

TEACHER CAPACITY AND ASSESSMENT REFORM:  
ASSUMPTIONS OF POLICY, REALITIES OF PRACTICE

Rebecca H. Berger

Submitted to the faculty of the University Graduate School  
in partial fulfillment of the requirements  
for the degree  
Doctor of Philosophy  
in the School of Education  
Indiana University  
November, 2005

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

---

Cary Buzzelli, Ph.D.

---

Ginette Delandshere, Ph.D.

---

David Flinders, Ph.D.

---

Jesse Goodman, Ph.D.

October 3, 2005

© 2005  
Rebecca H. Berger  
ALL RIGHTS RESERVED

## Dedication

This work is dedicated to the memory of my father, W. Fred Howat:

“How much Social Security do students get when they retire, Beck?”

and to Allen,  
for everything.

## Acknowledgements

There are many individuals whose assistance greatly enhanced my ability to complete this dissertation. I would like to take this opportunity to thank them.

I would first like to thank the members of my committee. All of them contributed richly to my intellectual development long before they agreed to serve on my dissertation committee. Space does not permit me to list all the reasons I am grateful to these individuals, but I remember the following particularly: Cary Buzzelli's mentoring as we team-taught a course; Jesse Goodman's book discussion group; David Flinders's critiques and support of my early attempts to define a research problem; and Ginette Delandshere's willingness to serve as my dissertation director and then to assist me, long distance, through the dissertation process. The benefits I gained from the seminars, classes and insightful discussions I have had with these individuals are too numerous to mention.

A number of people within the University of Maine at Farmington community also aided my research. My thanks go to Dr. Margaret Arbuckle of the Western Maine Partnership for her advice on site selection. I would also like to thank Dr. Kathy Yardley and others at U Maine-Farmington who arranged for release time for dissertation work. My colleagues in the College of Education, Health and Rehabilitation, particularly those in the Early Childhood/Elementary Education Department, have been a source of tangible and intangible support through this process.

I cannot name them, but I am very grateful to the teachers and administrators who participated in this research. They welcomed me into their professional lives for

an extended period, and they were always willing to engage in conversations that were clarifying and enlightening throughout the research process.

My husband, Allen Berger, assisted me in ways too numerous to mention, both professionally and personally. I am particularly grateful (now) for his editing assistance. I would also like to thank our children, Joshua and Moriah, who were ever supportive through this endeavor.

## Abstract

Rebecca H. Berger

### TEACHER CAPACITY AND ASSESSMENT REFORM:

### ASSUMPTIONS OF POLICY, REALITIES OF PRACTICE

An aspect of the standards movement in the late 1980s and early 1990s was a focus on assessment that dominated all levels of formal education. Policy makers attempted to influence educational reform through the adoption of standards that were used to drive curriculum and assessment decisions. Concurrently, other advocates for assessment reform criticized the reliance on standardized testing as a measure of student achievement and proposed the use of other methods to capture student performance. Expanded use of classroom assessment was also encouraged.

Unexamined assumptions about teaching and learning underlie assessment reform efforts; this leads to proposed reforms that serve different ends with little shared understanding of assessment purposes or desired outcomes across populations. Using a qualitative design, this research examined how teachers interpret and enact assessment reform given the unexamined assumptions and contradictions that beset reform efforts. Observation, interviews and document collection formed the bulk of the data. Data were collected and analyzed from individual teachers, administrators, and teacher committees. The analysis utilized the concept of teacher capacity in order to examine both individual and collective responses to reform implementation.

Currently, the educational policies of most states link standards, curriculum and assessment in an attempt to control all aspects of the educational enterprise. The research findings indicate that use of this hyperrationalized approach adversely affects individual and collective teacher capacity and can constrict or eliminate practitioners' individual and collective attempts at creativity and innovation.

## Contents

Dedication .....	iv
Acknowledgements.....	v
Abstract.....	vii
Chapter One: Introduction .....	1
<i>Assessment Reform in a Public Policy Context</i> .....	3
<i>Assessment Reform: Alternative Approaches to Standardized Testing</i> .....	5
<i>Learning and Assessment</i> .....	6
<i>Teachers' Conceptions of Assessment</i> .....	8
<i>Teacher Capacity</i> .....	8
<i>Research Focus</i> .....	10
Chapter Two: Reforming Assessment: An Examination of the Issues.....	12
<i>Assessment Reform: Rationale, Issues and Concerns</i> .....	15
<i>What Does it Mean to Know? Theories of Cognition</i> .....	20
<i>What does it mean to know? Empiricist epistemology</i> .....	21
<i>What does it mean to know? Constructivist perspectives.</i> .....	23
<i>Standards and Assessment as a Policy Tool</i> .....	28
<i>Teacher Capacity: Definitions and Understandings</i> .....	35
<i>Studies related to teacher capacity.</i> .....	39
<i>Research on teacher beliefs.</i> .....	43
<i>Research Questions</i> .....	46
Chapter Three: Situating the Study: Context and Methodology.....	48
<i>Recent History of Maine Assessment</i> .....	50
<i>The Local Context of the Study</i> .....	56
<i>The site.</i> .....	57
<i>Individual teacher and administrator participants.</i> .....	59
<i>Methodology</i> .....	63
<i>Data collection</i> .....	63
<i>Interviews.</i> .....	63
<i>Observations.</i> .....	65
<i>Data analysis.</i> .....	67
<i>Trustworthiness of the study.</i> .....	68
Chapter Four: Findings.....	69
<i>Teacher Capacity and Classroom Practitioners</i> .....	71
<i>Conceptions of learning and students as learners.</i> .....	72
<i>The capacity for reform: The Learning Results.</i> .....	77
<i>Teacher capacity and assessment.</i> .....	84
<i>District Capacity: The Beaver Pond Assessment Committee</i> .....	93
<i>Chapter 127, Policy, and Effects on Capacity</i> .....	103
<i>The assessment committee and its work.</i> .....	104
<i>Individual participants and the LAS: Issues and concerns.</i> .....	107
<i>Legislative intent and hyperrationalization.</i> .....	114
<i>Summary</i> .....	116

