the abstract thinking that she describes twice as “the evolution of an idea.” All of these examples point to teachers reporting the positive impact of “local” control over the creation and use of assessments through the PADI project. In addition to

being unconstrained by time and testing, they describe a validity based on the purpose of their work.

Empowerment from design and making judgments. The Shore Hill teachers report that they also seek validity in assessment by what students are asked to do. The teachers at Suburban Woods seem to be aligned with “professional accountability” in their descriptions of their ability to make decisions and judgments. This aligns with Ball’s (2009) notion of an “authentic professional,” whereby teachers are allowed to and expected to make complex judgments. This begins when Charles and Lynne talk about the attraction for teachers of being the “designers” of assessments through the PADI project. In this way, their work avoids the “meaninglessness” of standardized testing described by Wolf et al. (1991), as well as “the constriction of teacher’s professional judgment ... a system of assessment that is curriculum dependent. Such assessments reconnect effort, teaching, assessments and results” (p. 19). In addition to teacher judgment, Campbell (2012) describes how teachers who have the agency to make judgments can contribute to increased student agency. Though rarely observed in my findings and specific to judgments around assessments, the students at the Shore School began to use assessment rubrics in their work as the teachers developed and implemented new rubrics. The PADI efforts to deepen teachers’ understanding of standards should also be emphasized as a step toward improving instructional outcomes (Darling-Hammond, 2006).

“Localness” as a Mechanism for Changes in Practice and Outcomes

While the findings were inconclusive in demonstrating evidence of changes in practice and in understanding caused by the PADI work or its “local” nature, they do make important observations around how “professional accountability” and its “formative assessment processes” relate to the literature of agency and sense making. For example, Priestley et al. (2012) and Birdwell-Mitchell (2015) posit that the idea of teacher agency can result both in change and in resistance to change, and Weick (1995) and Coburn (2006)

both posit that sense making is a factor in how teachers react to policy and new initiatives in relation to their everyday practice. These observations come into focus by first recalling that “professional accountability” stresses the importance of assessments and assessment-based decisions made by teachers as they support and help define the learning that occurs in classrooms. This approach is put succinctly by Aness and Darling-Hammond (1996), who say that “professional accountability”

seeks to induce change by building knowledge among school practitioners and parents about alternative methods and by stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling and to experiment with new approaches. (p. 57)

While this study may not have clearly demonstrated that “reforms” have been induced by the PADI work, the evidence of “knowledge building” and the use of “alternative methods” is noteworthy. Specific to the question of change, only two of the four subjects talked directly to that question. In addition, direct evidence of change described by a subject occurred only in one of the four subjects. Tina and Amy from the Hill School are the only two teachers to talk directly about changes or the lack of changes in their thinking and in their practice. And they do so as opposites. Both are examples of “agency orientation.” Tina speaks about how the PADI process has changed her thinking about assessment and about her practice. Specifically, Tina talks about changing what she is assessing and how she has shifted the focus of her instruction from product to process. In contrast, Amy is quite insistent that her thinking and her practice have not changed. This difference in outcomes between two teachers working on the same project in the same school is in line with the literature on agency and change. For example, Priestley et al. (2012) found that the teachers they observed achieved agency in their work unevenly and that this unevenness was not completely attributable to the single factor of external policy demands. They summarized their conclusions as follows: “The analysis does not explain why one teacher with rich prior experience and strongly held views about education was able to translate this strongly into her teaching, whereas the other was less successful” (pp. 210-211).

In short, some teachers who achieve agency will change their practice, while others may be resistant to change. Thus, accountability structures alone cannot explain why teachers do or do not change their practice or what level of agency they develop or demonstrate. Similarly, Birdwell-Mitchell (2015) concurs that the achievement of agency does not always equate to change in practice. Because teacher agency is influenced by social and environmental contexts, “teacher agency can both change and maintain institutional instructional practices in schools” (p. 140). In this context, it could be argued that we have not actually revealed new mechanisms but have just explained old mechanisms through the newer discourse of data use.

Influences of Environment on Teacher Agency

Biesta and Tedder (2006) talk about how the achievement of agency is an ecological phenomenon, in that it is highly contextual, not to the teacher, but to their environment. This phenomenon is seen through the impact of culture and facilitation on teacher agency at Suburban Woods.

Culture and agency. Agency is defined by many scholars in terms of permission and ability to act (Birdwell-Mitchell, 2015; Campbell, 2012; Elmore, 2009; Popkewitz, 2008). Since the culture of a school will define the actions of teachers, the culture of Suburban Woods clearly plays a role in helping teachers achieve a sense of agency around the work of local assessment creation and use. One possible explanation for the difference in the level of teacher agency in Suburban Woods, therefore, could be the culture of the district—specifically the district’s emphasis on an “agency orientation” for all teachers. The notion of culture is also expressed by both teams when they discuss the supportive nature of the district and of their principals. Again, this description of culture aligns with a study of school systems globally (Mourished et al., 2010), where teachers are given more autonomy and decision-making authority in the highest functioning systems. For example, both teams express an overall sense of support to do the PADI work. Specifically, both teams

acknowledge that their facilitators give them time within the schedule to work together. In addition, they express how parents appreciate the kinds of projects and performances that students participate in. It was even noted that the Suburban Woods Board of Education passed a resolution stating that the community values multiple measures of performance and achievement and prefers them to standardized testing.

The power of the Suburban Woods culture as a source of agency and change can also be seen in one school by comparing someone new to the culture with a veteran teacher. In the Hill School, Tina is the newest teacher; at Suburban Woods, of all four teachers in the case, Amy is the most senior. In talking about their time in the district, both Tina and Amy invoke “agency orientation” (Sloan, 2006) language as they make sense of their experiences. “Agency orientation,” like “professional accountability” in the second theory of action, places the focus on student learning as well as on teacher learning through the process of instruction and reflection on student work—not on test results. Tina describes the differences between her old district and her new district. In doing so, Tina attributes (or makes sense of) her ability to act on assessment decisions in part to the PADI training, but also to the culture of her new district, which encourages teachers to seek outcomes for students beyond test scores. In this way, being new to the district causes Tina to see more changes to her practice—perhaps because she had farther to change. In contrast, Amy makes sense of the PADI experience differently. Unlike Tina, Amy has been in the district over 15 years longer than Tina and has often experienced an agency orientation toward curriculum and assessment. Amy explains that in the past she experienced even more agency. She sees the PADI projects as a return to the way she used to work.

Impact of differences in facilitation on agency. When it comes to agency building in the specific context of assessment practice, the findings point to the impact of differences in support among the two facilitators. The teachers from Shore School describe several ways that the facilitator supports specific aspects of the work. The teachers from the Shore School also describe Charles analyzing the data with them both during the pre-assessment

and the post-assessment. One teacher, Lily, had him in to observe lessons. He also brought Giselle Martin-Kniep to observe and work with the team in their classrooms. But it also seems that the facilitation in the Shore School goes farther in the ways Charles participates in the planning and assessment work directly. For example, he participated in lessons that focused on introducing the rubric to students. Where Lynne from the Hill School simply creates the space for the work to occur, Charles has not only created that space, but also, he has helped the team scaffold the goals and the work itself through his focus on rubrics and clearer standards to measure growth to deepen the work. Moss (2012) describes the role of facilitators as one of three possible mechanisms impacting teacher agency in and around assessment use. Charles's approach may also explain why the teacher from the Shore School used summative data, standards, and rubrics more than the other team.

In both of these examples, "professional accountability," in conjunction with the culture of Suburban Woods and the specific supports of facilitators, is arguably an environmental source of agency. When the context of the Suburban Woods culture is combined with the other positive aspects of the second theory of action—the validity of purpose and the increase in teacher and student engagement and judgment—the findings could be seen as drivers of reform. However, while "professional accountability" as applied in Suburban Woods may have demonstrated the achievement of agency, it has not been proven to be the cause of any changes in practice. Just as the discourse on data could not prove a linear connection between data use and better decision making (Coburn & Turner, 2012), a direct connection cannot be drawn between agency achieved and better decision making. The findings have not taken us into the "black box" (William, 2004) to truly understand the "complex mechanisms" (Moss, 2012) at play in assessment practice reform.

Limited Impact of “Professional Accountability” on Collectives and Organizations

So far, the discussion of “professional accountability” and agency has focused on individuals. As mentioned earlier, it has focused on only the first two of the three elements of professional accountability outlined by O’Day (2002)—namely, the process of instruction and teachers acquiring and applying knowledge. This final section discusses the third element and explores the impact of local assessment development on collective groups and on the organization as a whole. In the language of O’Day(2002), the third element of professional accountability is “the norms of professional interchange” (p. 20), or how teachers share and exchange ideas as they grow their instructional knowledge and practice. This idea is also expressed clearly by Aness and Darling-Hammond (1996), who say that “professional accountability” “seeks to induce change by ... stimulating organizational rethinking through opportunities to work together on a design of teaching and schooling” (p. 57).

Just as the notion of collective is in O’Day’s (2002) definition, the same idea is repeated by scholars who write about “professional accountability,” sense making, and agency. The findings of the case revealed very little of this collective or collaborative work either in what I observed or in what teachers reported. For example, Wolf et al. (1991), talk of “intense discussions of standards and evidence among all of the parties” (p. 59), and Darling-Hammond (1995, 2009) talks about the sharing of knowledge collectively to inform systems. Weick (1995) talks of collective sense making in addition to individual sense making. Even data use scholars discuss collective thinking, collaboration, and organization change as important mechanisms for creating outcomes (Coburn & Turner; Moss, 2012). Campbell (2012) and Birdwell-Mitchell (2015) talk of the agency of groups and collectives. For example, Campbell talks about how group agency can drive changes in practice. Likewise, collective actions are also one of Birdwell-Mitchell’s (2016) proposed mechanisms for agency in school that create the conditions for reform. Despite this

emphasis on the potential impact of data and assessment use on groups and organizations, this study found little evidence of such collective impact when the teachers in this study were specifically asked about sharing their work with other colleagues. When asked about sharing with others, teachers from both teams did not see how they would do so. Some went as far as to suggest that other teachers in their grade or in their school might not be interested in the work. Put simply, teachers from both teams were comfortable reporting to me on the impact PADI had on their practice but did not make connections to the practice of other teachers. This observation is somewhat in contrast to both facilitators who talked at least as an aspiration of sharing the work at faculty meetings and with other teacher teams.

There is only one example in my findings that touches on this “mechanism” of agency that Birdwell-Mitchell (2016) emphasizes—the concept of “peer learning” (p. 141). Birdwell-Mitchell’s mechanism can be applied to explain why Tina might have changed while Amy did not. From Birdwell-Mitchell’s perspective, changes in practice are more likely to take place when peer-to-peer learning is occurring. Amy’s position is clear in her statements; she believes that she is not learning anything new through PADI. Tina seems more open to explaining her learning both in terms of her assessment practice as well as seeing more connections between the work done in her art class with the work done by students in social studies. She is learning from working with her peer, Amy, in interdisciplinary context. It is interesting to note again that Tina sees these connections as new or as changes while Amy recognizes them as a return to previous practices. In this way, Amy does not see them as new.

The idea of collective learning for organizational change also touches on the role of facilitation (Moss, 2012) in the PADI project. While the facilitators themselves are not the direct focus of this study, their actions and the impact of their actions are discussed by both teams. All four teachers talked about how the facilitators helped each pair work together. But the teachers did not discuss how or if the facilitators helped them share their work with others or encourage others to learn from their work on local assessments. The only time

other teachers were mentioned was by both facilitators, and then it was mostly aspirational. For example, Lynne talked about sharing the project with the other fourth grade teachers. Charles had already discussed the PADI projects from several teams at a faculty meeting and said he planned to do more the next year. In the end, the findings are mostly silent to the idea of collective learning and organizational change, even though the PADI project and the study design intended to include this work, and so much of the literature on “professional accountability,” sense making, agency, and assessment speaks to its potential.

Limitations

I have placed a discussion of this study’s limitations into three categories. The first category of limitations reflects the issues in the overall design and execution of the case study. Second, there are limitations to the generalizability of the study based on the selection of the Suburban Woods District itself. Third, there is a brief discussion of the limitations of the theoretical framework that guides the examinations of data use and agency. In this section, I explore these conceptual limitations from the perspectives of a practitioner and a researcher.

Limitations of Design

This study’s design is built on the specific recommendations from other research studies of data use in order to address potential shortcomings of earlier research into assessment practices. One obvious set of limitations comes in the ways the execution of the case fell short of the design. The study’s data sources and the order of their collection, as well as the decision to add the direct observation of teacher work sessions, came directly from the data use literature. Spillane (2011) and Warren-Little (2012) both warn against studies that rely simply on “ex situ accounts of practice,” such as surveys and interviews, to gain insight into teachers’ use of data and assessments to inform practice. But the majority

of my findings were based on interview data and not in-depth observations over time. Although I did carry out some direct observations, my position in the study as a principal from outside the district may have contributed to my inability to observe a sufficient number of classroom work, student work product, and teacher work sessions. Put another way, a different researcher might have been able to spend more time embedded in the school and in the classrooms and faculty workspaces and perhaps gain a clearer picture of the teachers' choices as they made sense of their work. This limitation was compounded by the lack of student work products collected. Without seeing the student work that the teams were discussing, it was difficult to assess how the teachers were making sense of student growth. Therefore, a stronger focus on collecting student work outcomes would have been more in line with the kind of observations outlined in the literature.

The second design limitation is the time span of the study: simply put, changes in practice often take longer than a single school year to take hold. While one year did allow me to follow a single PADI assessment cycle, extending the observations into a second cycle of assessment creation and use would have been productive. In particular, I might have been able to see if the teams actually made the changes in practice in response to the mistakes they identified in the first cycle.

An additional limitation in design relates to changes I could have made in my data sample. First, the nature of the collaborations was small—pairs—and I did not see a group dynamic on the scale of a team of teachers on a grade level. Also, three teachers are similar in profile (20+ years teaching in the district). Only one is new. A broader, more diverse sample of teachers and PADI projects may have shown more differences among teacher learning and practice. The limited baseline data I was able to collect also made it difficult to demonstrate definitive changes in practice and teachers' understanding. Had I gathered more baseline data on teacher practice prior to the PADI training, I might have been able to see more detailed examples of how the teacher practices and understanding were or were not changing. Again, the fact that both teams did not create original performance

assessment tasks is a departure from the norm in the four years of PADI projects. This anomaly in my data sample contributed to my limited perspective as well as to my struggle to collect student outcome evidence. Had the study spanned more PADI teams—especially teams that completed and gave original performance assessments—my finding may have been broader and perhaps more conclusive in terms of the impact of assessment creation and use on teacher practice.

Taken together, these design limitations explain the extent to which the “processes” described by the teachers might actually result in long-term changes to practice. While the practitioner in me was excited to learn of these processes, and while the facilitators saw them as changes in practice, the teachers’ descriptions over the course of one year were not sufficient to provide definitive and triangulated evidence of change. This shortfall was particularly evident in the area of collective or organizational changes. To truly see this phenomenon, research would need to intentionally focus on larger groups of teachers and on specific actions where they are observed working together over longer periods of time.

Limits to Generalizability

A second set of limitations are those limits to generalizability created by the specific profile of the Suburban Woods District. First, the combination of the district’s funding and resources along with high-performing students may skew the results. For example, the district can pay for multiple teams of teachers to be trained on multiple days (\$5000+ per team). Perhaps the resources and supports that teachers have in a district like Suburban Woods make it difficult to generalize my findings onto school systems with few resources. The second limitation is the fact that the Suburban Woods district may be unique in its position as being open to alternative assessments. The culture that the school has created, which encourages teacher agency and downplays accountability to testing, may be rare in public schools. Granted, I did choose this district because of its clear stance on testing for accountability, as well as its BOE support of alternative forms of assessment. But these

unique characteristics of the school district did contribute to difficulty in determining the impact of the PADI training and the “localness” of the assessments over other contributing factors around assessment practice. While the strengths of the “localness” of the assessments were uncovered, these qualities may be more attributable to the culture that exists at Suburban Woods. Therefore, it cannot be determined if creating local assessments like the PADI project would have the same result in other school cultures.

The Discourse of Data and Change: Limitations as Seen Through a Post-Structuralist Lens

The third potential limitation in this study comes from the lens of a post-structuralist researcher. I must recognize that the discourse of “data use” may be imposed on the language of the subjects by the process of their participation in the study. Post-structuralists like Luke (1996) and Cruikshank (1999) would acknowledge that I have observed a process at Suburban Woods. But, unlike my suggestions derived through the positivist discourse of data use, they would argue that this process is not actually a new one. They would argue that this process is not unique to data use or assessment but rather is a more common practice of teachers being transformed through the imposition of a new discourse onto or into the classroom.

This critical observation is made despite my attempts to address it methodologically. The open-ended interview style set up by sense making theory to frame questions is designed to allow subjects to name their world and their experiences (Dervin & Clark, 1999). My process tried to mitigate suggestive uses of language in questions so that the vocabulary the participants used reflected their own understanding and that of the PADI training. Nevertheless, the context of the study is clear; the discourse of data use as a “text” was ever-present. The teachers’ discussions reflected the dominant narratives in policies and practices related to data use and assessment. Luke (1996) explains that any text shapes our thinking and positions subjects in relation to power. Assessments as “texts” define “what counts,” but I must also acknowledge that “data use” provides its own “text” or

discourse. That, in turn, may be shaping what my subjects say and also what I, as a researcher and as a practitioner, observe. Therefore, if post-structuralists ask of assessment practice who decides the conditions and categories that exist (Cruikshank, 1999), then the more critical researcher in me is left to question my findings. Specifically, my data use findings suggest that teachers prefer data derived from classroom observations over summative assessments. But teachers have always made classroom observations and adjusted their practice based on those observations. Despite using the discourse or language of data, my subjects, for the most part, may simply be describing actions that most teachers take. As such, my conclusions based on my findings may be far more limited to specific data use behaviors, such as rubric use and an increased focus on specific standards in action.