Chapter Five: Discussion .....	119
<i>Hyperrationalization Revisited</i> .....	120
<i>Hyperrationalization as capacity inhibiting: Hanging on while letting go.</i> ....	121
<i>Teacher response: Defensive simplification.</i> .....	123
<i>Cautious implementation.</i> .....	126
<i>Relationship to capacity.</i> .....	129
<i>Usefulness of the Research in Other Contexts</i> .....	135
<i>Limitations of the Study</i> .....	136
<i>Further Research</i> .....	139
References .....	143
Appendix A: List of Interviews, Observations and Meetings.....	159
Appendix B: Sample Interview Questions for Teachers.....	160
Appendix C: Beaver Pond Assessment Committee.....	161

## Chapter One: Introduction

In the popular *Harry Potter* books, first-year students at Hogwarts School for Wizardry and Magic get assigned to their houses by means of a Sorting Hat. Apparently the Hat uses some criteria as it performs its work, but we are not made privy to its rationale. How and why does it assess students for placement in the various houses? How are decisions reached? How useful are they? Does the Hat use certain standards? What aspects of a student's background does it take into account? Do hierarchies of knowledge or experience come into play? Who is holding the Sorting Hat accountable for its decisions? Does the sorting process need to be reformed? Generations of students in public schools have been similarly perplexed by attempts to sort them into categories based on assessments of various types. The Sorting Hat's case-by-case pronouncements of its decisions are an illustration of the obscure assessment processes that can befuddle us all. We are all affected by decisions concerning assessments, interpretation of assessment results, and uses to which assessment data are put. We do not have a Professor Dumbledore with an infallible Sorting Hat whose decisions we trust in matters of assessment.

The term "assessment" is used by policy makers, politicians, administrators, teachers, researchers and members of the general public in diverse and sometimes contradictory ways. Inconsistencies in definitions, assumptions and application among the various groups touting assessment reform have worked to the disadvantage of those most responsible for reform implementation such as teachers and local administrators. This study is designed to uncover the assumptions underlying assessment reform efforts and explore the contradictions and misjudgments that

characterize assessment policy. Its focus is on classroom teachers and local-level administrators as interpreters and enactors of reforms that originated from several sources. The common purpose of these reforms is the use of assessment.

Broadly, at least two groups with conflicting agendas are involved in educational reform that focuses on assessment. One group, policy makers, is interested in assessment as a source for data that will be used to hold schools accountable. This group includes state and federal level policy makers and political leaders. The other group sees assessment reform as a vehicle for acknowledging and advancing conceptions of student knowledge and achievement that have been overlooked within the context of traditional educational assessment (Wiggins, 1989; Shepard, 1989; Stiggins & Bridgeford, 1985). This group includes teachers, policy makers, and researchers who are interested in expanding conceptual understanding of the term assessment with the goal of improving the educational experience for all students.

Both groups see revisions in assessment practices as a vehicle for improved teaching and learning. However, we cannot assume that both groups are working from the same conceptions of improved teaching, learning, and assessment. Underlying the different approaches to assessment are very different assumptions about teaching and learning that can be traced to different learning theories (Delandshere, 2002; Shepard, 2000). The dominant learning paradigm of the 20<sup>th</sup> century, based on behaviorism and rooted in essentialist views of learning, is being challenged by constructivist ideas of learning. Constructivists share the precept that learning is a product of the individual mind and is not imposed from without.

Teachers are not the all-knowing masters who impart knowledge to students, but mediators who assist students as they move from “novice” to an understanding of the established meanings of the wider society in their field(s) of study (Lave & Wenger, 1991).

Untangling the threads of the different discussions about assessment and implicit assumptions about teaching and learning is not easy. Yet teachers are asked to make sense of all the reform rhetoric that comes their way from multiple sources and put it into practice in their classrooms. Teachers’ personal and professional resources will affect the decisions they make about their practice. Interactions among persons in the teacher-colleague-administrator relationship also contribute to the interpretation and implementation of reform.

This introduction presents the context for the study by briefly noting two different policy perspectives on assessment reform. Next, I consider assessment and its relationship to other important aspects of the educational enterprise. Finally, I advocate for the use of teacher capacity as an explanatory lens for examining assessment in a local context.

### *Assessment Reform in a Public Policy Context*

Efforts to reform public education in the United States are on-going. Current reform efforts, with their emphasis on educational standards and standardization, as well as increased attention to assessment, have been gathering steam since the 1983 publication of *A Nation at Risk*. This document presented a bleak picture of economic decline supposedly directly related to the state of education in the United

States. The states responded to this siren call by promulgating more educational laws and regulations from 1983 to 1990 than they had in the previous twenty years (Tyack & Cuban, 1995).

A primary focus of all these laws and regulations has been the promotion of academic excellence through adherence to higher educational standards. The emphasis on higher standards has brought with it a new emphasis on educational accountability with assessment, primarily through multiple-choice standardized testing, as the means by which attainment of standards is determined. The federal government has also entered the picture with legislation such as President Clinton's Goals 2000 and No Child Left Behind, enacted by Congress in 2001 under President Bush (PL 103-227, 1994; ESEA, 2001).

This use of testing as a tool for accountability is unchallenged in most public discussions of education. Assessment from this perspective is a mechanism by which policy makers can exert considerable influence over what happens in classrooms (McDonnell, 1994). Students who perform well on state-mandated assessment instruments are assumed to have mastered the appropriate curriculum and to have attained "higher standards." Assessment in this instance is part of a tight mechanism of goals and control that serves the political purpose of appearing to hold educational entities accountable (Wise, 1979). From the standpoint of educational policy, assessment reform means using educational assessments, in most cases standardized tests, to hold schools, students, and teachers accountable to state-defined higher standards (ESEA, 2001).