A second critique from critical theory looks at the conceptual frame of agency. As a practitioner, I was looking for specific changes in assessment practice. Deleuze (1992) would argue that, with its focus on standards and “best practices” working toward the specific goal of improved student outcomes, this study into classroom formative assessment is still very much part of a “culture of control.” Deleuze talks of “molds” versus “modulation” (p. 4). Where the examination, as a technology of discipline, encloses or molds the teacher and the student, the perpetual assessment in the classroom is a modulation—a technology that controls teachers and students in ever-changing forms. Again, agency is not achieved. Thus, like Popkewitz (2008), Deleuze would question if, in the purest philosophical sense of the term *agency*, any reform can claim to foster it. Thus, the uncertainties of this case study’s findings around causation can be traced back to the Agency Paradox found in any change process. In short, they uncover the fact that there may be no connection at all between agency and reform. I must take this into consideration as I examine critically any notion of causation in what I observed or in what I was told by teachers in the study. With the exception of an increased use of rubrics on the part of teachers and students at the Shore School and some changes in the focus of her assessments

on the part of Tina at the Hill School, most of the changes I observed could be attributed to changes in teacher talk more than action. Or their behaviors could not be changes at all, but rather the ordinary actions of the teacher infused with the PADI language. Granted that the talk toward the purpose of their work and to student outcomes was positive, it did not reveal specific actions or mechanisms of change in their practice.

Pushing Back on the Post-structural Critique

To close, I want to put in a final word on these theoretical limitations from the perspective of a practitioner working in schools. While I recognize that the demands of scholarly research make it essential to point out these limitations, as a practitioner in the field, I offer one other theoretical observation by way of pushing back on the post-structuralist critique. As in my pilot study, I found that the subjects of this study appreciated being part of the project. They expressed to me the fact that the questions not only helped them reflect on their work, but also helped them to see their work as valuable to others. This kind of reciprocity or “transformational validity” (Cho & Trent, 2006; Marshall & Rossman, 2010) has value for me as a practitioner. Teachers are not passive agents in the current data discourse. They, too, can choose to see value in these ideas as they seek to improve their practice. The fact that the teachers and facilitators see value in their work and in the opportunity to reflect on that work is a source of validity for practitioners in the field. So, while this case could not demonstrate any definitive examples of changes in practice or understanding that were caused by the creation and use of local assessments, the observations do contribute to the literature on data use and assessment practice and can inform practitioners who are looking for ways to use assessment to improve student outcomes. Perhaps just as important, and while post-structuralists would disagree, the process of participating in this study may have contributed to the growth of four practitioners.

Conclusions

The conclusions that follow are built around four ideas that I set out to understand by studying the creation and use of local assessments. In these conclusions, I draw upon my stance as both a practitioner and a researcher in order to explore how my findings and conclusions relate to contributions to assessment literature as well as to my work in schools.

Data Use Assumptions Challenged

The first conclusion touches on the study's questions around data use. Digging deeper into the technical questions of data use, this study was not expected to demonstrate that "localness"—the locally produced quality of an assessment process—caused changes in teacher behavior around data use. However, the study does yield valuable observations about patterns of data use around locally produced assessments.

This study found that teachers often value the data made from classroom observations and that they do not wait for summative assessment data to make decisions. This can lead to two predictable behaviors. The first is the tendency to extend instruction and delay assessment. The second is their difficulty in demonstrating growth over time due to lack of a clear instructional target or goal set by standards and measured by rubrics. One reason for these behaviors is a gap between teachers' initial understanding of the instructional goal or standard and what they see in the classroom. The "process" of assessment making requires time for teachers to fill these gaps in their understanding. The time taken by teachers as they make sense of their work and make judgments about what steps to take can look like resistance to change.

Whether this process represents a stage of development as several subjects suggest, or if it represents how the subjects make sense of their work, it is clear that these teachers are challenged by "closing the assessment loop." Even with clear goals, instruction, and training, as well as facilitation throughout the year, the work of designing, implementing,

and using local assessments is challenging. It takes constant attention and effort to effect changes to understanding and practice. Teachers described the experience as shaping or aligning their work to standards in order to have clear targets for them and for their students to work toward. This step is essential to help teachers, students, and perhaps even systems to measure student growth over time in complex skills and dispositions valued by the community.

Despite the positive claims made by the teachers in Suburban Woods, it was not clear from this one-year study that any lasting changes to practice have occurred. While some evidence of teacher talk around data is noted, that talk was not connected to many observed changes in practice. Here again, a causal connection is inconclusive. This study could not determine if the PADI work alone or the culture was driving the reported changes in practice.

Influences of Accountability of Formative Assessment Processes

In the spring of 2018, the New York State Legislature passed bill a.10475 “to eliminate the mandate that state created or administered tests be used to determine teacher or principal’s evaluation.” Given what I observed at Suburban Woods as well as where we are at this point in time with New York State accountability structures, it appears that the dichotomy set up at the beginning of this study pitting two theories of action against each other will not play out, at least in New York State. The work at Suburban Woods has demonstrated that places can exist in the current accountability environment where teachers report to have achieved the agency to design, to implement, and, in some instances, to even learn from local assessments. To the extent that the agency—the ability and the permission—to develop local curriculum and assessments has always been a part of the Suburban Woods culture, I cannot conclude that the findings at Suburban Woods are generalizable to all schools.

In Suburban Woods, state accountability requirements through annual testing do not represent an absolute barrier to local curriculum development. More broadly, high-stakes testing for teachers will likely not be a mechanism for, nor a barrier to, reform. The case study at Suburban Woods also revealed that while high-stakes, outsourced testing itself may not be a barrier, other constraints or barriers still exist to the creation and use of local performance assessments. Furthermore, these barriers may have emerged out of the accountability environment of the past two decades. Those two constraints are the increase of mandated curriculum and pacing calendars that are often shaped by the needs and timing of state assessments. While it is true that teachers will no longer be evaluated based on test scores in New York State, this change does not completely eliminate the impact of accountability structures that have emerged in the past two decades. In particular, changes to standards as well as new curriculum aligned to these standards have created new pressures on teachers. These pressures were seen at Suburban Woods in the form of an increase in prescribed curriculum units and strict pacing calendars. These shifting pressures do impact teacher agency in and around the PADI project and may limit local curriculum work to subjects not tested annually by New York State. Therefore, processes like PADI aimed at teaching teachers the formative assessment process will likely still be impacted by externally imposed accountability structures. In addition, moments of agency and therefore experimentation may be limited in terms of (1) the subjects where the assessments are developed, (2) their frequency, and (3) the time of year when they are implemented.

To be clear about a point I made in Chapter I, this case study did not set out to judge the merits of outsourced testing and other accountability structures as good or bad for student learning on their own. The study does, however, point out that these structures do present limits to the scale and scope of what local performance assessments teachers and schools can create should they choose to do so.

Influences of Facilitation of Formative Assessment Processes

Just as this study looked at barriers, it also found ways that the work of local assessment creation was supported. In particular, this study found that when it comes to teachers learning to create and use performance assessments, facilitation matters. This study has shown that there is an expressed need for facilitation in several forms. Time and space for teachers to experiment and make instructional judgments are important but, on their own, do not cause change. It would appear that the types of facilitation also matter when it comes to supporting assessment practice. Facilitation that supports specific practices, such as the development of clear goals around standards and rubrics to measure growth, are most likely to “close the assessment the loop.” From the perspective of research into teacher agency, however, the findings about support do not tell a complete story. While teachers’ talk was mostly positive about facilitation, the findings do not point to specific changes in practice or understanding that arose because of the facilitation. While the teachers in the study did claim to be learning an assessment process, it is not clear from a research perspective what is actually new about that process.

Finally, despite the expressed enthusiasm on the part of both facilitators, this study has not demonstrated that teachers working in teams to develop assessments has any impact on the larger organization. If anything, teachers themselves do not see changing the practice of others to be part of their role in the work.

Professional Accountability and the Role of “Localness”

Where the first two conclusions look primarily through the lens of a practitioner, the third looks more deeply into the strengths and limitations of the second theory of action and therefore into contributions to the research literature in this area. In particular, this third finding focuses on the quality of “localness” that is both a central question to this study and in important ideas at the heart of the “professional accountability” and the “formative assessment processes” it creates.

Positive impacts of localness. In the case of the two teams at Suburban Woods, the ability to design and create performance assessments was motivating and engaging for both teachers and their students. It allowed them to develop activities that supported standards and learning outcomes that were valued by them and by their community. For all four teachers, the work reinforced their thinking in terms of the projects they do with children and in terms of the kind of thinking the children were asked to do. Teachers valued the experiences that asked their students to “use their minds well” (Wolf et al., 1991). In this conclusion, we see the motivational power of curriculum and assessments that are locally created. For many, but not all, of the teachers, this motivation led them to spend more time understanding specific standards and goals around assessment use in their classrooms, all of which suggest a connection between “localness” and agency. In those cases, the teachers themselves and their facilitators reported change and growth. They claimed to be learning a process. This reinforces something Ainess and Darling-Hammond (1996) claim was described almost 80 years ago. More recently, Mourished et al. (2010), also claimed it to be a global phenomenon of high-performing systems. Schools that engage in creating local curriculum and assessment are successful not because they find a single approach or technique that works but because the teachers grow from being engaged in the search for practices and processes that work. They appear to find inherent value in the process of local assessment.

Localness as a mechanism for change. However, despite this positive talk around their work, a closer examination of the findings raises important questions about what we have learned. The research focus in this study set out to understand the mechanisms that drive changes and improvements in assessment practice. In this, “localness” specifically and “professional accountability” more generally fall short in two ways.

First, “local” performances, while engaging, seem to be infrequent events and are made even more so by the barriers of mandated units of study and strict pacing calendars. If the goal of the PADI work was to increase teachers’ work and understanding around

standards, rubrics, and assessments to inform their practice, projects that are limited to a few subjects and to a few times per year may be an inefficient method for teaching teachers how to “close the assessment loop.” This is not to say that local assessments can’t be the source of assessment work, but nothing in the case suggests that “localness” is driving that work either. In this way, both Preistley et al.’s (2012) and Birdwell-Mitchell’s (2016) observations are unchanged. The increased agency achieved through engagement in local assessment does not guarantee changes in assessment practice. Therefore, “localness”—the fact that assessments are locally produced—cannot be the mechanism for reform that the second theory claims it to be.

Second, just as a connection to facilitation and organizational change could not be made, the findings of this case have not demonstrated any causal connection between “localness” and an increase in collective or organizational change. Despite the aspirations of both facilitators, I was not able to observe any changes to the organization. In fact, the teachers did not see the connection between their work and the assessment work of their grade or their school as a whole. In part, these limited findings are caused by limitations in the design of the study. That said, in terms of understanding the mechanisms of reform, this study has not been able to show a connection between “localness,” “professional accountability,” and organizational change. So, while this study as well as Aness and Darling-Hammond (1996) and others may argue that engaging in the creation of local curriculum and assessment is a good practice, it has not put us into the “black box” to better understand the mechanisms that facilitate organizational change around assessment practice.

Over 25 years ago, Haney (1991) suggested that assessment has many purposes, and he offered that “what we need—more than bigger accountability schemes and greater sanctions attached to test results—are better ways to help children learn and better ways to help teachers and parents to help them do so, then we ought to focus time and energy devoted to assessment directly on those ends” (p. 159).

At about the same time, LeMahieu and Eresh (1995) made this second important observation about education reform: “The classroom is the crucible. It is the place where every idea about the best or most powerful form of education finds its expression” (p. 126). While the two quotations above still ring true for the practitioner in me, the researcher in me must conclude that my findings have not pointed to many new understandings to inform the literature on data use or assessment practice by teachers. Without those mechanisms understood, practitioners are left to move forward with the understanding that formative assessment practices based on locally produced performance assessments are still an “article of faith” (Coburn, 2006) more than a proven fact. Nevertheless, the practitioner in me has found that the PADI work observed for over a year in this case study at the Suburban Woods School District has offered a glimpse into a school system that is trying to live up to its values and has produced “outcomes” that Coburn and Turner (2012a) would deem worthy of study and perhaps even of emulation.

Implications

This final discussion keeps in mind the divisions in my thinking as a practitioner and as a researcher where practitioners knowingly make decisions about teaching and learning based on “articles of faith”—based on their experiences and observations of what works in their classrooms and schools without fully understanding the mechanisms driving those choices. In contrast, researchers are driven to look deeper into understanding those mechanisms and must remain skeptical of results that exist only in the ostensive realm of teacher talk when they cannot be observed in the performative realm of practice. In truth, throughout my experiences working in the Department of Curriculum and Teaching, I have strived to use both perspectives to improve my practice and support teachers in the systems where I work. I also maintain the belief I stated at the outset of this study. I believe that teachers and principals alike are neither opposed to nor afraid of accountability and

reform. In fact, as practitioners, we all seek to validate and improve our work.

Understanding the mechanisms in and around assessment practice that improve outcomes for students matters to practitioners and researchers. From my perspective, the past few years have practitioners at the point where three phenomena are occurring with regard to assessment practices in schools. First, most teachers are assessing more than they used to; second, some are assessing better; and third, too few are using assessments well to drive their instruction and improve outcomes. The following implications reflect this tension between the perspectives of practice and research and may offer a way forward.

Implications for Practitioners

First, I agree with Charles and Lynne, who claim that their teachers are growing from the PADI experience even though the projects diverged from the original PADI design to create original performance assessments. Like Charles explained, practitioners should understand that the implementation of new practices around assessment will often take longer than one cycle of instruction and assessment. Teachers need to plan with standards in mind but then also take time and to observe and reflect on those standards in action—in the behaviors and performances of their students during both instructional lessons and assessment activities. Only then will they be able to refine their understanding of those assessments and make better instructional decisions. This need for time should be a factor in the planning of assessment implementation and professional development.

To this end, practitioners should rethink the relationship between “local” and “outsourced” assessments. This study in its conceptualization, literature review, and design has cast “local” and “outsourced” as a binary where one is pitted against the other in two theories of action. Perhaps practitioners should see these types of assessments not as a binary choice but rather as a spectrum of options for practitioners to choose and to use to drive their assessment and instructional practice. With this in mind, I now question when “localness”—the creation and use of local assessments—is actually helpful and when it is

not, especially if the primary goal of professional development is to reform assessment practices. If the goal is to increase teacher practice in the process of using assessment information to drive instructional choices, then local performance projects on their own may be an inefficient way going forward. It would make more sense to have teachers work through multiple assessment cycles every year. In addition, it would help for teachers to do this work in areas where the standards and the assessments are already set out—for example, mathematics and language arts, which have clearly developed learning continuums and units of study. In this way, the work of looking at student assessment data could be curriculum-embedded (Palmer-Wolfe & Richardson, 1996). without being locally produced and could happen multiple times per year across the curriculum.

At the same time, I would not abandon “localness” altogether. Teachers are motivated by being designers and by developing outcomes that reflect the values of their community. When possible, practitioners should find projects and performances that allow this design process to occur. This could include, like both schools did at Suburban Woods, the re-tooling of existing projects and performances to reflect new assessment practices, such as the use of rubrics by students. It could also include modifying units from outsourced curriculum to reflect local interests and topics. If “localness” creates enthusiasm and engagement, then those attributes should be cultivated.

Therefore, I believe that while the work I observed fell short of this goal, both Charles and Lynne believe in the power of collective thinking to drive organizational change. While I was not able to prove it, I believe that Birdwell-Mitchell (2016) echoes the thinking of many scholars that collective phenomena such as “peer-learning” hold the key to work on assessment. As such, practitioners’ professional development should focus more efforts on work that encourages teams of teachers to examine their instruction and the impact of their instruction together. The facilitation of this work should also focus on these collective actions. That being said, when the goal of professional development is to produce better assessments and better assessment practice, then more support may be required across a

spectrum of assessments to ensure that “meaningful” outcomes are met. I close this section by offering advice to projects like PADI and to practitioners like myself.

My advice to PADI, and projects like this, is to push teachers and facilitators to make and use the assessments that they planned. This is not a criticism of Giselle Martin-Kniep’s role. Like me, she does not have influence over the teachers once the workshops are finished. But if the goal is to “close the assessment loop,” the need to push teachers would then rest with the facilitators to recognize where teachers need support and encouragement to see the work through. So, to coach facilitators into this role could be a consideration for future PADI cohorts. I know in my own work as a principal and as an instructional leader, it can be challenging to push teachers out of their comfort zone to work directly with assessments and assessment data, especially in groups where their work and their students’ work is being shared. But if this work is done regularly and in settings where trust is high, progress can be significant.

More broadly, those who facilitate assessment work with teachers must trust with the assumption that agency always begins with trust and support. Practitioners should recognize that teachers will be uncomfortable at first and that they may not always know what to do with assessment data. “Closing the assessment loop” requires that teachers work through sets of student work alongside standards and learning continuums to help them calibrate their understanding and set goals for their students. In time, these goals can also include students looking at their work with the same tools as they learn to evaluate their work and set goals for improvement. As we create structures and trust around this work the way Charles has done, teachers can learn how to make student learning more visible.

Implications for Research

Be they locally produced performance assessments or assessment activities from externally produced curriculum units of study, more research into how teachers use assessments in classrooms and in schools to drive instructional choices and improve

student outcomes is warranted. There is still a need to see into the “black box.” In some ways, the shortcomings and limitations of my own research design can point to directions for future research.

More research is needed specifically into the collective practices of teachers with the specific goal of understanding how those group practices lead to organizational change. Specific activities could include practices where teachers look at student work, participate in group deliberation and decision making around instructional practices, and in peer observations of practice.

More research is needed on the behaviors and practices of facilitators, coaches, and building leaders in order to understand what specific actions influence assessment practice. The actions studied could include the onboarding and support of new teachers, the creation of structures and schedules that facilitate individual and collective agency around assessment, as well as supporting the collaborative activities among teachers mentioned above.

If instructional changes happen slowly and over many years, more long-term study of assessment practice is needed. While my study demonstrated that it is very difficult for school administrators to lead “embedded” research into classroom practices, one thing that school leaders can do is help facilitate others to conduct that kind of research in their schools. School leaders such as myself should encourage research in their schools. Not only could this inform both researchers and practitioners; it could have the added impact described earlier as “transformative validity” (Cho & Trent, 2006), whereby teachers are engaged and energized around the work of studying and sharing their own assessment practices. This is the kind of transformation through collegial or collective learning that the second theory of action strives to achieve.

Implications for Policy as it Relates to School Reform

Following the advice of Vidovich (2009) and Weiss (2012), if the goal of policy is to “improve” teacher practices and ultimately outcomes for all students, then policy must continue to shift away from systems designed simply to “prove” results. True accountability will come when policies create structures that look for evidence of system-wide improvement of practices instead of structures of sanctions and rewards. For example, policies that encourage or even require assessment literacy could help drive systemic changes. Specifically, assessment literacy should include understanding how standards work across a learning continuum and how those standards look in classrooms. Policy in this area could focus on changes in requirements for pre-service teaching programs, recertification requirements, and criteria for leadership certification.