### *Assessment Reform: Alternative Approaches to Standardized Testing*

Although testing for educational accountability became the dominant paradigm in a vision of school reform promulgated by lawmakers and people concerned with educational policy in the 1990s, other ideas of assessment reform were taking root at the same time. Researchers began to challenge the narrow view of learning that was represented by multiple choice standardized tests (Wiggins, 1989; Shepard, 1989; Archbald & Newmann, 1988). Standardized tests were criticized for inadequately measuring important learning outcomes such as students' abilities in problem solving, "real life" situations, and sustained intellectual endeavors (Wiggins, 1989). The interest of these reformers was in broadening the ways we look at student knowledge and performance.

Initially the term "alternative assessment" referred to assessments other than standardized multiple choice tests. Other more descriptive phrases were coined, such as "performance assessment" and "authentic assessment." "Performance assessment aims to model the real learning activities that we wish pupils to engage with so that assessment does not distort instruction" (Gipps, 1994, p. 11). Some researchers preferred the use of the term authentic assessment, which is performance assessment carried out in an authentic context (Gipps, 1994). That term was criticized by others for introducing value-laden vocabulary into an already heated discussion (Terwilliger, 1997; Madaus & O'Dwyer, 1991). The terms are interchangeable to the extent that they all refer to methods of assessment other than multiple choice tests (Wiggins, 1998).

As the research on performance assessment expanded, advocates began to articulate a vision of assessment that would shape teaching and learning through its effective daily use in the classroom, not just its use as an indicator of student achievement (Shepard, 1989; Wiggins, 1989). The most important locus of assessment reform for these advocates of change is the classroom. Assessment should be integrated with instruction, congruent with important learning goals and anchored in meaningful tasks. Students should engage in self-assessment, and students and teachers should be collaborators in the learning process, discussing together aspects of an individual's progress in learning such as learning targets and next steps (Wiggins, 1998; Wolf, 1993; Davies, Cameron, Politano & Gregory, 1992). The underlying idea was that educational evaluation would make a transition from a testing culture to an assessment culture (Wolf, Bixby, Glenn, & Gardner, 1991).

Thus we have two views of the purposes of assessment: One sees assessment as an instrument of policy, with an emphasis on its use as a mechanism of control and accountability; the other focuses on the pedagogical potential of assessment to improve teaching and learning as we broaden ways to look at student achievement and performance that go beyond the narrow constraints of multiple choice testing.

### *Learning and Assessment*

Central to the issue of assessment is its connection to learning. In the traditional model with which all students are familiar, material to be learned is presented, then students practice the material, then they are assessed. Assessment has

been seen as something apart from teaching and learning, separate in time and purpose (Shepard, 2000). Advocates for alternative assessment propose instead that assessment be viewed as an occasion for learning and hence part of the instructional cycle (Wiggins, 1998; Wolf, 1993, Shepard, 2000).

Unacknowledged in the discussion of assessment reform are issues about competing perspectives of learning (Delandshere, 2002; Hargreaves, Earl & Schmidt, 2002). Most state standards and state assessments are firmly rooted in an essentialist view of learning that assumes that requisite knowledge is a body of material to be learned, and that the learner can then be assessed on the knowledge he has acquired. If we look at other perspectives on learning that are currently popular, such as constructivism, we find a variety of viewpoints and differing interpretations about how humans acquire knowledge (Slife & Williams, 1995; Delandshere, 2002). Investigations of classroom practices, including assessment, might note significant differences of opinion within constructivism, but these differences are glossed over (Shepard, 2001; Gipps, 1999; Windschitl, 2002). Certain stock phrases, such as “teaching for understanding,” (Windschitl, 2002, p.131) are bandied about, but questions such as “What does it mean to understand?” or “What does it mean to know?” are never asked, much less discussed (Delandshere, 2002). Thus assessment reform, however it is envisioned, is expected to occur amid unchallenged assumptions about learning and unacknowledged contradictions as to purpose. This is not helpful for teachers and others who must make responsible decisions about students and their learning based on assessment.

### *Teachers' Conceptions of Assessment*

Assessment of students has been identified as one of the six core job functions of teaching (Schafer, 1991). Yet teachers rate their training in assessment techniques to be inadequate, and give a low rating to their interest in and enjoyment of assessment (Plake & Impara & Fager, 1993; Stiggins & Bridgeford, 1985). According to one study, less than thirty percent of teacher education programs require any coursework that is focused solely on assessment issues (Schafer, 1991). Teachers learn assessment in an idiosyncratic manner and apply it in the same fashion (Stiggins & Conklin, 1992).

Teachers' critical attitudes towards standardized testing have been documented by research (Smith, 1991; McNeil, 2000). Research about teachers' conceptions of alternative assessments and other aspects of assessment reform is not as abundant. Certainly the ambiguities involved in different conceptions of assessment reform have an impact on how teachers understand and implement new forms of assessment. Some evidence exists that the lack of consensus about reform terminology and purpose is detrimental to implementation efforts (Snow-Renner, 1998; Matthews, 1995; Delandshere & Jones, 1999).

### *Teacher Capacity*

Implementation of educational reform never occurs precisely as policy makers envision (Tyack & Cuban, 1995; Spillane & Jennings, 1997; Snow-Renner, 1998; Datnow & Castellano, 2000). The grand vision that typically informs the legislation at the capitol building must be translated into practice at the local level. A key to

reform implementation, then, is the local actors' ability to serve as interpreters, and not just implementers, of policy (Datnow & Castellano, 2000). In particular, teachers' capacity to understand, interpret, and implement the many changes involved in a reform effort will affect the outcome. Important aspects of teacher capacity include dispositions, beliefs and motivation. Teachers' initial understanding of the concepts underlying a reform policy as well as their grasp of the desired outcomes should be a consideration for policy implementation.