Funding of assessments should not stop at the creation and distribution of large-scale accountability testing. The cost of testing every child every year in this way is a drain on resources for all schools. Instead or in addition, policymakers must consider how to fund the development of curriculum-embedded assessment materials and resources, especially for higher-need schools. There now exist many examples of curriculum units that have rich assessments, rubrics, and learning continuums contained in them. But these materials and the training required to adopt them are expensive. Policy changes could put these resources in reach for all students.

Finally, in terms of “localness,” perhaps policymakers must consider the findings of research like that found in the McKinsey study (Mourshed et al., 2010) and by teams like Priestley et al. (2012), who call for systems that acknowledge local flexibility and the development of local solutions built around a common set of standards for all children.

REFERENCES

- Addie, K. L. (2005). *Proficiency-based standards reform: Implications for teacher professionalism and accountability* Unpublished doctoral dissertation: University of Oregon.
- Adie, L. (2012). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91-106. doi: 10.1080/0969594X.2011.650150
- Adler-Kassner, L., & Harrington, S. (2010). Responsibility and composition's future in the twenty-first century: Reframing "accountability." *College Composition and Communication*, 62(1), 73-99.
- Agarwal, N. K. (2012, October). Paper presented at ASISandT conference. Retrieved from: <http://gslis.simmons.edu/blogs/naresh/files/2013/01/Agarwal-ASIST-History-preconf-2012-author-formatted-6Jan2013.pdf>
- Agee, J. (2009). Developing qualitative research questions: A reflective process. *International Journal of Qualitative Studies in Education*, 22(4), 431-447.
- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments*. Thousand Oaks, CA: Corwin Press.
- Allal, L. (2012). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 20-34. doi: 10.1080/0969594X.2012.736364
- Anderson, L. (2010). Embedded, emboldened, and (net)working for change: Support-seeking and teacher agency in urban, high-needs schools. *Harvard Educational Review*, 80(4), 541-572, 586.
- Anyon, J. (1980). Social class and the hidden curriculum. *Journal of Education*, 62(1), 67-92.
- Aoki, T. (2004). *Curriculum in a new key: The collected works of Ted Aoki*. Mahwah, NJ: Erlbaum.
- Apple, M. W. (1997). Justifying the conservative restoration: Morals, genes, and educational policy. *Educational Policy*, 11(2), 167-182.
- Apple, M. W. (2004). Creating difference: Neo-liberalism, neo-conservatism and the politics of educational reform. *Educational Policy*, 18(1), 12-44.
- Apple, M. W. (2009). Producing difference: Neoliberalism, neoconservatism and the politics of educational reform. In M. Simons, M. Olssen, & M. A. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 625-649). New York, NY: Teachers College Press.
- Arter, J. A. (2003). *Assessment for learning: Classroom assessment to improve student achievement and well-being*(ED62480068). Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/62182226?accountid=62114258>.

- Aydeniz, M. (2007). *Understanding the challenges to the implementation of assessment reform in science classrooms: A case study of science teachers' conceptions and practices of assessment*. Unpublished doctoral dissertation: Florida State University.
- Baker, L., et al. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper #278). Washington, DC: Economic Policy Institute.
- Ball, S. J. (2009). Education reform, teacher professionalism and the end of authenticity. In M. Simons, M. Olssen, & M. A. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 667-682). New York, NY: Teachers College Press.
- Banks, J. A. (1993). The canon debate, knowledge construction, and multicultural education. *Educational Researcher*, 22(5), 4-14.
- Bao, C. (2009). *Within the classroom walls: Critical classroom processes, students' and teachers' sense of agency, and the making of racial advantages and disadvantages*. Unpublished doctoral dissertation: Boston College.
- Barron, B. J. S., Schwartz, D. L., Vye, N. J., Moore, A., Petrosino, A., Zech, L., Bransford, J. D., & the Cognition and Technology Group at Vanderbilt. (1998). Doing with understanding: Lessons from research on problem and project-based learning. *Journal of the Learning Sciences*, 7(3&4), 271-311. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/62487307?accountid=14258>
- Benson, K. (2011). *Teacher collaboration in context: Professional learning communities in an era of standardization and accountability*. Unpublished doctoral dissertation: Arizona State University.
- Berger, R. H. (2005). *Teacher capacity and assessment reform: Assumptions of policy, realities of practice*. Unpublished doctoral dissertation, Indiana University.
- Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18-20.
- Bernhardt, V. L. (2009). Data use: Data-driven decision making takes a big-picture view of the needs of teachers and students. *Journal of Staff Development*, 30(1), 24-27. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/61904598?accountid=61914258>.
- Biesta, G. (2009). Education between accountability and responsibility. In M. Simons, M. Olssen, & M. A. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 650-666). New York, NY: Teachers College Press.
- Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, and democracy*. Boulder, CO: Paradigm.
- Biesta, G., & Tedder, M. (2006). *How is agency possible? Towards an ecological understanding of agency-as-achievement* (Working paper 5). Exeter, England: The Learning Lives Project.

- Birdwell-Mitchell, E. N. (2015). Theorizing teacher agency and reform: how institutionalized instructional practices change and persist. *Sociology of Education*, 88(2), 140-159. Retrieved March 17, 2015, from soe.sagepub.com at Teachers College PARENT
- Blachford, P. (1997, September). Students' self assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison group. *Educational Psychology*, 17(3), 345. Retrieved November 13, 2008, from Education Research Complete database.
- Black, P. (2001). Dreams, strategies and systems: Portraits of assessment past, present and future. *Assessment in Education*, 8(1), 65-85. Retrieved December 5, 2008, from Research Library database. (Document ID: 105199708)
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9-21.
- Black, P., & Wiliam, D. (1998, March). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7. Retrieved November 13, 2008, from Education Research Complete database.
- Bloomberg, L. D., & Volpe, M. (2012). *Completing your qualitative dissertation: A road map from beginning to end*. Los Angeles, CA: Sage.
- Boberg, J. E. (2013). *High school principal transformational leadership behaviors and teacher extra effort during educational reform: The mediating role of teacher agency beliefs*. Unpublished doctoral dissertation, University of Texas at Arlington.
- Bogdan, R., & Biklen, S. (2007). *Qualitative research in education: An introduction to theory and practice* (5th ed.). Boston, MA: Allyn and Bacon.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 03-15. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/812186378?accountid=812114258>.
- Borjes, R. L. (2012). *Teacher technology use and state accountability scores: A correlational study*. Unpublished doctoral dissertation: University of the Incarnate Word.
- Bower, H. A. (2012). *"It's all about the kids": School culture, identity, and figured worlds*. Unpublished doctoral dissertation: University of North Carolina at Chapel Hill.
- Bredo, E. (2006). Philosophies of education research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 3-32). Washington, DC: American Educational Research Association.
- Brito, Q. D. (2009). *Improving classroom practices from the inside out: A case study of teacher agency in school reform in Belo Horizonte, Brazil*. Unpublished doctoral dissertation: University of California, Los Angeles.

- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education*, 11(1), 7-26. Retrieved December 5, 2008, from Research Library database. (Document ID: 811434251)
- Brookhart, S. M. (2012). The use of teacher judgment for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90. doi: 10.1080/0969594X.2012.703170
- Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Education Studies in Mathematics*, 57, 213-227.
- Brookhart, S. M., & Bronowicz, D. L. (2003). "I don't like writing. It makes my fingers hurt": Students talk about their classroom assessments. *Assessment in Education*, 10(2), 221-242. Retrieved December 5, 2008, from Research Library database. (Document ID: 629996461)
- Bruce, L. B. (2001). *Student self-assessment: Encouraging active engagement in learning* (Doctoral dissertation, Fielding Graduate Institute). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3012318)
- Butler, L. F. (2009). *Charter schools, data use, and the 21st century: How charter schools use data to inform instruction that prepares students for the 21st century*. Unpublished doctoral dissertation: University of Southern California.
- Buzzelli, C., & Johnston, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. New York, NY: RoutledgeFalmer.
- Campbell, C., & Collins, V. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 27(1), 9-18. Retrieved April 25, 2009 from Education Research Complete database.
- Campbell, E. (2012). Editorial: Teacher agency in curriculum contexts. *Curriculum Inquiry*, 42(2), 183-190. Retrieved from: <http://www.jstor.org/stable/23253806>.
- Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 12(1), 39-54. Retrieved April 18, 2009, from Education Research Complete database.
- Carlson, D. L. (2005). *Producing a sovereign self: Portfolios and the "ownership" of scholastic bodies*. Unpublished doctoral dissertation, Teachers College, Columbia University.
- Carroll, D., & Carini, P. (1991). Tapping teachers' knowledge. In V. Perrone (Ed.), *Expanding student assessment*. New York, NY: ASCD.
- Chapman, M. (2008). *Assessment literacy and efficacy: Making valid educational decisions* (Doctoral dissertation, University of Massachusetts). Retrieved April 18, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3325124)

- Cho, J., & Trent, A. (2006). Validity in qualitative research revisited. *Journal of Qualitative Research*, 6(3), 319-340.
- Choi, J. (2005). *Elementary social studies teachers' implementation of curriculum-embedded performance assessment in South Korea: A mixed method study*. Unpublished doctoral dissertation: Michigan State University.
- City, E. A., Elmore, R. A., Fiarman, S. E., & Teitel, L. (2009). *Instructional rounds in education: A network approach to improving teaching and learning*. Cambridge, MA: Harvard Education Press.
- Clark, D. C., & Clark, S. N. (2000). Appropriate assessment strategies for young adolescents in an era of standards-based reform. *Clearing House*, 73(4), 201-204.
- Clark, R. K. (2007). *"Top down": An analysis of state implementation of a federal teacher accountability policy*. Unpublished doctoral dissertation: University of Oregon.
- Clifford, L. S. (1995). *The impact of state-mandated student performance assessment on instructional practices*. Unpublished doctoral dissertation: University of Louisville.
- Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation. *American Educational Research Journal*, 43, 343. Downloaded from <http://aerj.aera.net> at COLUMBIA UNIV on August 6, 2013.
- Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, 112(4), 469-495.
- Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives*, 9(4), 173-206. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/964184960?accountid=964114258>.
- Coburn, C. E., & Turner, E. O. (2012a). Interventions to promote data use: An introduction. *American Journal of Education*, 118(2), 99-111.
- Coburn, C. E., & Turner, E. O. (2012b). The practice of data use: An introduction. *American Journal of Education*, 118(2), 99-111. doi: 10.1086/663272
- Cochran-Smith, M., & Lytle, S. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, 24, 249-305.
- Cofield, C. (2013). *Agentic typologies: Teacher agency and urban renewal as tools of school reform*. Unpublished doctoral dissertation, State University of New York at Buffalo.
- Colyvas, J. A. (2012). Performance metrics as formal structures and through the lens of social mechanisms: When do they work and how do they influence? *American Journal of Education*, 118(2), 167-197. doi: 10.1086/663270

- Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be?: The use of value-added measures of teacher effectiveness in policy and practice* (Education Policy for Action Series). Providence, RI: Annenberg Institute for School Reform at Brown University.
- Costanza, V. J. (2008). *Creating spaces for an ownership profession*. Unpublished doctoral dissertation, Rutgers The State University of New Jersey.
- Craw, K. G. (2009). *Performance assessment practices: A case study of science teachers in a suburban high school setting*. Unpublished doctoral dissertation, Teachers College, Columbia University.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Los Angeles, CA: Sage.
- Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Cruikshank, B. (1999). *The will to empower: Democratic citizens and other subjects*. Ithaca, NY: Cornell University Press.
- Daley Peterson, K. J. (2001). *Surprise and sense making: The organizational socialization of first-year teachers in intermediate grades four through six*. Unpublished doctoral dissertation, University of Nebraska at Omaha.
- Daniel, B.-J. M. (2003). *Cohort group membership and individual agency in teacher education: Implications for addressing issues of race, gender and class*. Unpublished doctoral dissertation, University of Toronto (Canada).
- Darling-Hammond, L. (1988). Policy and professionalism. In A. Lieberman (Ed.), *Building professional culture in schools* (pp. 55-77). New York, NY: Teachers College Press.
- Darling-Hammond, L. (2008). *Powerful learning: What we know about teaching for understanding*. San Francisco, CA: Wiley.
- Darling-Hammond, L., & Aneess, J. (1995) Authentic assessment and school development. In J. Baron & D. P. Wolf (Eds.), *Performance based student assessment: challenges and possibilities* (pp 52-83). Chicago, IL: University of Chicago Press.
- Darling-Hammond, L., Aneess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York, NY: Teachers College Press.
- Darling-Hammond, L., Wei, R., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX: NSDC.
- Datnow, A., Park, V., & Kennedy-Lewis, B. (2013). Affordances and constraints in the context of teacher collaboration for the purpose of data use. *Journal of Educational Administration*, 51(3), 341-362. Retrieved from: <http://ezproxy.cul.columbia.edu/login?url=http://search.proquest.com/docview/1355279398?accountid=1355210226>.

- DeHart, J. C. (2007). *The effects of student input into school organization, operation and governance on dropout rate in Iowa alternative schools and programs* (Doctoral dissertation, Iowa State University). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3259462)
- Dekker, T., & Feijs, E. (2005). Scaling up strategies for change: Change in formative assessment practices. *Assessment in Education: Principles, Policy & Practice*, 12(3), 237-254. Retrieved April 18, 2009, from Education Research Complete database.
- Deleuze, G. (1992, Winter). PostScript on the societies of control. *October*, 59, 3-7. Retrieved April 17, 2012, from Jstor database. <http://links.jstor.org/sici?sici+0162-2870%28199224%2959%3C3%3APOTSOC%3E2.0.CO%3B2-T>
- Dervin, B. (1998). Sense making theory and practice: An overview of user interests in knowledge seeking and use. *Journal of Knowledge Management*, 2(2), 36-46.
- Dervin, B., & Clark, K. D. (1999). Exemplars of the use of sense making methodology (meta-theory and method): In-depth introduction to the sense making issues of the electronic journal of communication. *Electronic Journal of Communication* [On-line serial], 9(2, 3, &4).
- Dervin, B., & Foreman-Wemet, L. (Eds.). (2003) *Sensemaking methodology reader: Selected writings of Brenda Dervin*. Cresskill, NJ: Hampton Press.
- Dorman, E. H. (2007). *The role of context, identity, and pedagogical tools in learning to teach for social justice and equity*. Unpublished doctoral dissertation, University of Colorado at Boulder.
- Dwyer, C. A. (1998). Assessment and classroom learning: Theory and practice. *Assessment in Education*, 5(1), 131-137. Retrieved December 6, 2008, from Research Library database. (Document ID: 30132782)
- Elwood, J. (2006). Formative assessment: possibilities, boundaries and limitations. *Assessment in Education: Principles, Policy & Practice*, 13(2), 215-232. doi: 10.1080/09695940600708653
- English, S. J. (2007). *Teacher education program accountability: In search of meaningful outcomes evidence*. Unpublished doctoral dissertation, Capella University, Ann Arbor.
- Erickson, F., & Gutierrez, K. (2002). Culture, rigor, and science in educational research. *Educational Researcher*, 31(8), 21-44.
- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4), 395-430. Retrieved November 13, 2008, from Education Research Complete database.
- Falkenberg, T. (2006). *Caring and human agency: Foundations of an approach to teacher education*. Unpublished doctoral dissertation, Simon Fraser University (Canada), A.

- Fenwick, L. (2012). Limiting opportunities to learn in upper-secondary schooling: differentiation and performance assessment in the context of standards-based curriculum reform. *Curriculum Inquiry*, 42(5), 629-651. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=83405132&site=ehost-live>.
- Feuer, M. J., Towne, L., & Shavelson, R. J. (2002). Scientific culture and educational research. *Educational Researcher*, 31(8), 4-14.
- Flowers, C., & Hancock, D. (2003). An interview protocol and scoring rubric for evaluating teacher performance. *Assessment in Education: Principles, Policy & Practice*, 10(2), 161-168. Retrieved April 18, 2009, from Education Research Complete database.
- Floyd, J. K. (2009). *Fort Zumwalt School District Professional Development Plan: A qualitative study of current district perceptions and proposed changes*. Unpublished doctoral dissertation, Lindenwood University.
- Forbes, E. (2007). *Improving the knowledge and use of formative assessment: A case of a model of formative assessment in a K-3 science curriculum* (Executive Position Paper, University of Delaware Amherst, Delaware). Retrieved April 18, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3267194)
- Foreman-Wimet, L. (2003). Rethinking communication: Introducing the sense making methodology. In B. Dervin & L. Foreman-Wemet, with E. Lauterback (Eds.), *Sense making methodology reader: Selected writings of Brenda Dervin* (pp. 3-16). Cresskill, NJ: Hampton Press.
- Foucault, M. (1979). *Discipline and punish* (A. Sheridan, Trans.). New York, NY: Harper.
- Frazier, C. H. (2007). *Investigating teachers' self-perceived applications of classroom assessment practices using the Assessment Practices Inventory Revised (APIR)*. Unpublished doctoral dissertation, University of Alabama.
- Frisbie, D. (2005). Measurement 101: Some fundamentals revisited. *Educational Measurement: Issues and Practice*, 24(3), 21-28. Retrieved April 25, 2009 from Education Research Complete database.
- Fuhrman, S. (2010). 2010-11: The year of research at TC. *Inside*, 16(3).
- Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York, NY: Teachers College Press.
- Gantt, P. E. C. (2012). *Teachers' perceptions of the impact of performance-based accountability on teacher efficacy*. Unpublished doctoral dissertation, Walden University.
- Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform, and equity*. Philadelphia, PA: Open University Press.