Issues of teacher agency in reform implementation are frequently left unaddressed (Hargreaves, 1997; Smyth, 1995). Policy makers often proceed as if their proposals and legislation will be enacted in classrooms exactly as envisioned. Hence reform implementation is often seen operating in one direction only, from the top down (Conley & Goldman, 1998). For many reasons, teachers are seen as cautious rather than enthusiastic implementers of reforms conceived by policy makers (Wilson & Floden, 2001). What might be seen initially as recalcitrance on the part of teachers to adopt reform may instead be an issue of capacity as teachers attempt to retain the ability to exercise their professional judgment in ways that are congruent with their beliefs and understandings.

Individual capacity is a vital component of reform implementation, but use of capacity as an explanatory tool will be extended to demonstrate that collaboration among teachers and administrators as they set about the task of interpreting policy is an important element of school-level capacity. Tangible and intangible aspects of district-level capacity affect reform implementation. Tangible attributes include resources such as money, time, astute leaders, and personnel with capacity to

implement reform. Intangible aspects include the collaborative environment within the system, coherence of the proposed reform, and a critical mass of individuals who are capable of implementing the reform (Fullan, 2001; Floden, Goertz & O'Day, 1995).

### *Research Focus*

This research examines how teachers interpret and enact assessment reform. Qualitative research methods such as observation and interviewing are used to gain insight into participants' understanding of the various aspects of assessment reform, its ambiguities and contradictions. The broad questions relate to local enactors' capacity to understand and implement assessment reform individually and collectively while untangling the contradictions in purpose that may be part of the stated reforms. What conceptions of teaching and learning influence teachers' capacity in assessment? What contributes to the development of teachers' understanding and use of assessment, particularly when official expectations are unclear or contradictory? How do teachers' individual and collective capacities influence attempts to understand and implement policy? How does individual capacity intertwine with communities of practice (Lieberman, 1996) to support teachers' collective work in assessment?

There are several tasks for this study. First we must look at assessment policy from differing perspectives, and examine the assumptions about knowledge and learning that underlie each perspective. The ambiguities, contradictions, and unchallenged assumptions of current assessment reform efforts extend beyond the

theoretical and become especially problematic as teachers and administrators undertake the task of turning policy into practice. Capacity will be defined and used as the lens for examining local enactors' attempts to grapple with competing and contradictory versions of assessment reform. Finally, assessment reform in a particular context will be examined and the data used to illuminate teachers' individual and collective reactions to assessment reform.

## Chapter Two: Reforming Assessment: An Examination of the Issues

The purpose of this chapter is to examine the salient research on assessment reform. The issues to be considered include: different perspectives on assessment and its reform, underlying assumptions about teaching and learning, and teacher capacity and its role in reform implementation. An investigation of the pertinent research on these issues and their connectedness will form the basis for the questions that inform this study.

Advocates of the use of assessment as a means to educational reform have called for both increased use of multiple choice tests (ESEA, 2001) and less emphasis on testing (Gipps, 1994). Critics of multiple choice testing maintain that we cannot evaluate whether students have reached high standards if we do not evaluate their capabilities in other ways, as these tests assess only a small amount of student knowledge (Wiggins, 1998; Gipps, 1994). Some educators have turned to what is commonly referred to as performance assessment as an alternative to multiple choice testing. They maintain that performance assessment could potentially tap into student knowledge in ways that were not likely to happen with multiple choice tests (Wiggins, 1998; Shepard, 1989; Gipps, 1994).

I shall begin this chapter with a brief overview of the emergence and development of performance assessment, including aspirations for its potential to change classroom practice and improve learning. Advocates of performance assessment saw the use of this approach as an antidote to the perceived narrowness of traditional instruments of educational assessment such as standardized tests (Madaus

& O'Dwyer, 1991). However, increased use of performance assessment led to growing concerns regarding its suitability in some assessment contexts.

The research of the early 1980s on assessment reform did not examine basic assumptions about learning that underlie the educational enterprise. In the second section I address some of these assumptions by examining the perspectives on teaching and learning and their variations that continue to dominate educational discourse. This has implications for both classroom practice and educational policy.

In the third section of this chapter, I examine assessment reform as an educational policy tool (McDonnell, 1994). Standards and alternative assessment became prominent in the mid 1980s as aspects of educational reform (Ravitch, 1995; Wiggins, 1989). The adoption of standards became a centerpiece of educational policy in almost every state (Fullan, 2001). The increased use of assessment as an accountability tool tightened the bonds between assessment and state standards, and assessment became a critical element of standards-based reform. These strong links between standards and assessment led to a hyper-rationalized system; that is, the outcomes of such a system are contrary to what the designers intended (Wise, 1979). Even though their approaches to assessment reform are quite different, both advocates of standards-based reform and performance assessment share concerns about public education's ability to provide an appropriate education for all students.

The chapter concludes with an examination of teachers' conceptions of assessment and its function as an aspect of their work. This discussion clarifies the need for a framework that makes teachers and their role in reform implementation visible. Teacher capacity will provide the framework that will enable us to

understand teachers' use of assessment in the face of unclear and contradictory expectations that stem from a variety of sources. The research questions that will inform this study will then be seen as a synthesis of the previous discussions of the issues.

It is important to note that the advocates of various approaches to educational reform often share terminology, but not necessarily meaning. This contributes to the confusion surrounding assessment reform efforts. As previously mentioned, efforts labeled "assessment reform" have their roots in differing conceptions of "assessment," "reform," "learning," and "standards," to name just a few of the contested terms. The locus of reform is also part of the debate, with some groups advocating for top down reform while others argue that true reform will not occur if local enactors have no sense of ownership.

Before I proceed, I would like to clarify some key definitions. I will use the term *assessment* to mean "the gathering of information to aid in decision making" (Davies et al., 1992). Inherent in the assessment process is the "formation of value judgments and interpretations that determine the significance, the importance, and the value of learning and knowing" (Delandshere, 2001, p. 117). Testing is the traditional way of assessing and evaluating students in school. It is designed to yield a score or measurement, whereas assessment does not demand quantification. Indeed, part of the early rationale for the use of assessments other than standardized tests was the argument that such assessments would provide more evidence of student performance and hence judgments stemming from assessment would be more defensible. Testing

is only a small part of the methodology available for assessment, but it continues to be the dominant form of educational assessment (Airasian, 1991).