- Gilmore, A. (2002). Large scale assessment and teachers' assessment capacity: Learning opportunities for teachers in national education monitoring programs in New Zealand. *Assessment in Education: Principles, Policy & Practice*, 9(3), 343-361. Retrieved April 18, 2009, from Education Research Complete database.
- Glaser, R., & Silver, E. (1994, June). *Assessment, testing, and instruction: Retrospect and prospect* (CSE Technical Report 379). Pittsburgh, PA: CRESST/Learning Research and Development Center, University of Pittsburgh.
- Good, A. G. (2011). *A seat at the table: Teacher agency in educational policy design*. Unpublished doctoral dissertation, The University of Wisconsin-Madison.
- Good, R. (2006). *Analyzing the impact of data analysis process to improve instruction using a collaborative approach* (Doctoral dissertation, Texas A&M University). Retrieved April 18, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3245230)
- Goren, P. (2012). Data, data, and more data—What's an educator to do? *American Journal of Education*, 118(2), 233-237. doi: 10.1086/663273
- Grantz, D. G. (2005). *Analysis of patterns of data use by teachers at Seaford Middle School*. Unpublished doctoral dissertation, University of Delaware.
- Gratz, D. B. (2009). Purpose and performance in teacher performance pay. *Education Week*, 28(24), 40 & 32.
- Greenstein, L. (2004). *Finding balance in classroom assessment: High school teachers' knowledge and practice* (Doctoral dissertation, Johnson & Wales University). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3124561)
- Groh, A. M. (2013). *Structures and supports for data use in schools: A qualitative case study of one urban elementary school*. Unpublished doctoral dissertation, University of Wisconsin-Milwaukee.
- Guisse, M. E. (2009). *How three English Language Arts teachers negotiate their beliefs and instructional practices in three educational contexts*. Unpublished doctoral dissertation, University of Pittsburgh.
- Guskey T. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 27(1), 19-27. Retrieved April 25, 2009 from Education Research Complete database.
- Guskey T. (2009). *Practical solutions for serious problems in standards based grading*. Thousand Oaks, CA: Corwin Press.
- Hackett, J. L. (2003). *Reading teachers reading students: Exploring teachers' sense making about student resistance*. Unpublished doctoral dissertation, Mount Saint Vincent University (Canada).

- Haertel, E. H. (1999). Performance assessment and education reform. *Phi Delta Kappan*, 80(9), 662-666. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=503624112&site=ehost-live>.
- Hagedorn, E. A. (1999). *Development of a measure of student self-evaluation of physics exam performance* (Doctoral dissertation, University of Wisconsin). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 9926929)
- Haigh, M., & Dixon, H. (2007). 'Why am I doing these things?': Engaging in classroom-based inquiry around formative assessment. *Professional Development in Education*, 33(3), 359-376. Retrieved April 25, 2009, from Education Research Complete database.
- Haney, W. (1991). We must take care: Fitting assessments to functions. In V. Perrone (Ed.), *Expanding student assessment*. New York, NY: ASCD.
- Hargreaves, A., & Shirley, D. L. (2009). *The fourth way: The inspiring future for educational change*. Thousand Oaks, CA: Corwin Press.
- Harrison, C. (2005). Teachers developing assessment for learning: Mapping teacher change. *Teacher Development*, 9(2), 255-263. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/61897165?accountid=61814258>.
- Hart, C. (1999). *Doing a literature review: Releasing the social science research imagination*. London, England: Sage.
- Hayward, L., & Hedge, N. (2005). Traveling toward change in assessment: Policy practice and research in education. *Assessment in Education: Principles, Policy & Practice*, 12(1), 55-75. Retrieved April 18, 2009, from Education Research Complete database.
- Holloway, T. M. (2006). *The effect of principals' leadership style on student growth and teacher behavior in the accountability era*. Unpublished doctoral dissertation, University of Southern Mississippi.
- Holstrom, L. A. (1998). *Resistance and agency: A case study of a cohort of older post-baccalaureate women in a teacher education program*. Unpublished doctoral dissertation, University of Cincinnati.
- Honig, M. I., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational Policy*, 22(4), 578-608.
- Honig, M. I., & Venkateswaran, N. (2012). School-central office relationships in evidence use: Understanding evidence use as a systems problem. *American Journal of Education*, 118(2), 199-222. doi: 10.1086/663282
- Hoyt, K. (2005). *Teacher voice and participation in shaping large-scale standards-driven testing: The case of teacher involvement in the design and construction of a third year high school French end-of-course exam, based on the Indiana Academic Standards for Foreign Languages*. Unpublished doctoral dissertation, Indiana University.

- Huai, N., Braden, J., White, J., & Elliott, S. (2006). Effect of an Internet based professional development program on teachers' assessment literacy for all students. *Teacher Education and Special Education*, 29(4), 244-260. Retrieved April 25, 2009 from Education Research Complete database.
- Humphries, E. K. (2012). *Novice social studies teachers' sense making of their emerging identities as civics teachers*. Unpublished doctoral dissertation, University of Florida.
- Hyde, A. M. (2007). *Self-constitution as resistance to normalization: Educator agency in the era of accountability*. Unpublished doctoral dissertation, University of Pittsburgh.
- IES What Works Clearing House. (n.d.). Retrieved December 5, 2008, from <http://ies.ed.gov/ncee/wwc/references/standards/>
- Jahns, A. J. (2009). *I can, we can, and you can: An examination of how teacher agency develops learner agency*. Unpublished doctoral dissertation, Marian University.
- James, J. K. (2007). *The influence of No Child Left Behind and the Wyoming Comprehensive Assessment System on curriculum and instruction: Perceptions of Wyoming's elementary teachers*. Unpublished doctoral dissertation, University of Wyoming.
- Jennings, J. (2012, November). The effects of accountability system design on teachers' use of test score data. *Teachers College Record*, 114, 110304.
- Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial & Special Education*, 21(5), 261-267 Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=507714878&site=ehost-live>.
- Johnston, M. T. (2009). *Supports that facilitate teacher data use in schools*. Unpublished doctoral dissertation, University of Texas at Austin.
- Jones, K., & Whitford, B. L. (1997). Kentucky's conflicting reform principles: High-stakes school accountability and student performance assessment. *Kentucky Instructional Results Information System (KIRIS)*, 79, 276-281. Retrieved from <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.spx?direct=true&db=eft&AN=503443669&site=ehost-live>.
- Kane, M. B., & Khattri, N. (1995). Assessment reform: a work in progress. *Phi Delta Kappan*, 77, 30-32. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=503341983&site=ehost-live>.
- Keazer, L. M. (2012). *Mathematics teachers investigating reasoning and sense making in their teaching*. Unpublished doctoral dissertation, Purdue University.
- Kennedy, M. M. (2004). Reform ideals and teachers practical intentions. *Education Policy Analysis Archives*, 12(3). Retrieved March 1, 2013 from <http://epaa.edu/eppa/v12n13/>.
- Kennedy, M. M. (2005). *Inside teaching: How classroom life undermines reform*. Cambridge, MA: Harvard University Press.

- Kesler, T. B. (2008). *Complexities of negotiating balanced literacy and high-stakes testing: Exploring the discourses and enactments of two urban teachers and their students labeled "at-risk."* Unpublished doctoral dissertation, Teachers College, Columbia University.
- Klenowski, V. (1995, August). Student self-evaluation processes in student-centered teaching and learning contexts. *Assessment in Education: Principles, Policy & Practice*, 2(2), 145. Retrieved November 13, 2008, from Education Research Complete database.
- Klenowski, V. (2013). Investigating the complexity of judgement practice. *Assessment in Education: Principles, Policy & Practice*, 20(1), 1-4. doi: 10.1080/0969594X.2012.745702
- Kliebard, H. M. (1993). *The struggle for the American curriculum 1893-1958* (3rd ed.). New York, NY: Routledge.
- Kliebard, H. M. (2002). *Changing course: American curriculum reform in the 20th century*. New York, NY: Teachers College Press.
- Knight, M. G., Norton, N. E. L., Bently, C. C., & Dixon, I. R. (2004). The power of Black and Latina/o counterstories: Urban families and college-going. *Anthropology and Education*, 31(1), 99-120.
- Koo, W. S. (2010). *The sense making process of teachers in institutional change in curriculum: A case study on the implementation of the subject liberal studies in Hong Kong*. Unpublished doctoral dissertation, Chinese University of Hong Kong (Hong Kong).
- Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Kort, T. L. (2008). *Teachers making sense of data within a response to intervention model: A case study*. Unpublished doctoral dissertation, University of Florida.
- Ladewski, B. G. (2006). *Making sense of shared sense making in an inquiry-based science classroom: Toward a sociocultural theory of mind*. Unpublished doctoral dissertation, University of Michigan.
- Lai, E., & Waltman, K. (2008). Test preparation: examining teacher perceptions and practices. *Educational Measurement: Issues and Practice*, 27(2), 28-45. Retrieved April 25, 2009 from Education Research Complete database.
- Landis, B. C. (1998). *Looking at the effective schools characteristics and performance assessment in Pennsylvania elementary schools*. Unpublished doctoral dissertation, Lehigh University.
- Lee, C., & Wiliam, D. (2005). Studying changes in the practice of two teachers developing assessment for learning. *Teacher Development*, 9(2), 265-283. doi: 10.1080/13664530500200244

- LeMahieu, P., & Eresh, J. (1995). Coherence, comprehensiveness, and capacity in assessment systems: the Pittsburgh experience. In J. Baron & D. P. Wolf (Eds.), *Performance based student assessment: Challenges and possibilities* (pp. 125-142). Chicago, IL: University of Chicago Press.
- Lewis, L. M. (2010). *Teachers' perceptions of the effectiveness of benchmark assessment data to predict student math grades*. Unpublished doctoral dissertation, Northcentral University.
- Lindsey, M. A. (2000). *A constructivist study of developing curriculum to teach Internet information literacy to Navajo high school students* (M.A. thesis, Prescott College). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 1401532)
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16. Retrieved from: <http://www.jstor.org/stable/1177052>.
- Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, 118(2), 143-166. doi: 10.1086/663271
- Liu, S. (2006). *A description and analysis of the congruence between teachers' classroom assessments and the recent curricular reform initiatives* (M.A. thesis, University of Toronto, Canada). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT MR16231)
- Llosa, L. M. (2005). *Building and supporting a validity argument for a standards-based classroom assessment of English proficiency*. Unpublished doctoral dissertation, University of California, Los Angeles.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- Luke, A. (1996). Text and discourse in education: an introduction to critical Discourse analysis. *Review of Research in Education*, 21(1995-6), 3-48.
- Luke, A., Green, J., & Kelly, G. J. (2010). What counts as evidence and equity? *Review of Research in Education*, 34(1), vii-xvi. doi: 10.3102/0091732x09359038
- Luttrell, W. (2000). "Good enough" methods for ethnographic research. In W. Luttrell (Ed.), *Qualitative educational research: Reading in reflexive methodology and transformative practice* (pp. 258-278). New York, NY: Routledge.
- Luttrell, W. (Ed.). (2009). *Qualitative educational research: Reading in reflexive methodology and transformative practice*. New York, NY: Routledge.
- Mackay, L. P. (2002). *A comparison of instructor and student self-evaluations of student leadership potential* (Ph.D. dissertation, Texas A&M University). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3072497)

- Maltbia, T. (2013). *Qualitative research methods: Data analysis & reporting*. PowerPoint presentation, Teachers College, Columbia University.
- Marsh, J. (2012, November). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, 114, 110303.
- Marshall, C., & Rossman, G. B. (2010). *Designing qualitative research* (5th ed.). Thousand Oaks, CA: Sage.
- Marshall, D., Sears, J., Allen, L. A., Roberts, P. A., & Schubert, W. H. (2006). *Turning points in curriculum: A contemporary American memoir* (2nd ed.). New York, NY: Prentice-Hall.
- Martin-Kniep, G., & Picone-Cochia, J (2009). *Change the way you teach, improve the way students learn*. Alexandria, VA: ASCD.
- Martin-Kniep, G., & Wilson, D. (2015, January). Authentic assessment and learning for students and teachers: A path to engagement. *Curriculum, Instruction, and Assessment Insights*, pp. 1-8. Downloaded on February 5, 2015 from <http://www.lcilt.org>.
- Mason, G. (2013). *Education reform and teacher accountability in the Obama era*. Unpublished doctoral dissertation, George Washington University.
- Mason, J. H. (2010). *Exploring the influence of high-stakes testing and accountability on teachers' professional identities through the factors of instructional practice, work environment, and teacher efficacy*. Unpublished doctoral dissertation, Western Carolina University.
- Mazzie, D. (2008). *The effects of professional development related to classroom assessment on student achievement in science* (Doctoral dissertation, University of South Carolina). Retrieved April 18, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3321420)
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10(2), 209-220. Retrieved December 5, 2008, from Research Library database. (Document ID: 629996451)
- Meier, S., Rich, B., & Cady, J. (2006). Teachers' use of rubrics to score non-traditional tasks: factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy & Practice*, 13(1), 69-95. Retrieved April 18, 2009, from Education Research Complete database.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Mertler, C. (2005). Secondary teachers' assessment literacy: does classroom experience make a difference? *American Secondary Education* 33(2), 76-92.

- Mertler, C., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the "Assessment Literacy Inventory"* (ERIC Report ED490355). Online submission of paper presented at AERA. Downloaded from ERIC, Resources in Education April 18, 2009.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Moore, R. (2005). Advising students in developmental education: How accurate are developmental education students' self-assessments? *RTDE*, 22(1), 53-58.
- Moss, P. A. (2012). Exploring the macro-micro dynamic in data use practice. *American Journal of Education*, 118(2), 223-232. doi: 10.1086/663274
- Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in educational assessment. *Review of Research in Education*, 30(1), 109-162. doi: 10.3102/0091732x030001109
- Mourished, M., Chijioke, C., & Barber, M. (2010). *How the world's most improved school systems keep getting better*. Boston, MA: McKinsey. Retrieved November 29, 2010 from http://www.mckinsey.com/client-service/Social_Sector/our_practices/Education/Knowledge_Highlights/Best_performing_school.aspx
- Munns, G., & Woodward, H. (2006). Student engagement and student self-assessment: The REAL framework. *Assessment in Education*, 13(2), 193-213.
- Munroe, P. (1998). Engendering curriculum history. In W. Pinar (Ed.), *Curriculum toward new identities* (pp. 263-294). New York, NY: Garland.
- Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, 22, 155-175.
- Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *Part of a special issue: First Cycle of PISA (2000-2006)—International Perspectives on Successes and Challenges: Research and Policy Directions*, 8(3), 545-563. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=508168928&site=ehost-live>.
- Nevo, D. (1983). The conceptualization of educational evaluation: An analytical review of the literature. *Review of Educational Research*, 53(1), 117-128. doi: 10.3102/00346543053001117
- New York State Legislation. (2012). *A11171-2009 S 3012-C annual professional performance review of classroom teachers and principals*. Retrieved December 3, 2010 from <http://open.nysenate.gov/legislation/bill/A11171>.
- Noll, R. E. (2007). *Teacher job satisfaction in Kentucky: The impact of accountability, principals, and students*. Unpublished doctoral dissertation, University of Cincinnati.

- O'Day, J. A. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72(3), 1-37. Retrieved November 25, 2014 from <http://gseweb.harvard.edu/-hepg/oday.html>.
- O'Leary, M. (2008). Towards an agenda for professional development in assessment. *Journal of In-Service Education*, 34(1), 109-114. Retrieved April 18, 2009, from Education Research Complete database.
- Olina, Z. (2002). *Teacher evaluation, student self-evaluation and learner performance* (Doctoral dissertation, Arizona State University). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3043826)
- Olina, Z., & Sullivan, H. J. (2004). Student self-evaluation, teacher evaluation, and learner performance. *Educational Technology Research and Development*, 52(3), 5-22.
- Oliveira, A. W. (2012). Teacher agency in the performance of inquiry-oriented science curriculum reform. *Cultural Studies of Science Education*, 7(3), 569-577. Retrieved from: <http://ezproxy.cul.columbia.edu/login?url=http://search.proquest.com/docview/1038760590?accountid=1038710226>.
- Palmer-Wolf, D., & Reardon, S. F. (1996). Access to excellence through new forms of student assessment. In J. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities* (pp 1-31). Chicago, IL: University of Chicago Press.
- Park, V. (2008). *Beyond the numbers chase: How urban high school teachers make sense of data use*. Unpublished doctoral dissertation, University of Southern California.
- Parr, J. M., & Timperley, H. S. (2008). Teachers, schools and using evidence: Considerations of preparedness. *Assessment in Education: Principles, Policy & Practice*, 15(1), 57-71. Received from: <http://eduproxy.tc-library.org/?url=/docview/61956425?accountid=61914258>.
- Peressini, D. D., & Knuth, E. J. (1998). The importance of algorithms in performance-based assessments. *Yearbook (National Council of Teachers of Mathematics)*, 1998, 56-68. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=507609744&site=ehost-live>.
- Perkins, M. F. (2001). *The Maryland School Performance Assessment Program testing and mathematical instruction: A comparative study*. Unpublished doctoral dissertation, Walden University.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning process: Toward a wider conceptual field. *Assessment in Education: Principles, Policy & Practice*, 5(1), 85-102. Retrieved April 18, 2009, from Education Research Complete database.
- Peshkin, A. (1988) In search of subjectivity - one's own. *Educational Researcher*, 17(7), 17-22.

- Peterson, S. S. (2012). An analysis of discourses of writing and writing instruction in curricula across Canada. *Curriculum Inquiry*, 42(2), 260-284. Retrieved from: <http://www.jstor.org/stable/23253810>.
- Pham, C. C. (2011). *An investigation of how elementary school teachers make data-driven instructional decisions in literacy*. Unpublished doctoral dissertation, University of San Diego and San Diego State University.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48, 4-11. Retrieved April 18, 2009 from Education Research Complete database.
- Popkewitz, T. S. (2008). *Cosmopolitanism and the age of school reform: Science, education, and making society by making the child*. New York, NY: Routledge.
- Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Postman, N. (1985). *Amusing ourselves to death: Public discourse in the age of show business*. New York, NY: Penguin.
- Priestley, M., Edwards, R., Priestley, A., & Miller, K. (2012). Teacher agency in curriculum making: Agents of change and spaces for manoeuvre. *Curriculum Inquiry*, 42(2), 191-214. Retrieved from: <http://www.jstor.org/stable/23253807>.
- Quinn, J. R. (2009). *Teacher sense making and policy implementation: A qualitative case study of a school system's reading initiative in science*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Rawson, K. A., & Dunlosky, J. (2007). Improving students' self-evaluation of learning for key concepts in textbook materials. *European Journal of Cognitive Psychology*, 19(4&5), 599-579.
- Resnick, L. (2006). Making accountability really count. *Educational Measurement: Issues and Practice*, 25(1), 33-37. Retrieved April 25, 2009 from Education Research Complete database.
- Rhude-Faust, M. K. (2011). *Piloting without a flight plan: Data use in professional learning communities*. Unpublished doctoral dissertation, University of Pennsylvania.
- Riveros Barrera, A. (2013). *Beyond policy implementation: Policy sense making and policy enactment in schools*. Unpublished doctoral dissertation, University of Alberta (Canada).
- Rollins, D. J. H. (2003). *Impact of North Carolina accountability program: Teacher beliefs and perceptions*. Unpublished doctoral dissertation, East Carolina University.
- Root, D. A. (2013). *Purpose, policy, and possibilities: Social studies teachers' sense making of curriculum*. Unpublished doctoral dissertation, University of Texas at San Antonio.