For teachers, a whole range of activities, formal and informal, fall within the category of assessment. The decisions that stem from this information gathering may be informal and used primarily to feed back into the teaching-learning process, or they may be formal in the sense that the information gathered is used to make an evaluation regarding the quality of a student's work and its sufficiency in meeting requirements for, e.g., promotion or graduation. The traditional way of evaluating students is to give them grades. Grading, for better or worse, is usually an important outcome of the assessment process, but assessment encompasses much more than assigning a grade. Indeed, a letter grade obscures the complexities of the assessment process because of the limited amount of information it conveys and the assumptions of objectivity that have traditionally accompanied letter grade use. Teachers' assessments of cognitive growth, affective growth, and improvement in skills that may have occurred over a period of time get reduced to a letter on a report card that is assumed to have equal meaning across contexts but in reality does not.

#### *Assessment Reform: Rationale, Issues and Concerns*

The terms "alternative assessment," "authentic assessment," and "performance assessment" are used more or less interchangeably in spite of slight differences in emphasis (Gipps, 1994; Terwilliger, 1997). What has come to be known as authentic or performance assessment started receiving increased attention in the early 1980s (Stiggins & Bridgeford, 1985). Measurement experts and others

became concerned that the emphasis placed on multiple choice standardized tests limited the ability to measure students' competency and knowledge. Multiple choice tests may not assess important learning outcomes because their format constrains the content that can be included, nuances of interpretation are not honored, and emphasis is placed on certain aspects of psychometric theory, such as reliability and a narrow view of validity, at the expense of covering a wider range of instructional objectives (Shepard, 1989).

In light of these growing concerns, performance assessment was seen as an antidote to the narrowness of multiple choice standardized tests. By engaging in activities that more closely replicated "real world" activities, learners would be able to demonstrate a wider range of knowledge and competencies. Performance assessment would encourage students to think, engage in real-world problem solving through access to an interdisciplinary curriculum, and then demonstrate their ability to use that knowledge in situations that replicated real world contexts. Performance assessments were seen as the answer to the narrowing of the curriculum that was part of the problem with selected response standardized testing; it was said that with performance assessment, teachers could develop a "thinking curriculum" with tests worth teaching to (Resnick & Resnick, 1992, p. 37).

The supporters of performance assessment began advocating for change at the state level. The appeal of a more challenging curriculum for all students was hard to resist. Some states were eager to reform their assessment policies in order to answer the legitimate criticisms of selected response standardized testing. California, Connecticut, Washington, and Arizona, to name a few, began work on programs

designed to implement performance assessment state-wide (Shepard, 1989). States did not want to be accused of lagging behind in the teaching and assessing of a richer curriculum. The message that “what you test is what you get” was starting to make an impression on state legislators and administrators as well as educators (Shepard, 1989).

Much of the early enthusiasm about performance assessment concerned its use as a summative form of assessment, that is, as a way of certifying a level of achievement or performance at the end of a specified period or course of study. This has been referred to as “assessment *of* learning” (Assessment Training Institute Foundation, 2002). This is a typical function of both classroom and standardized tests. In test theory developed by the measurement community over the course of the last century, validity, reliability, and generalizability were important attributes of test scores that determined the uses to which the scores could be put (Messick, 1989; Linn, 1995). Using a test score to determine if a student could be promoted to the next grade, for example, requires more evidence as to validity, reliability, and generalizability than using a test score to move a student from one task to another in the classroom.

Concerns soon arose, however, about the wisdom of using performance assessment in the same manner and for the same purposes that selected response tests were used. Evidence to support validity and reliability in the traditional sense was difficult if not impossible to produce (Gipps, 1994; Madaus & O’Dwyer, 1991; Terwilliger, 1997; Koretz, McCaffrey, Klein, Bell & Stecher, 1996). Some measurement experts acknowledged that psychometric notions of validity and

reliability may not retain their traditional relevance, but that other technical standards related to reliability such as standard error of measurement were still very important in determining the dependability of decisions made about students' placements (Linn, 1995). In addition to these concerns, performance assessments on a large scale were expensive, time-consuming, more difficult to grade, and more inefficient to administer (Madaus & O'Dwyer, 1991).

Another caveat concerning performance assessments also surfaced. Initially, performance assessments were viewed as being more equitable than traditional standardized tests because students had an opportunity to demonstrate a wider range of knowledge and achievement. However, as Linda Darling-Hammond pointed out (1994), the mere use of performance assessments would do little to promote educational equity if the purpose of assessment did not change from sorting mechanism to diagnostic support. "A fundamental question is whether assessment systems will support better teaching and transform schooling for traditionally underserved students or whether they will merely reify existing inequities" (p. 7).

As policy makers, teachers, administrators, educational researchers and others wrestled with the dilemmas inherent in large-scale performance assessment, some advocates for assessment reform proposed that the locus of reform be shifted from large-scale, state-mandated efforts to classroom assessment (Wolf & Reardon, 1996; Gipps, 1994; Shepard, 1989; Wiggins, 1989). Classroom assessment can either be summative, that is, certifying a level of achievement at the completion of a specified unit of study, or it can be formative, that is, used in a manner that informs and guides further instruction. In contrast to assessments that are required by others and that

occur outside the context of classroom instruction, classroom assessments are more likely to be the domain of the classroom teacher.