- Rose, K. P. (2004). *Using assessment conversations to promote teacher and student learning* (Doctoral dissertation, Northwestern University). Retrieved November 20, 2008, from Dissertations & Theses: Full Text database. (Publication No. AAT 3132591)
- Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self evaluation in grade 5-6 mathematics effects on problem solving achievement. *Educational Assessment*, 8(1), 43-59.
- Rothstein, R. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.
- Ruggles, K. C. (2009). *Teacher agency in the process of state mandated reform*. Unpublished doctoral dissertation, University of California, San Diego and California State University, San Marcos.
- Salyers, B. L. (2011). *Experiencing the crux of agency and structure in urban education: On becoming a teacher educator*. Unpublished doctoral dissertation, University of North Carolina at Charlotte.
- Sanger, M. (2012). The schizophrenia of contemporary education and the moral work of teaching. *Curriculum Inquiry*, 42(2), 285-307. Retrieved from: <http://www.jstor.org/stable/23253811>.
- Sarour, E. O. (2007). *The moral agency of family and consumer sciences teacher candidates: A grounded theory*. Unpublished doctoral dissertation, Iowa State University.
- Schafer, W. (1993). Assessment literacy for teachers. *Theory Into Practice*, 32(2), 118-126. Retrieved April 18, 2009 from Education Research Complete database.
- Scheurich, J. J., Skrla, L., & Johnson, J. F. (2000). Thinking carefully about equity and accountability. *Phi Delta Kappan*. 82, 293-299. doi: 10.1177/003172170008200411.
- Schissel, J. L. (2012). *The pedagogical practice of test accommodations with emergent bilinguals: Policy-enforced washback in two urban schools*. Unpublished doctoral dissertation, University of Pennsylvania.
- Schmidt, M. J. (2001). *Toward teacher inquiry when implementing classroom assessment*. Unpublished doctoral dissertation, University of Toronto (Canada).
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books.
- Seashore, K. L., Febey, K., & Schroeder, R. (2005). State-mandated accountability in high schools: Teachers' interpretations of a new era. *Educational Evaluation and Policy Analysis*, 27(2), 177-204. Retrieved from: <http://ezproxy.cul.columbia.edu/login?url=http://search.proquest.com/docview/197237728?accountid=197210226>.
- Sebatane, E. (1998, March). Assessment and classroom learning: A response to Black & William. *Assessment in Education: Principles, Policy & Practice*, 5(1), 123. Retrieved November 13, 2008, from Education Research Complete database.

- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499. doi: 10.3102/0034654307310317
- Sharkey, N. S., & Murnane, R. J. (2006). Tough choices in designing a formative assessment system. *American Journal of Education*, 112(4), 572-588. doi: 10.1086/505060
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher*, 21(4), 22-27. Retrieved from: <http://www.jstor.org/stable/1177207>.
- Shavelson, R. J., & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of Educational Research*, 51(4), 455-498. doi: 10.3102/00346543051004455
- Shaw, J. M., & Nagashima, S. O. (2009). The achievement of student subgroups on science performance assessments in inquiry-based classrooms. *Electronic Journal of Science Education*, 13(2), 6-29. Retrieved from: <http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=508035160&site=ehost-live>.
- Shulman, L. S. (1999). Professing educational scholarship. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp. 159-165). San Francisco, CA: Jossey-Bass.
- Silva, R. D. C. (2001). *Teacher-as-researcher: Contested issues of voice and agency in the creation of legitimised knowledge in teaching*. Unpublished doctoral dissertation, University of Houston.
- Sloan, K. (2006). Teacher identity and agency in school worlds: Beyond the all-good/all-bad discourse on accountability-explicit curriculum policies. *Curriculum Inquiry*, 36(2). Pp. 119-152. Retrieved March 23, 2014 from JSTOR 72.32.119.162.
- Smith, M., Appelman, D., & Wiliem, J. (2015) *Uncommon core: where the authors of the standards go wrong about instruction- and how you can get it right*. Thousand Oaks, CA: Corwin Press.
- Sonnek, B. K. (2003). *How teachers and administrators perceive a language arts change initiative in one middle school*. Unpublished doctoral dissertation, University of Iowa.
- Sonu, D. (2012). Illusions of compliance: Performing the public and hidden transcripts of social justice education in neoliberal times. *Curriculum Inquiry*, Retrieved from: <http://www.jstor.org/stable/23253809>.
- Spillain, J. P. (2012). Data in practice: conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113-141. doi: 10.1086/663283
- St. Pierre, E. (2002). Comment: "Science rejects postmodernism." *Educational Researcher*, 31(8), 25-27.

- Statewide performance assessment. (1997). *Symposium*, 69, 6-40. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=507598138&site=ehost-live>.
- Steiner, D. (2010). How to sort good apples from bad: State education chief lays out plan to judge teachers using tests. *Daily News*. Retrieved November 2, 2010 from <http://nydailynews.com/fdcp?1288731879934>
- Stewart Rose, L. J. (2008). *The lived experience of learning to teach in inner city schools: Negotiating identity and agency in the figured worlds of teacher candidates*. Unpublished dissertation, University of Toronto (Canada).
- Stiggins, R. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3). Retrieved April 18, 2009 from Education Research Complete database.
- Stiggins, R. (1998). *Classroom assessment for student success* (ERIC Report ED429990). Downloaded April 18, 2009 from ERIC, Resources in Education.
- Strong, S., & Sexton, L. C. (2000). A validity study of the Kentucky's performance based assessment system with National Merit Scholars and National Merit Commended. *Journal of Instructional Psychology*, 27(3), 202-206. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=507715886&site=ehost-live>.
- Supovitz, J. (2012, November). Getting student understanding- the key to teachers' use of test data. *Teachers College Record*, 114, 120301.
- Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: Teachers' perceptions and practices in mathematics*. Unpublished dissertation, Virginia Polytechnic Institute and State University.
- Tan, K. H. K. (2008). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research & Development*, 27(1), 15-29.
- Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, 28(5), 549-565.
- Taubman, P.M. (2011). *Teaching by numbers: Deconstructing the discourse of standards and accountability in education*. New York, NY: Routledge.
- Terwilliger, J. (1997). Semantics, psychometrics, and assessment reform: A close look at "authentic" assessments. *Educational Researcher*, 26(8), 24-27. Retrieved from: <http://www.jstor.org/stable/1176303>.
- Tierney, R. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education*, 13(9), 239-264. Retrieved April 18, 2009 from Education Research Complete database.
- Tierney, R. D. (2010). *Insights into fairness in classroom assessment: Experienced English teachers share their practical wisdom*. Unpublished dissertation, University of Ottawa (Canada).

- Tomlinson, C. A. (1991). *Teacher and student sense making about differentiation of instruction for gifted learners*. Unpublished dissertation, University of Virginia.
- Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal*, 27(5), 615-631. doi: 10.2307/1501956
- Tucker, M. S. (2012). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Cambridge, MA: Harvard Education Press.
- Tyack, D. B. (1974). Inside the system: The character of urban schools, 1890-1940. In *The one best system: A history of American urban education* (pp. 177-228). Cambridge, MA: Harvard University Press.
- United States Department of Education. (n.d.). *No Child Left Behind Act of 2001*. Retrieved December 5, 2008 from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- United States National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and the Secretary of Education*. Washington, DC: United States Dept. of Education.
- Van Maele, D. (2007). *Data use by teachers in high-performing, suburban middle schools to improve reading achievement of low-performing students*. Unpublished dissertation, University of Pittsburgh.
- Vidovich, L. (2009). "You don't fatten the pig by weighing it": Contradictory tensions in the policy pandemic of accountability infecting education. In M. Simons, M. Olssen, & M. A. Peters (Eds.), *Re-reading education policies: A handbook studying the policy agenda of the 21st century* (pp. 549-567). New York, NY: Teachers College Press.
- Villanueva, A. (2013). *Implementing a district-wide professional development initiative: What it means to educate for the 21st century*. Unpublished dissertation, Columbia University.
- Vogel, L., Rau, W., Baker, P. & Ashby, D. (2006). Bringing assessment literacy to the local school: A decade of reform initiatives in Illinois. *Journal of Education for Students Placed at Risk*, 11(1), 39-55. Retrieved April 18, 2009 from Education Research Complete database.
- Vogler, K. E. (2000). *The impact of high-stakes, state-mandated student performance assessment on 10th grade English, mathematics, and science teachers' instructional practices*. Unpublished dissertation, University of Massachusetts-Lowell.
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, 34(2), 239-255. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=525894401&site=ehost-live>.

- Volante, L., & Fazio, X. (2010). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749-770. Retrieved April 18, 2009 from Education Research Complete database.
- Volante, L., & Jaafar, S. B. (2010). Assessment reform and the case for learning-focused accountability. *Journal of Educational Thought*, 44(2), 167-188. Retrieved from: <http://eduproxy.tc-library.org/?url=http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=508188688&site=ehost-live>.
- Wallenstein, J., & Hatch, T. (2014). *The instructional value of performance assessments within an accountability context*. Paper presented at the annual conference of the American Educational Research Association, Philadelphia, PA.
- Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, 78(3), 516-551. doi: 10.3102/0034654308320292
- Warren-Little, J. (2012). Data use practice among teachers: the contribution of micro-process studies. *Journal of Education*, 118(2), 143-166.
- Wassell, B. A. (2004). *On becoming an urban teacher: Exploring agency through the journey from student to first-year practitioner*. Unpublished dissertation, University of Pennsylvania.
- Watson, A. (2006). Some difficulties in informal assessment in mathematics. *Assessment in Education: Principles, Policy & Practice*, 13(3), 289-303. Retrieved April 18, 2009, from Education Research Complete database.
- Wayman, J. C., & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education*, 112(4), 463-468. doi: 10.1086/505055
- Weick, J. (1995). *Sensemaking in organizations*. Thousand Oaks, CA: Sage.
- Weick, J., Sutcliff, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409-421. Retrieved from <http://www.jstor.org/25145979>.
- Weiss, J. A. (2012, November). Data for improvement, data for accountability. *Teachers College Record*, 114, 110307.
- Whitty, G. (2010, November). *Policy borrowing. Transcript of speech given at Teachers College*. Retrieved December 5, 2010 from <http://blip.tv/file/4369718>
- Wiliam, D. (2010a). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
- Wiliam, D. (2010b). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, 34(1), 254-284. doi: 10.3102/0091732x09351544

- Wiliam, D., Lee, C, Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49-65. Retrieved December 5, 2008, from Research Library database. (Document ID: 811434271)
- Williams, K. L. (2002). *Accountability and professional prerogative: The impact of high stakes testing on teacher agency*. Unpublished dissertation, California State University, Fresno and University of California, Davis.
- Wilson, M. E. (2009). *Using data to inform curriculum, instruction and professional development in science education* Unpublished doctoral dissertation, Teachers College, Columbia University.
- Wilson, N. (2008). Teachers expanding pedagogical content knowledge: Learning about formative assessment together. *Professional Development in Education*, 34(3), 283-298. Retrieved April 25, 2009, from Education Research Complete database.
- Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34(1), 1-24. doi: 10.3102/0091732x09350472
- Wolf, D, Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74. Retrieved February 14, 2014 from www.jstor.org.
- Wozny, P. D. (1998). *A view into performance assessment in science*. Unpublished dissertation, University of Alberta (Canada).
- Yamashita, M. Y. (2011). *How does high stakes testing influence teachers' classroom instruction?: Institutional pressures and classroom instruction*. Unpublished dissertation, University of Pittsburgh.
- Young, V. M. (2005). *Teachers' use of data: Embedded contexts and institutional logics*. Unpublished dissertation, Stanford University.
- Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education*, 112(4), 521-548. doi: 10.1086/505058
- Young, V. M., & Kim, D. H. (2010). Using assessment for instructional improvement: A literature review. *Education Policy Analysis Archives*, 18(19). Retrieved March 24, 2012, from <http://epaa.au.edu/ojs/article/view/809>.
- Zanazanian, P. (2012). Historical consciousness and the structuring of group boundaries: A look at two francophone school history teachers regarding Quebec's Anglophone minority. *Curriculum Inquiry*, 42(2), 215-239. Retrieved from: <http://www.jstor.org/stable/23253808>.
- Zhang, Z. (1995). *Investigating teachers' perceived assessment practices and assessment competencies on the Assessment Practices Inventory (API)*. Unpublished dissertation, University of Alabama.

- Zwick, R., & Sklar, J. (2008). Instructional tools in education measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, 27(2), 14-27. Retrieved April 25, 2009 from Education Research Complete database.
- Zwick, R., Sklar, J. C., Wakefield, G., Hamilton, C., Norman, A., & Folsom, D. (2008). Instructional tools in educational measurement and statistics (ITEMS) for school personnel: Evaluation of three web-based training modules. *Educational Measurement: Issues and Practice*, 27(2), 14-27. Retrieved from: <http://eduproxy.tc-library.org/?url=/docview/61981100?accountid=61914258>.

Appendix A

Informed Consent

Teachers College, Columbia University
525 West 120th Street
New York NY 10027
212 678 3000
www.tc.edu

DESCRIPTION OF THE RESEARCH: You are invited to participate in a pilot research study on teacher agency in the creation and use of local assessments. You will be asked to participate in two 30 minute interviews and one 45 minute focus group that will be audio recorded. The researcher will all make non-participant observations of your PADI team during several training sessions throughout the course of the year. The transcripts of these interviews and the notes taken during these observations will be kept confidential and will only be reviewed by the research and his advisor for the purpose of refining his dissertation. The research will be conducted solely by the primary researcher, Duncan Wilson. The research will be conducted at various locations including Wainwright House in Rye, New York, Mercy College, and the North Shore School District.

RISKS AND BENEFITS: Participants will experience minimal risk similar to the risk of boredom, fatigue, or professional embarrassment associated with any professional development training that involves working with teachers from multiple districts. These risks will be minimized by the confidentiality measures mentioned above. In addition, participants will have the right to review and comment on interview transcripts for accuracy prior to their being used in the research. There is no direct benefit to this research for the subjects. If at any time, should a participant become distressed or simply choose to back out of the process, the interviews can cease, the data collected can be returned, and appropriate assistance will be called upon based on the participant's requests.

DATA STORAGE TO PROTECT CONFIDENTIALITY: All audio recordings and transcripts will be kept confidentially in the researchers office. They will be destroyed following the defense of the dissertation.

TIME INVOLVEMENT: Your participation will take approximately 60 to 120 minutes.

HOW WILL RESULTS BE USED: The results will only be reviewed by the researcher and his advisor for the sole purpose of a dissertation research study.

Teachers College, Columbia University
 525 West 120th Street
 New York NY 10027
 212 678 3000
www.tc.edu

Principal Investigator: Duncan Wilson

Research Title: Unpacking Assessment: Understanding Teacher Agency in the Development and Use of Local Assessments

- I have read and discussed the Research Description with the researcher. I have had the opportunity to ask questions about the purposes and procedures regarding this study.
- My participation in research is voluntary. I may refuse to participate or withdraw from participation at any time without jeopardy to future medical care, employment, student status or other entitlements.
- The researcher may withdraw me from the research at his/her professional discretion.
- If, during the course of the study, significant new information that has been developed becomes available which may relate to my willingness to continue to participate, the investigator will provide this information to me.
- Any information derived from the research project that personally identifies me will not be voluntarily released or disclosed without my separate consent, except as specifically required by law.
- If at any time I have any questions regarding the research or my participation, I can contact the investigator, who will answer my questions. The investigator's phone number is (914) 490-9260.
- If at any time I have comments, or concerns regarding the conduct of the research or questions about my rights as a research subject, I should contact the Teachers College, Columbia University Institutional Review Board /IRB. The phone number for the IRB is (212) 678-4105. Or, I can write to the IRB at Teachers College, Columbia University, 525 W. 120th Street, New York, NY, 10027, Box 151.
- I should receive a copy of the Research Description and this Participant's Rights document.
- If video and/or audio taping is part of this research, I () consent to be audio/video taped. I () do NOT consent to being video/audio taped. The written, video and/or audio taped materials will be viewed only by the principal investigator and members of the research team.
- Written, video and/or audio taped materials () may be viewed in an educational setting outside the research
 (X) may NOT be viewed in an educational setting outside the research.
- My signature means that I agree to participate in this study.

Participant's signature: _____ Date: ____/____/____

Name: _____

Appendix B

Accountability Assessment: Sorting the Bad Apples

As mentioned in Chapter I, O'Day (2002) has framed accountability assessment using the term "bureaucratic accountability" and has posited that the primary difference of the "new accountability" that has continued to evolve in the decade since she conducted her study is that schools and teachers are held accountable for the outcomes and not for delivering inputs and processes. Embedded in this approach to assessment-driven reform is the theory of action that change in practice and outcomes occurs "through extrinsic rewards and sanctions for both schools and students on the assumptions that the fundamental problem is a lack of will to change on the part of educators" (Ancess & Darling-Hammond, 1996, p. 57). In her 2002 comparative study of Chicago schools and Baltimore schools, O'Day puts forward Chicago school reform as an example of bureaucratic accountability in a four-part frame. She focuses on how the reform policy looks at attention, motivation, knowledge development, and resource allocation. In so doing, O'Day demonstrates the flaws of bureaucratic accountability—how a discourse of accountability shifts the focus of reform away from any discussion of teacher agency or what teachers learn and do in order to raise standards to improve performance toward simply measuring results and punishing failure. These same four attributes described in Chicago in 2002 can be seen in a newspaper editorial written by New York State Education Commissioner David Steiner, who outlined the emerging statute 3012-C. in 2010. The title, "How to Sort Good Apples from Bad," implies that the problem of failing schools is ineffective teachers and that the solution is simply to "sort" teachers into four categories from which "school districts can better target incentives to reward their most effective teachers and principals and provide targeted professional development" (p. 1). By looking at the Chicago study from 2002 in parallel with the New York statute from 2012, the similarities reveal that little has changed in almost two decades of bureaucratic accountability policies. What Steiner's editorial ignores

and O'Day's research, along with subsequent studies, reveals is the unintended consequences of bureaucratic accountability. In particular, they ignore the impact these policies have on teachers and on outcomes.

In essence, the discourse of accountability in Steiner's "Bad Apples" analogy labels teachers as a bad investment. Portraying teachers as unmotivated or as incapable of self-improvement, that is, as bad apples, is also described well as "maladaptive" in O'Day's (2002) study of Chicago schools for two reasons.

First, attention in these schools became focused not so much on student learning per se, but on getting off or staying off probation. This goal places adult desires over the needs of students. Second, to achieve this goal, probation schools exhibited an emphasis on strategies to produce immediate increases in test scores, often to the neglect of long terms success. (p. 17)

O'Day's conclusions from the Chicago study are not entirely negative. She did find significant variability between and among various schools in the Chicago system. However, her explanation for these differences does not support the goals of bureaucratic accountability. "Multilevel analysis of survey data for this rapidly improving group suggests that the differed significantly from other probation schools along several dimensions of initial school capacity; peer collaboration, teacher-teacher trust, and collective responsibility for students learning" (p. 14). Despite the methodological limitation that this observation was made only through survey data, it raises the important point that that these variations between and among schools connect with the elements of teacher agency more than elements of the accountability structure.

Just as O'Day's research in Chicago in 2002 began to uncover the weaknesses of the bureaucratic reform model, similar weaknesses that are seen in the current NYS 3012-C as well as throughout the literature on accountability assessment. In another important literature review from 2003, Laura Hamilton outlines at least four overlapping negative impacts of high-stakes accountability assessment: (1) an impact on instructional practice, (2) particularly a narrowing of practice toward test preparation and basic skills, (3) an

impact on teacher beliefs, and (4) an impact on classroom climate. In two more recent reviews, other authors frame these issues in slightly different ways. Baird (2013) uses the metaphor of “currency” to describe the value assessment has for various groups within a system. As such, Baird suggests that while large-scale testing does have value or currency for policymakers, it often has low value or currency for teachers and students. In fact, Baird suggests that, in some cases, some high-stakes test have no real value in terms of student learning whatsoever and push students and teachers to perform in ways that may have little utility to anything other than the test. Elwood (2013) calls on Messick’s (1989) notions of test validity in order to look at assessment through an ethical lens, examining both the uses and impacts of any testing. Hamilton’s four themes, along with elements of Elwood and Baird, are echoed throughout the many studies into accountability assessment in the last two decades, including many international studies, as well as recent doctoral dissertations in this country.

Internationally, Kellaghan (2001) created a review of literature that demonstrated how much the “audit community” has become a global phenomenon. One study by Dominguez, Vieira, and Vidal (2012) even showed how the increase of international testing has impacted how academic journals themselves have been transformed by the discourse of accountability assessment. In the past decade, the following studies are indicative of the kind of research being done internationally on the impact of accountability assessment on various school systems (Torrence, 2009, in Howie, 2012). Testing on “both side of the Atlantic ... does impact on the curriculum but ... it narrows the curriculum to that which is tested” (p. 488). Six other examples of studies since 2007 demonstrate the impact of accountability assessment in nations that have adopted this policy. Scott (2007) Brown and Gebril (2013) Howie (2012) Klenowski and Wyatt-Smith (2011) Jager et al. (2012).

Perhaps the unifying observation here is that the theory of action associated with accountability assessment not only does not work, but also it has adverse effects on practice. Many of those adverse effects can be described through the lens of teacher agency

in that they describe conditions that reduce teachers' belief that they have the ability and authority to make changes in their practice. These conditions include instructional time lost to testing, an emphasis on national or outsourced curriculum over locally developed curriculum, as well as an emphasis on workbook-based test preparation over curriculum designed to develop more complex thinking and expression.

In terms of dissertations, researchers have been examining the impact of high-stakes testing on teacher and students since the mid 1990s. Like many of the dissertations that focus on agency, the methodology of these studies tends to impact the nature of the findings. For example, Clifford (1995) and Volger (2000) are illustrative of this trend. Both studies rely on survey information to demonstrate a connection between testing policy and teacher-reported changes in behavior. The results are findings that echo the larger body of criticism. Teachers report that they are changing practice to meet the needs of test performance; some report this as a positive refocusing of goals and practice. James (2007) and Yamashita (2011) also conducted survey-only studies and had a focus on elementary teachers only and secondary teachers, respectively. Two studies from Teachers College examined directly the impact of accountability assessment on teachers. Hassler (2011) conducted qualitative research that revealed how teachers see and react to a variety of "accountabilities" in their daily work. Similar to the assessment literacy research, Hassler found that teachers lacked some understanding of where the accountability was coming from and, overall, felt disempowered by assessment linked to accountability. Wills and Sandholtz (2009) coined the phrase "constrained professionalism" to describe how teachers felt about their role in the classroom and in the curriculum within the current policy environment. Their ability to make choices based on professional judgment does exist but is constrained by the external demands on their curriculum choices and on their time. While both studies clearly show how teachers and other stakeholders think and behave within the current assessment climate, they fall short of exploring the possibilities of teacher agency and performance in environments that support formative assessment processes.

Perhaps this quotation by Linn in 2000 still captures the inconclusive nature of studies on accountability testing.

I would like to conclude by summarizing a compelling case showing the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in dramatic ways. Unfortunately, that is not my conclusion. Instead, I am led to conclude that in most cases the instruments and technology have not been up to the demands that have been placed on them by high stakes accountability. Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended negative effects of the high-stakes accountability uses often outweigh the intended positive effects. (p. 14)

Applying the lens of data use research from earlier in this chapter, these studies of accountability assessment fall short in that they do not explain the “complex mechanisms” (Moss, 2012) that link the macro world of accountability policy to the micro world of classroom practice. In other words, these studies do reveal that a link exists between policy and practice—or that bad policy weakens teacher agency and hurts practice—but they never go beyond understanding the significance of that link in order to improve practice or outcomes. They offer evidence that a system of accountability and reform does not work, but they do little to suggest what does.

Appendix C

PADI Performance Task Template A

Cover Sheet

PADI Assessment for:	<u>Grade 4 Students</u>
School District	Suburban Woods Schools – Shore School
Author(s) of this assessment	Lily – Paula – Charles – Process Facilitator –
Teacher(s) who will implement the assessment (if different from above)	Same
Brief Description of the Assessment (what students will do) <ul style="list-style-type: none"> • type of assessment: product, demonstration, process, combination • moments: diagnostic, formative, summative, pre/post • student involvement in self- and/or peer assessment 	<p><u>Type:</u> <i>Performance and Product</i></p> <p><u>Moment:</u> <i>Culminating:; Students will create a single “media truth-navigation guide” in the form of an iBook or iMovie that is used with community members who are professional members of the news media (one community member with small group of students) who will discuss the draft of the media survival guide before it is distributed to the community.</i></p> <p><i>Diagnostic and Post-Assessment: response to pre-assessment questions and post assessment task.</i></p> <p><i>Essential questions:</i></p> <p><i>Where does the truth “lie”?</i></p> <p><i>How do we construct our own story of an event or issue?</i></p> <p><i>Student learning: instruction and formative assessment on conditions and experiences of Native Americans in the form of written responses to questions.</i></p> <p><i>Modeling of Multiple Perspectives in Native American Historical Accounts and authors craft techniques and student creation of multiple perspective text analysis.</i></p> <p><i>Students create three historical accounts from three different perspectives on the same topic within Native American History in groups of three.</i></p> <p><u>Student involvement:</u></p>

	<p><i>Students will....</i></p> <ul style="list-style-type: none"> -Co-construct a checklist for analyzing multiple viewpoints in history/media. -Co-construct a checklist for coming to one's own truth when reading multiple perspectives.
<p>Brief Description of how results/evidence will be used by the teacher</p>	<p><i>Teacher(s) will...</i></p> <p><i>Teacher will provide additional/differentiated feedback via conferencing and written comments on content learning and perspective analysis after design learning opportunities.</i></p> <p><i>Teacher will provide feedback on the article creation using a rubric and structure the final presentation groups based on patterns, which emerge in the final presentation groups.</i></p> <p><i>Teacher will provide written feedback on the "Media Survival Guide," and distribute it to the wider community.</i></p>
<p>Estimated duration of the performance task (from the time students begin to work on it to their final submission)</p>	<p><i>6 Weeks.</i></p>

Section 1: Standards/Outcomes Assessed by the Task(s)

Notes:

- If you are designing an **ELA** assessment, you will use the **CCSS for Literacy**. You may or may not have content standards.
- If you are designing a **Math** assessment, you will have the **CCSS for Mathematics and Mathematical Practice Standards**.
- If you are designing for **SS, Science, Performing Arts, Health, PE, LOTE**, you will have your **content standards** and the **CCSS for Literacy**.
- If you are an **IB school/district**, it is your decision as to whether the IB standards replace or supplement the content and/or CCSS.
- Include any relevant dispositional (i.e., perseverance, open-mindedness) and other significant outcomes and skills (i.e., perspective taking, self-management).

List outcomes, standards and indicators here, with codes:

Reading for Information and Literature

RI.4.9 Integrate information from two texts on the same topic in order to write or speak about the subject knowledgeably.

RI.4.6 Compare and contrast a firsthand and secondhand account of the same event or topic; describe the differences in focus and the information provided. (CCR Anchor Standard: Assess how point of view or purpose shapes the content and style of a text).

Writing

W.4.3 Write narratives to develop real or imagined experiences or events using effective techniques, descriptive details, and clear event sequences.

Speaking and Listening

SL.4.1

[CCSS.ELA-Literacy.SL.4.1](#)

Engage effectively in a range of collaborative discussions (one-on-one, in groups, and teacher-led) with diverse partners on *grade 4 topics and texts*, building on others' ideas and expressing their own clearly.

[CCSS.ELA-Literacy.SL.4.1.a](#)

Come to discussions prepared, having read or studied required material; explicitly draw on that preparation and other information known about the topic to explore ideas under discussion.

[CCSS.ELA-Literacy.SL.4.1.b](#)

Follow agreed-upon rules for discussions and carry out assigned roles.

[CCSS.ELA-Literacy.SL.4.1.c](#)

Pose and respond to specific questions to clarify or follow up on information, and make comments that contribute to the discussion and link to the remarks of others.

[CCSS.ELA-Literacy.SL.4.1.d](#)

Review the key ideas expressed and explain their own ideas and understanding in light of the discussion.

Social Studies

NYS K-8 Social Studies Skills – Historical Thinking – Grade 4 – Identify multiple perspectives from a historical event.

NYS Grade 4 Content Understandings – 4.2 a, b, c (*Native Americans*); 4.5 (*In Search of Freedom and a Call for Change*).

Dispositional

Perspective-Taking

Critical-Thinking (North Shore Shared Value Outcome – Rubric Under Development by NSCSD)

Section 2: Task(s) Description & Teacher Support

(with explicit alignment to standards and outcomes in parentheses)

Supporting rubrics: <i>- Rubric for Deconstructing/Constructing Multiple Perspectives</i> <i>-Media Literacy Guide Checklist/Rubric (Needs to be Created)</i>	Supporting Checklists: <i>-Perspective Taking Checklist</i> <i>-Collaborative Discussion Checklist</i> <i>-Perspective Writing Checklist</i>
---	---

If teachers need to prepare students for the assessment with instructions or resources, list these below as chronological steps.

<p>Pre-Assessment</p> <p><i>Students Respond to a Likert Scale Survey on the following four questions: (1 strongly agree to 5 strongly disagree)</i></p> <ul style="list-style-type: none"> • <i>What I read in newspapers is true?</i> • <i>What I read online is true?</i> • <i>What I watch on the news (TV) is true?</i> • <i>What I read in textbooks is true?</i> <p><i>Students write a brief response to the question: There are 359 million histories of the United States. Do you agree or disagree with this statement? Explain your thinking.</i></p> <p>Standards for Pre-Assessment</p> <ul style="list-style-type: none"> • NYS K-8 Social Studies Skills – Historical Thinking – Grade 4 – Identify multiple perspectives from a historical event. 	<p>Post Assessment</p> <p><i>Students Respond to a Likert Scale Survey on the following four questions: (1 strongly agree to 5 strongly disagree)</i></p> <ul style="list-style-type: none"> • <i>What I read in newspapers is true?</i> • <i>What I read online is true?</i> • <i>What I watch on the news (TV) is true?</i> • <i>What I read in textbooks is true?</i> <p><i>Students write a brief response to the question: There are 359 million histories of the United States. Do you agree or disagree with this statement? Explain your thinking.</i></p> <p>How do we construct our own story of an event or issue?</p> <p>Students write their own version of the “lunch-room fight” scenario in the form of a newspaper article and also describe their thought process in finding their truth.</p> <p>Standards for Post-Assessment</p> <ul style="list-style-type: none"> • NYS K-8 Social Studies Skills – Historical Thinking – Grade 4 – Identify multiple perspectives from a historical event. • RL.4.6 Compare and contrast the treatment of similar themes and topics (e.g., opposition of good and evil) and patterns of events (e.g., the quest) in stories, myths, and traditional literature from different cultures.
---	---

<p>8. Students will co-construct (as a whole class) a brainstorm list on the two questions. Disposition: Perspective Taking</p> <p>11. Students will construct a specific understanding of the causes and effects of key historical events such as:</p> <ul style="list-style-type: none"> • The Columbian Exchange • The Indian Removal Act/Trail of Tears • the establishment of reservations, the Indian Wars, and the genocide of Native Americans • Mascots <p>by analyzing images, primary source documents, and video. NYS Social Studies 4.2; 4.5</p> <p>12. Students will examine multiple textbook and video accounts which show different versions of the same events in Native American History and will explain how those narratives are different. NYS K-8 Social Studies Skills – Historical Thinking – Grade 4 – Identify multiple perspectives from a historical event.</p> <p>14. Students will then add to the double sided list of:</p> <ul style="list-style-type: none"> • moves/techniques that creators of historical narratives utilize in creating their “truth” • the list of techniques used to find their own “truth” when reading the sources. RI.4.9; RL.4.6 	<p>9. This brainstorm will be given to students and modified throughout the project time period; students will be given a “mini-version” of this chart.</p> <p>10. Content instruction and formative assessment on conditions and experiences of Native Americans before the arrival of European Settlers (1500–1900).</p> <p>11. Teacher provides documents, images, videos which illustrate the multiple perspectives of specific events in Native American history. Teacher will model the use of the brainstorm lists by using two different textbook accounts of Native American culture.</p> <p>13. Teacher will provide instructions on how to use the one side of the double sided chart pertaining to the techniques that creators of historical narratives utilize, in order to provide instructions on how to create a perspective based historical narrative/account. Teacher will assess the students using the perspective-taking rubric as they come to their own story of the theme that they created a perspective based writing for.</p>
--	--

<p>15. Students create three historical accounts from three different perspectives on the same topic within Native American History in groups of three. Students will use a checklist to create this activity and will explicitly state which techniques were used. These articles will be “published” in a classroom newspaper; students will then discuss topics in the classroom newspaper other than the one they worked on, to find their person “story” of that different theme. Students will provide feedback to each other using the perspective based writing checklist.</p> <p>Disposition: Perspective Taking W.4.3 Write narratives to develop real or imagined experiences or events using effective techniques, descriptive details, and clear event sequences.</p> <p>18. Product Based on all of these performance experiences, students will create a single “media truth-navigation guide” in the form of an iBook or iMovie that is distributed to the community. A checklist will be used by students for this assignment. Each student will create a section of the complete “media-truth-navigation guide” and each section will be assessed with a rubric by the “editorial board (principal and two teacher designers of this project); students will submit revised sections after the first assessment with the rubric. The class as a whole will “test drive” the guide in analyzing three perspectives on the same news story from three different news outlets, with community members who work in journalism. Small groups of students will discuss the media guide with one community member in the form of a socratic discussion and use the discussion checklist using the</p> <p>Disposition: Critical Thinking CCLS – Listening and Speaking Standard 4.1a-c</p> <p>20. Students reflect on discussion using the discussion SL4.1a-c and outline possible changes to the iBook.</p> <p>21. Students make final modifications to the iBook and distribute it to the community. SL4.1.d</p> <p>22. Students respond to Post-Assessment. Disposition: Perspective Taking</p>	<p>16. Teacher will provide instructions on the criteria for creating a component of the iBook.</p> <p>17. Teachers will assess students on their assigned portion of the iBook creation using the iBook creation rubric.</p> <p>19. Teacher summarizes feedback from community members and outlines suggested changes. SL4.1.d</p>
---	--

**Rubric for Formative Feedback and Summative Evaluation
(with explicit alignment to outcomes and standards)**

Rubric for Deconstructing/Constructing Multiple Perspectives

- Insert student-friendly titles for rubric levels.
- Name and define dimensions for students.
- Include outcome and standards alignment in dimension boxes, to ensure tight alignment.
- Supporting tool: *Rubric Checklist*

Dimensions	Beginning	Growing	Meeting Standard (Level 3)	Beyond Standard
Reading for Multiple Perspectives (RI 4.6)	Identify the main purpose of a text, including what the author wants to answer, explain, or describe.	Distinguish their own point of view from that of the author of a text.	Compare and contrast a firsthand and secondhand account of the same event or topic; describe the differences in focus and the information provided.	Analyze multiple accounts of the same event or topic, noting important similarities and differences in the point of view they represent.
Writing for Multiple Perspectives (W 4.3)	- summarizes details from the text without identifying key ideas - no explicit connections to other texts or to research	- describe ideas in the text in general terms - include only obvious connections between texts	clearly identify key ideas from the texts -include specific connections between texts that cite important similarities and differences	clearly identify key ideas from the texts with specific references -include connections between texts that cite important and sometimes subtle similarities and differences
Social Studies - Understanding Multiple Perspectives (Historical Thinking Standards)	Describe a historical event in a world community.	Identify similarities and/or differences between him/her and others.	Identify multiple perspectives from a historical event.	Describe and compare events in the history of the Western Hemisphere in societies in similar chronological contexts and in various geographical contexts.
Perspective Taking	Other people can think whatever they want - it doesn't change my mind about what I think or do Disrespectful of other points of view	I like to talk to people about what I am thinking or planning so that they can tell me if it's a good idea Respectful of some points of view	It helps me to read about or talk to people with different ideas so that I can see things in different ways Respectful of other points of view	I like to read or hear many different points of view so I can think of interesting, new ideas or questions Support demonstrates respect of other points of view

Perspective Taking Checklist

- I've read things that have helped me see things in a different way. Agree. Disagree
- I've talked to people with different ideas so that I can see things in different ways. Agree. Disagree
- I changed my mind because I listened to what others think or believe. Agree. Disagree
- I heard and was respectful of a different point of view. Agree. Disagree
- I can restate a point of view that is different from my own. Agree. Disagree

Collaborative Discussion Checklist

- Engage effectively in a range of collaborative discussions (one-on-one, in groups, and teacher-led) with diverse partners on *grade 4 topics and texts*, building on others' ideas and expressing their own clearly.
- Come to discussions prepared, having read or studied required material; explicitly draw on that preparation and other information known about the topic to explore ideas under discussion.
- Follow agreed-upon rules for discussions and carry out assigned roles.
- Pose and respond to specific questions to clarify or follow up on information, and make comments that contribute to the discussion and link to the remarks of others.