By giving formative classroom assessment a more prominent role in the assessment process, advocates of performance assessment could concentrate on the positive potential of this approach to assessment while working to eliminate some of the most intractable problems that dogged its use on a large scale. In the words of Stiggins in an ATI video (2002), teachers could concentrate on assessment *for* learning as opposed to assessment *of* learning. In the classroom, assessments could operate as diagnostic supports rather than sorting mechanisms (Darling-Hammond, 1994), and students could be involved in self-assessment, leading to more involvement in their schooling, a sense of ownership, and heightened metacognitive skills. Assessment would no longer only be something that was attached to the end of a unit of study as a means of evaluation; assessment would be an integral part of instruction, with responsibility shared more equally by teacher and student (Gipps, 1999; Shepard, 2000; Wiggins, 1998). It could function as the feedback loop that allows for the identification of the extent to which learning has taken place and lays the groundwork for the next stages of learning (Hargreaves, Earl & Schmidt, 2002).

Advocates of performance assessment were instrumental in helping to uncover some of the deleterious effects of standardized testing, but underlying assumptions about what it means to learn were not challenged (Delandshere, 2002). This is critical because teachers' assessment practices and beliefs will be linked to their beliefs about what is important for students to know and how students will demonstrate mastery of important knowledge. The prevailing discourse in education

assumes that knowledge is acquired empirically (Slife and Williams, 1995).

Changing assessment practices without a concurrent discussion of assumptions about teaching and learning will not reform educational practices in desired ways.

### *What Does it Mean to Know? Theories of Cognition*

Initial critiques of standardized testing did not focus so much on the psychology of learning as on the format of the assessment. People advocated for more challenging content in schools, for increased use of higher order thinking skills, for assessments that engaged students in higher level activities instead of merely testing them (Wiggins, 1989). The enthusiasts for assessment reform in the 1980s and 1990s concentrated on the evils of standardized testing rather than challenging implicit assumptions of teaching and learning.

However, we cannot fundamentally change the educational process without asking some basic questions about the nature and purpose of schooling. Standardized tests ask, “What does this student know about the constructs tested in comparison to others?” Performance assessments ask, “What does this student know within the constraints of the required task?” The basic question, “What does it mean to know?” is never investigated (Delandshere, 2002). We will now turn to this question and its ramifications for the teacher’s role in the classroom.

From the practitioner perspective, learning theories have been divided into two camps: empiricism and constructivism. I acknowledge that this division is somewhat simplistic but utilize it here because it captures the essence of current practice (see e.g., Shepard, 2000; Windschitl, 2002).

*What does it mean to know? Empiricist epistemology.*

*Epistemology* is a term found in philosophy that refers to ways of knowing, or the nature of knowledge. Slife & Williams (1995) maintain that the vast majority of behavioral scientists, and they include educators in that group, endorse empiricism as their epistemology of choice, whether knowingly or not. “Empiricism is the notion that our learning and memory are primarily derived from our experience of events of the world” (Ibid, p. 67). The mind is involved in the learning process, but the source of knowledge is information gained through sensory experience. This view of learning is so ingrained and has such a long history that we have difficulty envisioning other pathways to learning. Understanding the knowledge claims of empiricism is a necessary prelude to examining the educational status quo.

Theories of learning that are empirically based make a number of knowledge claims. Knowledge is out there, waiting to be discovered or uncovered, and is independent of the knower and ideologically neutral (Cherryholmes, 1988). Facts are discovered (science) or known (history), and issues of interpretation of this avowedly non-ideological knowledge are not raised (Gipps, 1999). Knowledge is considered to be transhistorical, replicable, and disinterested (Gipps, 1999). Learning theories based on this view of knowledge have been criticized for assuming the decomposability and decontextualization of important learning objectives (Resnick & Resnick, 1992). This means that all important learning objectives can be specified in discrete pieces in a hierarchical sequence that can then be taught, practiced, and tested (Shepard, 1989).

Empiricism has dominated educational discourse and life in classrooms through behaviorism as an approach to learning. Behaviorism defines learning as a more or less permanent change in behavior over a period of time (Driscoll, 1994; Shepard, 1989). Desired behaviors are elicited by manipulating environmental variables. All learners, it is assumed, will respond once the correct combination of variables is found.

The teacher's role within this deterministic approach to learning is familiar. Teachers present material with learning objectives in mind, reinforce it through exercises, homework, or other assignments, and then assess what has been taught. The curriculum is fixed, and teachers and others determine what knowledge is of the most worth. Knowledge is doled out in pieces and then assessed, usually through a test. Students who do not demonstrate specific mastery of the carefully sequenced material are not permitted to move on. Students are treated as empty vessels, ready to be filled up by the teacher, who possesses the knowledge that is necessary for students to become educated (Berlak, 1992; Gipps, 1994).

Assessment within this framework has traditionally been summative and in the form of tests—end of chapter, end of book, end of grade. The use of tests as valid indicators of learning is an offshoot of the assumptions underlying the empiricist perspective. That is, a mechanistic approach to teaching and learning leads to assessment practices whose purpose is to measure the ability of students to reproduce knowledge when asked to do so. In such a framework, tests take on a great deal of significance as the means through which students will demonstrate their knowledge. The curriculum narrows as teachers concentrate on the knowledge that is deemed

important, i.e., that information to be tested (Smith, 1991; Shepard, 1989; McNeil, 2000).

*What does it mean to know? Constructivist perspectives.*

A focus on constructivism as a learning theory has been parallel to the focus on alternative forms of assessment in both educational research literature and literature for practitioners (Shepard, 2000; Perkins, 1999; Brooks & Brooks, 1993). Constructivist beliefs can be placed on a continuum, from radical psychological constructivists to radical social constructivists (Delandshere, 2002). In part, disagreements concern the nature of learning: Is learning active cognitive reorganization, undertaken by the individual? Or is learning a process of enculturation into a community of practice and thus a social activity? Cobb (1996) has suggested that perhaps the perspectives shift, with one sometimes in the foreground, then the other.