Double Sided - iBook Creation Checklist/Rubric (Needs to Be Created)

Section 3. Design, Piloting and Implementation Process

Supporting tools: *Reliability and Bias Rubrics*

Designers and Design Process

<p>Identify the people who designed this assessment and the expertise that they brought to the design (level of teaching, disciplines, degrees) Example: <i>This assessment was designed by teachers from <u>district(s) X</u>. <u>All teachers were certified in Y content and level area</u>. The teachers in the program had a combined total of <u>X years</u> of teaching experience and the average length of the team's teaching experience was <u>X years</u>. The design process was facilitated by district personnel and an external support organization, Learner-Centered Initiatives, Ltd. LCI, Ltd. has a 20-year history assisting teachers in the design and use of diversified assessment measures. The assessments were designed as a part of a professional development program that supported the teachers' understanding of quality assessment design practices as well as the creation of this assessment.</i></p>	
---	--

Summarize the drafting and revision process.	
<p>Describe how designers attended to measurement error and other threats to validity. Consider the following strategies when you describe what you did.</p> <ol style="list-style-type: none"> 1. received feedback from different groups of stakeholders on the task design and components 2. sought out student feedback on the task during the pilot or field testing process 3. collected anchors and exemplars, as well as inter-rater reliability data, during field testing 4. analyzed student work from a heterogeneous group of students 5. participated in a “final eyes” review of the assessment task and components 6. if the task includes controversial topics, we sought out feedback from [REDACTED] 7. explicitly attended to differentiation and the needs of our diverse learners in the task design 8. considered Universal Designs for Learning attributes when creating documents for students 9. analyzed student work from a heterogeneous group of students 	

Appendix D

PADI Performance Task Template B—Pre/Post

Cover Sheet

PADI Assessment for:	<u>Art, Social Studies and Literacy: Grade 4</u>
School District	Suburban Woods Schools - Hill School
Author(s) of this assessment	Tina (Art Teacher) Amy (Classroom Teacher) Laura (Process Facilitator)
Teacher(s) who will implement the assessment (if different from above)	Unknown
Brief Description of the Assessment (what students will do) <ul style="list-style-type: none"> • type of assessment: product, demonstration, process, combination • moments: diagnostic, formative, summative, pre/post • student involvement in self- and/or peer assessment 	<u>Type:</u> Combination assessment: product and demonstration <u>Diagnostic:</u> Students will answer the essential question, How do our beliefs impact what we see and do? <u>Formative:</u> Student self and peer assessment of their work of art and essay Teacher feedback on.... Summative: Students will create work of art and essay for an exhibit of their choice that addresses the question How do our beliefs impact what we see and do? Students will: <ul style="list-style-type: none"> • generate criteria for each of the rubrics, and use it to evaluate their work and each others' work • Select their particular work of art • Write a persuasive essay responding to the essential question, How do our beliefs impact what we see and do?
Brief Description of how results/evidence will be used by the teacher	Teachers will lead students in creating an exhibit including artwork and writing for other students.

Section 1: Standards/Outcomes Assessed by the Task(s)

Notes:

- *If you are designing an **ELA** assessment, you will use the **CCSS for Literacy**. You may or may not have content standards.*
- *Include any relevant dispositional (i.e., perseverance, open-mindedness) and other significant outcomes and skills (i.e., perspective taking, self-management).*

List standards and indicators here, with codes and complete as many templates as needed if you have more than one pre- and post-

PRE POST

<p>Art: Students brainstorm and work as a group to create a work of art with various materials in the classroom.</p> <p>VA: Cr1.1.4: Brainstorm multiple approaches to creative art or design problem.</p> <p>VA: Cr1.2.4: Collaboratively set goals and create artwork that is meaningful and has purpose to the makers.</p> <p>VA: Cr2.1.3: Explore and invent art-making techniques and approaches.</p>	<p>Art: Brainstorm and work as a group to create a site-specific piece using materials from that specific environment.</p> <p>VA: Cr1.1.4: Brainstorm multiple approaches to creative art or design problem.</p> <p>VA: Cr1.2.4: Collaboratively set goals and create artwork that is meaningful and has purpose to the makers.</p> <p>VA: Cr2.1.3: Explore and invent art-making techniques and approaches.</p>
<p>Literacy: Students will write an on-demand persuasive essay to the prompt How do our beliefs impact what we see and do? (CCSS.ELA-Literacy.W.4.1a-d)</p> <p>CCSS.ELA- Literacy.RI.4.3 Explain events, procedures ideas or concepts in a historical, scientific, or technical text including what happened and why, based on specific information in text.</p> <p>CCSS.ELA- Literacy.RI4.7 Interpret information presented visually, orally or quantitatively and explain how the information contributes to an understanding of the text in which it appears.</p> <p>CCSS.ELA-Literacy.W.4.1 Write opinion pieces on topics or texts, supporting a point of view with reasons and information.</p>	<p>Literacy: Students will write a opinion essay How do our beliefs impact what we see and do? (CCSS.ELA-Literacy.W.4.1a-d) citing examples from various text, notes and other resources.</p> <p>CCSS.ELA- Literacy.RI.4.3 Explain events, procedures ideas or concepts in a historical, scientific, or technical text including what happened and why, based on specific information in text.</p> <p>CCSS.ELA- Literacy.RI4.7 Interpret information presented visually, orally or quantitatively and explain how the information contributes to an understanding of the text in which it appears.</p> <p>CCSS.ELA-Literacy.W.4.1 Write opinion pieces on topics or texts, supporting a point of view with reasons and information.</p>

<p>CCSS.ELA-Literacy.W.4.1.a Introduce a topic or text clearly, state an opinion, and create an organizational structure in which related ideas are grouped to support the writer's purpose.</p> <p>CCSS.ELA-Literacy.W.4.1.b Provide reasons that are supported by facts and details.</p> <p>CCSS.ELA-Literacy.W.4.1.c Link opinion and reasons using words and phrases (e.g., <i>for instance, in order to, in addition</i>).</p> <p>CCSS.ELA-Literacy.W.4.1.d Provide a concluding statement or section related to the opinion presented.</p>	<p>CCSS.ELA-Literacy.W.4.1.a Introduce a topic or text clearly, state an opinion, and create an organizational structure in which related ideas are grouped to support the writer's purpose.</p> <p>CCSS.ELA-Literacy.W.4.1.b Provide reasons that are supported by facts and details.</p> <p>CCSS.ELA-Literacy.W.4.1.c Link opinion and reasons using words and phrases (e.g., <i>for instance, in order to, in addition</i>).</p> <p>CCSS.ELA-Literacy.W.4.1.d Provide a concluding statement or section related to the opinion presented.</p>
<p>Social Studies: 4.2 Native American Groups and the Environment: Native American groups, chiefly the Iroquois and Algonquian-speaking groups, inhabited the region that became New York. Native American Indians interacted with the environment and developed unique cultures.</p>	<p>Social Studies: 4.2 Native American Groups and the Environment: Native American groups, chiefly the Iroquois and Algonquian-speaking groups, inhabited the region that became New York. Native American Indians interacted with the environment and developed unique cultures.</p>

Section 2: Task(s) Description & Teacher Support

(with explicit standard alignment in parentheses)

Complete as many templates as needed if you have more than one pre- and post-

If teachers need to prepare students for the assessment with instructions or resources, list these below as chronological steps.

Art: Students will be presented with the idea of earthworks and site -specific artwork. Students will interpret and explore meaning in various examples of earthworks.

Writing: Students will use persuasive writing to create a statement about their art and message.

All of the above work will be presented at a combination Art & Multimedia Exhibition.

PRE

<p>What will the students do or produce to demonstrate their understanding and abilities? What is the task?</p> <ul style="list-style-type: none"> Describe what students will do, in step-by-step fashion. Place standard codes in parentheses where appropriate. 	<p>What will the teacher(s) do to support student learning? How will feedback be embedded in the process?</p> <ul style="list-style-type: none"> Describe what the teacher will do to support students during the assessment. Describe differentiation strategies, feedback strategies and how assessment will inform instruction for all students.
<p>Art:</p> <p>1.Students answer the questions "What do you believe are materials that could be used to that could be used to create a work of art?" And "How do your beliefs affect your artistic choices?"</p> <p>2. Students discuss work of Andy Goldworthy</p> <p>4. Students will create a Rubric to evaluate further</p>	<p>Art</p> <p>1.Teacher presents the work of Andy Goldsworthy and Earthworks.</p> <p>3. Teacher will give students a framework for creating a rubric that measures the standards above.</p>

<p>work.</p> <p>7. In an outdoor environment, students will <u>explore and invent art-making techniques and approaches</u> using available materials. (VA: Cr2.1.4; SVO: 5.E)</p> <p>9. In small groups, students will <u>collaboratively set goals</u> for their end product and process and to decide on one meaning and method. (VA: Cr1.2.4, SVO: 1.A)</p> <p>10. Each group will <u>collaboratively create artwork that is meaningful</u> and intended to affect how the viewers see in the materials and the site? (VA: Cr1.2.4; SVO: 5.E) Students document their process and completed work through photographs.</p> <p>12. Individual students will respond to questions. "How has your belief about art materials changed based on your experiences with this project?"</p> <p>16. Students self assess using rubric and participate in a peer critique.</p>	<p>5. Teacher will review rubric with all students and teach how to use it to guide their work.</p> <p>6. Teacher will teach and model how to experiment with materials and use them in a site specific piece</p> <p>8. Teacher will assign groups according to like ideas.</p> <p>9. Teacher will support student problem solving by giving one on one support with possibilities, material usage.</p>
--	---

<p>ELA:</p> <p>On demand persuasive writing on How do our beliefs impact what we see and do ? (CCSS.ELA-Literacy.W.4.1) (SVO: Comm: items a, b and c/ Growth in Self: items c, f and h)</p>	<ul style="list-style-type: none"> • Introduce the Essential Question: How do our beliefs impact what we see and do? • Lead a discussion about what this may mean? • Allow time for students to complete a pre-assessment writing on-demand.
--	--

If teachers need to prepare students for the assessment with instructions or resources, list these below as chronological steps.

POST

What will the students do or produce to demonstrate their understanding and abilities? What is the task?	What will the teacher(s) do to support student learning? How will feedback be embedded in the process?
<p>Art:</p> <p>2. Students will review Rubric from pre to evaluate further work.</p> <p>3.. In an outdoor environment, students will <u>explore and invent art-making techniques and approaches</u> using available materials. (VA: Cr2.1.4; SVO: 5.E)</p> <p>4. In small groups, students will <u>collaboratively set goals</u> for their end product and process and to decide on one meaning and method. (VA: Cr1.2.4, SVO: 1.A)</p>	<p>Art</p> <p>1. Teach will lead a class discussion about effective collaboration and review project goals. Class will discuss the ethical choices they will have to consider (ex. picking a flower to use in your art will kill it; is that ok?)</p> <p>5. Teacher will teach and model how to experiment with materials and use them in a site specific piece</p>

<p>6. Each group will <u>collaboratively create artwork that is meaningful</u> and intended to affect how the viewers see in the materials and the site? (VA: Cr1.2.4; SVO: 5.E) Students document their process and completed work through photographs.</p> <p>8. Individual students will respond to questions. "How has your belief about art materials changed based on your experiences with this project?" "How did your experience now compare to your experience in the fall?"</p>	<p>7. Teacher will support student problem solving by giving one on one support with possibilities, material usage.</p>
<p>Writing:</p> <p>2. Students will create a checklist of required components of a persuasive essay, to be used as they work through and develop their own essays.</p> <p>1. Students will generate topics for persuasive essays.</p> <p>5. Students will draft during Writing Workshop. (CCSS.ELA-Literacy.W.4.1a-d L.4.2 L.4/6)</p>	<p>Writing:</p> <p>1. Teacher will give students examples of persuasive text. In groups, students will read and identify the essential components of a persuasive text. (CCSS.ELA-RI4.1 & 4.9)</p> <p>4. Teacher will conference and work with students to identify and refine "angles" for their pieces. (CCSS: ELA W.4.5)</p> <p><i>(Although the post assessment will be an opinion essay, the students will be acquiring the information needed to write it during ELA and social studies time. ELA time will be used to teach how to write an opinion piece and Social Studies time will be used to teach content as it relates to the Native Americans.)</i></p> <p>6. Teacher will review checklist with the group and make revisions (based on new information and student input), as necessary. (CCSS: ELA W.4.5)</p>

<p>7. Based on knowledge gained from notes, readings, videos, field trips, etc. students will write opinion essays to answer the question, How do our beliefs impact what we see and do? (CCSS.ELA-Literacy.W.4.1a-d L.4.6, L.4/2)</p> <p>9. Students will peer conference to edit and revise work. (CCSS: ELA W.4.5)</p>	<p>8. Teacher will review the rubric with students to review expectations for their writing. (CCSS: ELA W.4.5)</p> <p>10. Teacher will facilitate publishing of opinion pieces to incorporate them into the exhibit featuring earthworks.</p>
<p>Notes:</p> <p>How do our beliefs impact what we see and do?</p> <p>Include periodic reflections throughout the unit as opinions and beliefs change over time.</p> <p>End-of Year: How and why do beliefs change?</p>	

Opinion Writing - 4th Grade- Possible checklist

Planning Checklist

- ☐ I read and understood articles and/or books about my topic.
- ☐ I investigated different aspects about my topic.
- ☐ I took accurate notes about what I learned.
- ☐ I thought carefully about what I learned.
- ☐ I used what I read to form an opinion about my topic.
- ☐ I listed reasons for my opinion.
- ☐ I found facts and details that support my reasons for my opinion.
- ☐ I categorized the information I found and grouped related ideas together.
- ☐ I made a list of my sources.
- ☐ I made an outline that answers the requirements for the assignment.
- ☐ I considered my audience when planning my paper.
- ☐ I spent enough time researching and thinking about my opinion on this topic.

**Rubric for Formative Feedback and Summative Evaluation
(with explicit standards alignment)**

Rubric for _____

- Insert student-friendly titles for rubric levels.
- Name and define dimensions for students.
- Include standards alignment in dimension boxes, to ensure tight alignment.

Dimensions	Level 1- Novice	Level 2- Developing	Meeting Standard (Level 3)	Level 4- Highly Effective
CCSS R.4.1 Refer to details and examples in text when explaining what the text says explicitly and when drawing inferences from text.	Attempts to reference source material, but refers to few or no details or examples from the provided sources. References to source material demonstrate inaccurate understandings of the details or ideas in the text.	Refers to at least one detail or example from the provided source. Some details may not support the idea from the text the student is discussing. References to source material demonstrate a mostly literal understanding of the text.	Refers to more than relevant detail and/or example from the provided sources(s). References to source material demonstrate an accurate understanding of literal and inferential details from the text.	Selects the most relevant details and examples from the provided sources(s) to support the main claim. Elaborates on source material to demonstrate an accurate and insightful understanding of literal and inferential details from the text.
Write opinion pieces on topics or texts supporting a point of view with reasons and information.				

Position	Attempts to make a claim, but does not take up a particular side of the issue. May introduce the general topic rather than stating an opinion about the topic.	Makes a claim that connects to the given topic. The claim may not make clear which side of the topic the writer will support.	Makes a claim that connects to the given topic and takes a clear position.	Makes a claim that takes clear position/demonstrates the writer's understanding of the complexity of the issue
Structure: Introduces topic; provides a concluding statement	Gestures towards and introduction and/or a conclusion. These sections may go off on slight tangents from the main claim, relating to the topic generally but not addressing the main argument.	Provides a very brief introduction, which may not connect closely to the claim. Provides a conclusion that may restate the claim.	Provides an introduction to the claim, clearly announcing that this is an argument. Attempts to inspire readers to care about the topic and/or claim. Provides a conclusion that connects to the writer's main claim.	Provides an introduction that orients the reader to what is most important in the argument. Concludes the essay with a section that highlights important points and facts from across the rest of the piece or brings in new effective evidence.
Structure: Creates and organizational structure	Organizes reasons into a list-either through a preview of the reasons in an introduction or by creating body paragraphs. Some sections are better defined than others.	Reasons and examples are grouped so that readers can follow the writer's "train of thought." Sections are mostly well defined.	Uses paragraphing to group supporting ideas and their relevant evidence. It's clear how each section has been organized.	Orders paragraphs in a structure that is clearly planned; either demonstrates least to most importance, chronological order, or follows the flow of the research text.