Theorists have more nuanced interpretations of the various positions along the continuum, but what is important for our purposes here are the conceptions and articulations of practitioners. Most literature relevant to educators discusses constructivism as if there were two categories: psychological (sometimes referred to as cognitive) or social/cultural (Windschitl 2002; Cobb, 1996). It is important to note that differences in epistemological assumptions underlying constructivist learning theories are seldom articulated in the literature, even though the implications for practice are considerable (Windschitl, 2002; Delandshere, 2002). I will address the ramifications for teaching, learning, and assessment from each perspective.

In psychological constructivism, the phrase, “Students construct their own knowledge” refers to the Piagetian view of learning as development that occurs through an individual’s assimilation and accommodation to new experiences occurring in his environment, i.e. active cognitive reorganization (Ginsburg & Opper, 1988). Learners construct their own interpretations of concepts and relate them to their existing knowledge. This existing knowledge may be unsophisticated or inaccurate, which can influence the development of concepts and interpretations regarded as acceptable by experts in the field (Windschitl, 2002). It is important to note that psychological constructivists analyze learning within the individual, giving priority to pupils’ sensory-motor and conceptual activity (Gipps, 1999).

This perspective defines learning with a complexity lacking in behaviorism, but it is still located outside the individual and is controlled by the environment. The assumption is that the environment can then be structured to produce the desired outcomes or responses (Delandshere, 2002). Teachers and students are not engaged in co-construction of knowledge so much as the teacher serves as a coach/mediator/guide to the acquisition of “correct” knowledge. Knowledge is still a thing to be possessed rather than an activity to be engaged in (Delandshere, 2002).

How is this a constructivist approach? Here we are working from a practitioner’s perspective. This is considered a constructivist approach because, in contrast to behaviorism, learners are not seen as empty vessels waiting to be filled, but active organisms seeking meaning (Driscoll, 1994; Brooks & Brooks, 1993). Clearly this is a difference of degree of imposed knowledge, rather than kind, but one that nevertheless has implications for teachers.

A teacher who works from the perspective of psychological constructivism helps students attain disciplinary concepts that are more congruent with what has been validated by experts in the field. Teachers working from a psychological constructivist perspective are encouraged to elicit students' prior knowledge of topics and concepts in order to plan instruction that will lead to deep learning (Windschitl, 2002; Brooks & Brooks, 1993; Shepard, 2000). Teachers are also encouraged to have, as part of their instructional strategy, "clearly defined *conceptual goals* for learners and an understanding of how learners might progress toward these" (Windschitl, 2002, p. 140) (emphasis mine).

Thus students may take somewhat different learning paths, but educators still speak of "learning targets" that they are to "hit" and assessments that will determine how close they are (Ramirez, 1999; Hargreaves et al., 2002). Even if assessments are performance assessments that might tap into broader domains of learning than standardized tests, the underlying assumption that learning is controlled by the environment remains (Delandshere, 2002). While this approach to assessment is more sophisticated than the traditional psychometric paradigm, underlying theories of learning are much the same.

Both psychological constructivists and sociocultural constructivists assume human agency in the process of coming to know, but sociocultural constructionists do not view knowledge construction as the result solely of individual cognitive processes. Rather, knowledge construction depends on engagement in relational activities with others (Gipps, 1999). Models of learning used by sociocultural constructivists include mentors, apprenticeships, and scaffolding (Gipps, 1999).

“Learning is a process that takes place in a participation framework, not in an individual mind. It is distributed across co-participants,” according to Lave & Wenger (1991, p.15).

This view of learning holds very different implications for the role of the teacher. The model of schooling with which we are most familiar is, indeed, about coming to know the world, but in many instances, the world that we are to know has already been decided upon. If we are working from a sociocultural perspective, teaching practice would need to decenter learning from the current teacher-student dyad to focus on constructing learning communities with co-participation of both teacher and students. The role of the teacher is to mediate between students’ personal meanings and the established meanings of the wider disciplinary community of which they are a part (Cobb, 1996). Learning is a way of being in the social world, not a way of coming to know the world (Lave & Wenger, 1991).

In such a vision of education, the role of the teacher extends beyond knowing certain basic information and having the pedagogical skills to transmit that information. Classroom communities should be designed where students engage collaboratively in relevant, meaningful intellectual activity (Wells, 1999; Brooks & Brooks, 1993). Teachers must become more than “didactic caretakers,” and must shift from acting upon students as compliant learners to co-participating with students as makers of meaning (Lave & Wenger, 1991; Wolf & Gearheart, 1997). Teachers must also have a voice in creating learning communities that do more than reproduce the most basic sorts of knowledge (Bull, 2000).

Delandshere (2002) questions whether our current understandings of the word

‘assessment’ allow us to address the question, “What do students know?” from a sociocultural perspective. “Because truth cannot be established with any certainty, there are no standards of right and wrong, and therefore no criteria by which to make individual judgments” (p. 1474). She later states,

“If what we know cannot be separated from how we know and from the experiences and activities that shape it, then the assessment questions have to be framed in such a way as to be consistent with this theoretical perspective and include social, cultural, and ethical issues that have not typically been addressed....if learning is a “kind of doing” (Gill, 1993, p. 68) or the transformation and change of people (Lave & Wenger, 1991), then in documenting learning we have to take its dynamic nature into account. In this perspective, because individuals’ experiences will always be different, the current focus on the assessment of individual differences in learning outcomes at a given moment and regardless of the context and experiences of these individuals becomes irrelevant and meaningless” (p. 1479).

Both Delandshere’s (2002) use of the inquiry metaphor and Wolf’s designation of “assessment as an episode of learning” (1993) help us envision an approach to learning that is more generative, in contrast to the current mimetic model (Delandshere, 2002). Assessment in this context should be more interpretive and less concerned with high-quality technical matters that narrow the definition of learning. Perhaps it could truly be a demonstration of student understanding of a topic in such a way that would provide evidence as to what a student knows, the breadth and depth of that knowledge, and how closely that knowledge approximates the taken-as-shared knowledge of more competent members of the community of practice, assuming some consensus in this area. Students should be involved in the assessment and in determining how the products of such an assessment might be compiled and used. Such collaboration is important since teachers cannot “know” their students and need students’ insights in order to begin to understand what a student might “know”

(Hargreaves et al., 2002). Similar to Lave and Wenger's communities of practice (1991), Hargreaves et al. (ibid) propose that students engage in ongoing dialogue within various communities that would clarify their achievements and inform their learning.