Structure: Transitions	Attempts to use transitional words and phrases to connect opinion and reasons; may do so inconsistently, overuse them or use the inappropriately at times.	Ideas and reasons are connected using linking words (for example, because,)	Uses words and phrases to connect different parts of the piece together; to demonstrate shifting from reasons to evidence (e.g. "for instance") or to introduce a new point (e.g. "in addition").	Uses transition words and phrases to connect evidence to reasons using phrases like, this show that... <i>Help the reader move through the essay with phrases such as, another reason, the most important reason."</i>

Student Created Rubric for Art

	Beginner	Getting There	Got It!	Expert
Comes up with new ideas	I am unclear ideas. I am not sure where to begin.	I have some ideas but can't decide which is best. Some of my ideas are off topic but I like them anyway.	I have a variety of creative ideas that are realistic and will accomplish the task.	I have lots of unique ideas. and can use them. I am prepared and my plans are well thought out.
Flexibility of mind	I start over when things go wrong. I am focused on only one idea even if it is not working.	I get frustrated with problems.. I hears others ideas, but still think my ideas are the best.	I can work through problems. I value the team's opinions and am open to ideas.	I perseveres in spite of problems. I can problem solve. I can understand and use a number of different ideas at once.
Artwork is meaningful or has a message	I can make something, but there is no deeper meaning or I am unsure if there is a deeper meaning.	My meaning is unclear until I explain it. Some of my choices help communicate the meaning.	I can explain the meaning of my work and how my choices help to communicate that meaning.	I can make something that clearly represent my meaning. I can use tools, such as symbols, to help communicate.
Collaborative goal setting	I am focused on what I want to do more then the goal of the project or my teammates.	I have ideas in mind but am not sure about the steps to take. I am focused on getting the group to agree with my idea.	I am working with my team to set goals and consider options together with respect and interdependence.	I consider all team options and includes everyone's point of view in the finished work. Everyone is heard and everyone's skills are used.
Artwork and/or folder shows new techniques were invented or explored.	I am happy with my first idea. I do not need to explore new techniques, I already have an idea.	I can come up with one or two ideas but is unsure about techniques.	I came up with some new techniques. They are shown through examples and evidence of brainstorming in my folder.	I have detailed plans and unique ideas shown through examples. My folder shows that I brainstormed many techniques before settling on one.

Artwork communicates a new way of seeing the world, art or the materials used.	I can make something using the materials. I am not sure what it communicates.	I come up with an idea that I want to communicate, but the limited materials or time make it difficult.	I can communicate a new way of seeing, even though I have limited materials and time.	I can make choices about what materials would best communicate my idea. I can understand what each different material might communicate.
--	---	---	---	--

Appendix E

Pilot Interview 1, April 1, 2015

- I: The first questions are a little bit narrative. I actually have a copy of the blueprint as well which I'll ask some questions about. But can you guys take turns to describe the work that you guys have been doing since the pre-assessment in October?
- F1: Okay, I'll start. So looking back at my notes, I think we started with the essential question of what does it mean to make a difference. That's what we wanted to get from the students. And then we determined what it would look like, what did we want to see from them. So starting with that as a basis. There were two things for student performance. One was applying being able to be aware of what it means to be a caring citizen in the classroom, school, and community. And then secondly, to demonstrate how they can make a difference in another person's life and inspire them to do the same. So if that was our goals, then (pause) then we just built lessons to accomplish that. So we developed like lessons to do that.
- F2: Do you want specifics as to what we did?
- I: Yeah, so a couple specific examples would be helpful.
- F2: So in my classroom, and I know Lily used some of the same literature, I started off the year with books, literature, things like *Ordinary Mary's Extraordinary Deed* and just talked about how making a difference to someone inspires them to make a difference and then that kind of comes back to you. So it was a lot of conversation in the beginning and then what we did was talk about how can we—what would it look like and sound like to make a difference just in the classroom. And so I was there to guide them through and make reference if they did and say how that made a difference and how that made someone feel. And then we kind of expanded that to the

school. And then what we did was we made posters to display around the building to remind people to make a difference and be kind and it also tied into something that Peggy does at the beginning of the year, being(?) the Bucket Program. So then after that, we wanted to go out to the community and so at holiday time, my class did a presentation on how they made a difference to their families and what they then did was say to them, "We'd love you to make a difference to someone else," and they handed them a piece of paper and the parents would write in when they did something and then sent it back in. That's kind of just a little snippet of what we did, but we plan on ending the year with the students writing a letter to inspire someone else to do a difference, with an index card saying, "When you do make a difference, please send it back in," so then it's more global, not just within our community.

I: So that would be the performance, the final assessment for you guys.

Both: Right, yes.

F2: That would be the performance and then we'd go back to—then I would give the post of what I gave at the beginning or different ____.

I: And what did you guys give specifically? Just a little more detail on what you gave at the beginning of the year. What was the—

F1: It was a one-page (pause) this paper saying, "Draw a picture to show how you've made a difference in one of the places below: classroom, school or community. Tell in words why your actions made a difference." So that would be before any lessons. And then we gave, we do all of the lessons, they go through all their experiences and responses, and then we gave them the same thing after and then compared the two.

I: That's great. So I would describe this looking at it as like a narrative writing prompt, sort of typical for the beginning of the grade.

F2: And as facilitator, as we came back to school (pause)

I: ____ yes.

F2: Do you want to say more?

I: No, no, I just asked if I could have a copy and I wasn't listening, I'm sorry.

F2: That's okay. I was just checking in with my colleagues making sure everybody's okay and we needed some extra time to write our lessons because—and our rubric really wasn't set at the beginning of the year, so we felt like to jump in completely this year might not have been a great idea. We really weren't feeling ready. So I was able to get us release time. We met, we revised our rubric, we made lists of literature, resources. We revised our pre- and our post-tests. So we had that time together as a group outside the classroom to kind of regroup and catch up and make sure we felt good about things.

F1: And actually, the more we talked about it, the more it changed. Some of our lessons changed and the focus and the activities changed. Remember it started off very research-based and then we simplified it, yes. So the plan actually changed over time.

F2: It did. It changed this year what we—were we shadowed by Cohort Four this year? We were in the fall. So at our first meeting back in the fall, we said, "Wouldn't it be nice for Cohort Four to kind of watch us?" and so before they even met with Gisele the first time, they shadowed us at our meeting. And so we actually got some feedback from them about our project, which led us to revise and edit our whole project really, like Lily said. It was more research-based, teaching research skills, rather than this letter, this narrative or this opinion letter as our performance assessment.

I: So can you say a little bit about how the group came to that decision?

F2: I think Cohort Four gave us a lot of insight. By sharing it with them, they were questioning us which then made us question ourselves. Like was it too much for the second graders? Was it really them making a difference, you know? Where this became more personal a reflection and then inspiring someone else to do something. I just think it was really planned and it sounded great(?) where we were going to go out into the community, but it wasn't necessarily attainable in the amount of time that we had. I mean, this is (pause) them really reflecting, making the connection, and then trying to inspire others, and so it's still making a difference in the community, but in all different various ways, not in the way that we came up with.

I: Understood.

F1: And I'd like to piggy-back on what she was saying. You know, thinking back on this, I think what happened was in all those Gisele meetings, the summer intensive days of planning and following the blueprint outline, you get very involved in developing a plan and it's very ideal. When you bring it back to people who are not privy to that plan and just basic teachers, they're looking at it saying, "Really? You're going to do all that?" and then that was like, wow, do we really need to do all this? And that's how I think it got a little more simplified.

F2: I think we felt it as well when we sat down at the table and read it for the first time since the summer. We were reading it with those eyes, those new eyes, and we looked at each other and said, "What are we feeling(?)" (laughs)

F1: (overlapping) Right, exactly.

F2: And so because we felt that way and the people listening felt that way, then it led to the revision.

- F1: And I think it was meaningful, but this was more meaningful for the children. I think they felt they would feel more invested and involved than this grand-scale kind of—
- I: Sure. Now on the second grade, are you guys influencing other teachers to do this project or is this something that is really just contained with this group? How did that work?
- F2: At this point, it's just contained within our group. I think people make differences, they may not use that terminology. People talk about ____ and certainly there is the philosophy in our building with being respectful and responsible, but no one else is really doing lessons at this point.
- I: Was there ever a suggestion or expectation that that would do that? You know, you try here and others would try it? Or it was really—
- F1: It was quite the opposite, that we were told that this was for us, that it was never going to be pushed on to anybody else unless they were interested.
- I: Okay. So let's jump back to the pre-assessment. How did that information kind of influence the letter process? I mean, it sounds like you guys had some ideas for lessons, but how did that influence what you guys are sort of about to embark now or have just done?
- F2: I think we found—just jump in if I'm not correct. That pre-assessment is even a little bit different than what we actually administered this year, because based on what we saw, we were like, wait a minute, it's not getting what we want to. Is it the terminology they didn't understand? Is it really the act of making a difference? So we actually tweaked it, so I think using the pre-assessment helped us already reflect and tweak what our new pre-assessment would be next year.
- I: Okay, so just seeing some language issues and—
- F2: Or, right, it was just—

- F1: That the original pre-assessment was aligned with our original performance assessment.
- F2: Right.
- F1: And then when we met in the fall and revised our performance assessment, we had already given the original pre-assessment and we said, well, now that we're revising our performance assessment, we need to look back and say: does our pre-assessment give us the information we need for our now present? So we really weren't—this fall, we really weren't ready to kind of embark on this whole process. But meeting in the fall and re-aligning made us more ready, actually, than we would have been to start this year.
- F2: And what I think was great is our administration allowed us to dabble this year so that we could see what we really were trying to get and accomplish from the kids. So I think dabbling and trying things out and seeing is going to make it more powerful and meaningful next year.
- I: Okay.
- F2: So I'm glad we had that plan. What was your question? (laughs)
- I: No, I think you answered it. I'm on to a different one, which is—and part of me wants to come back and ask this again after you've given the final assessment, but do you feel you're on to something here that you'd like to share with colleagues? I mean, do you feel like, hey, this is something that other people should be seeing or doing?
- F2: I'm very excited by it, about it. It's my big thing of, yes, academics is huge, but I also think being a good person and inspiring others to be good people is important. The kids in my classroom are using this language all the time. We share every day after lunch about how we made a difference and we're tracking it on the chart that ___ on the chart at this point. And if I forget, they're the first ones to remind me, "I want to share how I made a difference."

So I think what I'm excited about is they're internalizing it. It's not just lessons. I feel like they're really trying to apply what they learn and that it's becoming more innate. And so I would love to, but at the same time, let's be honest. There's a lot of things in our curriculum already and I would not want to overwhelm any other teacher with what we're doing already.

F1: Well, it is tied to some of the actual social studies standards, but those haven't really been finalized. So I think it's, for me it would be a question of, well, what standards are actually finalized, what pieces are necessary at that point in time, and then how does this fit into that.

I: Okay, that actually leads to my question of how you guys see this work connecting to other district work.

F2: We use PBIS in the district so they are—the children all use the same language in terms of character education and understanding that we are respectful, responsible, and safe in our building. So this ties in really easily and well. So going back to your previous question about sharing, I agree with Lily and Nicole, but I also feel like it is nice when everybody's using the same language. With PBIS, everybody's using the language. The children go to music, they go to phys ed, whatever, everybody's saying "Respectful, responsible, and safe," and hopefully everybody's kind of understanding what that means. But if we bring in some of the language that we're using—(INTERRUPTION) in terms of language that I'm not saying that we should push this on anybody, but it would be nice to share the language.

F1: What I will say is I feel like I didn't have to really put—I mean, there's lots of big pieces that I added, but the language is just changing what I say. For example, when we're doing character study, it's a great place to say, "Well, how is the character feeling? Was that person making a difference to that person?" You know what I mean? So it kind of like, it really, it may just switch

what you're saying. It's into so many things that we do. In communities, you know, when we're doing rural, urban, and suburban, you can talk about it there too, like what, you know, if everyone's doing their job as a good citizen, does that make a difference to the community and to the people within it? So I think it lends itself nicely without feeling overwhelming.

I: Yeah. Now as a piece of persuasive writing, the letter format is something that would also be shared across the grade, across the grade?

F1: We do opinion writing, so what our letter is is basically sharing their opinion about what they did and at the end saying, "And we hope we can inspire you to do something similar."

F2: But not everybody writes letters for their opinion piece, right?

F1: No.

F2: So it could be in various formats.

I: So my sort of broader question, having been through the Patti(?) training is: how is Patti Smith(?) changing your thinking about some of these curricular areas or approaches? Or not?

F2: I have two ways that I could think of offhand. I constantly say to myself: Is this meaningful? And as Gisele would say, who cares?

F1: Who cares? (laughs)

F2: And when I'm doing things, is it really an authentic audience? So I mean I do walk, I have walked away with making sure when I'm being really reflective and saying, is this meaningful to the kids? Who would care about it, you know?

I: How about you, Lily? What would you say? Has Patti shifted your thinking, challenged your thinking or just ____

F1: I think it focused my thinking. You know, it sets up an outline of a way of developing plans and assessing yourself and I think that it's made me more

aware of assessing myself all the time as I go through the teaching process. So just constantly using feedback and talking and revising and, you know, adapting your teaching to your students.

I: So you guys mentioned changes in their classes in terms of the behaviors around making a difference. How about any academic changes? Have you seen anything shift as a result of this work? (pause)

F2: That's interesting how you would mention ____ (pause)

F1: Well, so the one piece that we added to the assessment was not just naming how you've made a difference but why it makes a difference. And so that level of thinking is a little bit more in depth. So it would support more inferential thinking and cause and effect.

I: Okay. (pause)

F2: For me, I think not necessarily the curriculum it's impacted, but the culture of my classroom, you know. Kids would talk about not only making a positive difference, but what is a negative difference and what would that look like? So it's just then becoming more aware, like if someone drops something on the floor, they're the first one over to help them pick it up rather than just step over it, where I've seen, you know, in the past, like—and yes, there will always be kids who have done that, but there's more this year I feel than there has been.

I: Cool.

F1: I've also made the tie to character development. When you mentioned reading and characters, when we talked about all of this and actions and what it means, bringing it to that level of understanding, "If you do this, this means this." So one of the things that we do in character study is understanding character traits, is to say if you're reading a book and a person says or does this, this shows this. So this is the flip side of that. What you're

doing in real life, if I help you, I am being caring. So when you read a book and the person is being helpful to someone else, you can pick that up about that character. So translating into character traits.

I: Okay, that's great, that's great. Can we talk a little bit about, staying on this thread of influencing the kids? You mentioned the idea of audience. Do you notice anything there in terms of the kids as they—again, I probably should clarify. I don't know where each of you are in terms of the actual letter writing process, whether you're there or not, but can you talk a little bit about how you presented that sense of audience to them and how they're reacting?

F2: I haven't done the letter yet because I want to bring the year to closure with that, like bring up the year reflecting on how it made a difference and then sending them off with enough time to get feedback from anyone who has been inspired. The authentic audience ____ the families, when we presented at the holiday time, which I thought was a good time because that's (pause) ____ talked about being giving and making a difference. And so that was the authentic audience for them. And then what was great was when the families did send that in, what they did to make a difference. For example, one little girl got a puppy but they ended up getting two because the mother is like, "Oh, we can't just leave this one there! They're listening to what you just said." So she wrote thank you to me. (laughs) You know, and how a child went home and made a video for his grandmother to make a difference because she's across the country. So I think that is their authentic audience. And watching what their acts did and influencing others. (pause)

F1: I forget the question.

I: Oh, sorry, it was about the kids' reaction to the sense of audience that the final assignment gives.

F1: Well, I sort of did a little bit differently in that we, we tied into a school event where parents were collecting donations of socks to give to people who needed it, and so when we had gotten to the part of how do you help in the classroom, how do you help in the school, how do you help in the community, we brought that up. And that was one way of helping out and so kids were able to do that. That becomes—so that has sort of grown and evolved accidentally. They brought in socks, we put it into the box, and we thought that would be the end. But then the parents came back and said, “Wow, that’s really great. We think we’re going to contribute them to this audience.” And then they decided to come back and give awards and acknowledge the kids for their caring. So it’s coming back in that way, so it wasn’t meant to be a real live audience, but it’s turning out to be that way.

I: Interesting, interesting.

F1: Yeah, so now the kids, they don’t even know about it yet, but it’s happening tomorrow, and that the parents are going to come in with these awards and acknowledge them for how you made a difference, which the kids completely forgot about. But yeah, that’s, that’s going to be their real live audience.

I: So I am definitely hearing that within this project, two individual teachers, you guys really do feel that you have a lot of choices to make, that you’re doing two major pieces in common but the other pieces are different. How much time do you have to share with each other what you’re doing? And how many other things have you sort of done in common through that sharing process?

F2: Well, like Bonnie said, we had that meeting at the ___ year, but you know, Lily and I move on the fly sometimes and share what we did, or you know, I’ll share resources with her, she’ll share with me, but not again in a format of a controlled study kind of thing.

- I: Okay, okay.
- F2: I just want to bring up too like what we're trying to drive home is doing this when no one's looking.
- I: Yeah.
- F2: That was just a big piece too, just like that they're not always doing it just for acknowledgment. Like I mean, that's what they started to do at the beginning of the year. "Well, I made a difference. See that?" At the beginning of the year, it's great because you want to—and I gave our Good Work awards every time I saw it. But now they're coming up just to get that Good Work award, and I'll say, "Remember, even when no one's looking and we're doing it just to—the reward is its, in itself making you feel good that you helped someone." So we're been trying to scaffold them to do it not just for the acknowledgment but because it makes you feel good.
- I: Okay. You guys have been amazing because in the way you answered some of my questions, you're hitting other ones. I guess I'm going to ask an abstract one at the end, which is, I think it builds a little bit on this idea that this kind of feels like the right stuff. Is this work helping you do work with kids that you feel is the work you should be doing? I know that's sort of abstract, but.
- F2: In terms of them making a difference or in terms of performance assessment?
- I: I guess both. But I wanted to leave it a little open because, you know, I think we live in a time where people are pulled in different directions and I'm just trying to get at is this—does this feel like this is pulling in a good direction for you? And (pause)
- F2: Like I said, I'm very passionate about this, so for me it's pulling me in a good direction and I feel like it pulls kids in a good direction. And I think it really is adding to the culture of my room and hopefully inspiring the kids to do it

outside of the room. But I do, I've always done it. It just (pause), it just the performance assessments are different and the language is different, but I've always been passionate about it, and so that's why I was so excited to do this work to make it even more powerful.

I: Okay, okay.

F1: I think the idea of performance assessments is something that I've always been involved in in that whenever kids learn about something, they need a place to show it. So this is just a formal way of putting it together.

I: Okay, good. Guys, thank you very much for your time.

F2: You're welcome.

I: And I know that it's a busy morning, so I don't want to keep you any longer unless you had a question for me or thought of something that I was not wise enough to ask. (pause)

F1: I'm just curious how this helped you figure out your next steps.

I: Good question. Again, my—and I'm going to turn this off at this point.

(END OF PILOT)