The question, "What does it mean to know?" has no simple answer. Schools have traditionally operated as if "knowing" was the same as scoring well on a test. This view has been challenged by many in the education community, including advocates of assessment reform, but without sufficient examination of the underlying assumptions regarding the nature of knowledge and what it means to learn, fundamental change will not occur. As a further complication, the issues and challenges surrounding how important knowledge gets determined and assessed are defined differently from the perspective of policymakers.

#### *Standards and Assessment as a Policy Tool*

Concern about the academic abilities of U.S. students extends back to the 1960s when scores on the SAT, a test taken by college-bound students, started to decline (Ravitch, 1995). However, most experts date the current reform movement from the 1983 report, *A Nation at Risk*. The composition of the commission and the strong language of the report guaranteed that this report and its recommendations would be widely disseminated. Most of the recommendations involved testing, and one could say that testing became the engine driving educational policy (Linn, 1995).

Education in the United States operates on an assumption of local control, thus hampering efforts to obtain nation-wide consensus on issues of educational

policy. However, the existence of a cabinet level Department of Education created a venue for entertaining the discussion about improving the quality of education in the U.S. In 1989, President G.H. Bush and the state governors drafted a list of goals referred to as America 2000 that implied that all children would master challenging content in schools. President Clinton's version of this proposal was called Goals 2000. While neither of these proposals called for federal standards, the need for some sort of standards was implied. It was hoped that the coalescence of concerned business people, members of the education community, parents, and policy makers would forge a consensus about standards that would be the impetus for educational improvement.

Neither G.H. Bush nor Clinton proposed national standards. The opposition to federal control of education was too great. However, professional fields, following the example of the National Council of Teachers of Mathematics and with assistance from the federal Department of Education, were encouraged to shape a consensus about what students should know and be able to do in their respective disciplines. Thus discipline-based groups in science, history, geography, foreign languages, the arts, English, and civics drafted standards that were voluntary. The hope was that these standards would eventually prove useful, gain acceptance among educators and the public, and find their way into the schools (Ravitch, 1995).

This idealistic concept of a voluntary standards movement that would gain acceptance at the grass roots level went awry as policy makers began to draft requirements for state standards and tie them to accountability measures. "A survey conducted at the end of 1992 found that most states had willingly joined the national

push toward more demanding discipline-based content standards, defined in terms of comprehensive curriculum frameworks and innovative assessment systems” (Ravitch, 1995, p. 160). All major instructional mechanisms, such as curriculum, instruction, and assessment were aligned with these challenging standards (Fuhrman, 2001; Goertz, 2001). The issue became, not one of standards, but of standardization (McNeil, 2000). The assertion was that challenging standards would ensure that all children in America would develop the skills that were once expected of our top students.

Tracking down the origins of such platitudes as “high standards for all,” “all children can learn to high levels,” and “no child shall be left behind” is difficult. These phrases may have their origins in two different critiques of the educational system: sociological and psychological. Wolf & Reardon (1996) decry what they see as the historic split in American education between educating for academic excellence and educating for social efficiency, i.e. training for blue collar jobs. In their sociological analysis, they maintain that standards and assessment will not be enough without “a conviction that high achievement is widely attainable” (p. 6). Marian Wright Edelman, chair of the Children’s Defense League, warned against “the soft bigotry of low expectations,” as she advocated for more challenging curriculum and better schools for African-American children, a traditionally underserved part of the school population. The standards that are at issue in this sociological critique have to do with opportunity to learn and access to more challenging curriculum, not more stringent content standards.

Similar phrases can be found in the literature referring to cognition. In her discussion on the epistemological underpinnings that underlie constructivism, Gipps (1999) states, “As the observer constructs reality, the learner constructs his or her own knowledge. In addition, constructivist theory makes the implicit assumption that all students can and will learn....” (p. 372). However, a standard curriculum with the same high content and performance standards for each and every child is clearly not the vision of experts working from a sociocultural perspective:

If we can stop focusing on who learns more or less of particular culturally well-defined fragments of knowledge, and ask questions instead about what is around to be learned, in what circumstances, and to what end, learning achievements would become statements about the points of contact available to persons in various social settings. (McDermott, 1996, p. 277).

Policy makers, however, have seized on phrases used to describe complex ideas of learning without fully understanding the ideas behind them and turned them into “platitudinous statements....[that] inevitably generate enthusiastic response from unreflective audiences” (Noddings, 1992, p.29). Critics of the mantra, “high standards for all” and similar statements are somewhat muffled in the public arena; it is difficult to object to the idea that all our students should receive a good education or that all children can learn to high levels. The ideas behind those slogans were never intended to result in the increasingly prescriptive standards and accountability mechanisms that are in force at the present time. “To say that there are no unteachable children is trivial until one fills out the statement with what is to be taught them and at what cost” (Noddings, *ibid.*). The use of the term “standards” has shifted from delineating important discipline-based knowledge to the concept that they are mandated knowledge-for-all with an accountability framework included.

In this hyperrational (Wise, 1979) system of educational accountability, with goals, standards, and assessments all determined at the state level, good teaching is linked to how well students perform on the assessments that comprise the state system. Implicit in this attitude toward teaching is that learning is controlled by the environment, and the teacher is responsible for structuring the learning environment so that all students achieve the desired results. Teaching and learning are correlated with test scores, and it is assumed that teachers will change their practices to bring about the desired outcomes once all external controls are in place, including rewards and sanctions (Hargreaves, et al., 2002). Teachers become deprofessionalized, with their capacities reduced to a knowledge base and pedagogical skills that lead to “predetermined results through predetermined needs” (Bull, 2000, p. 117). The situated contexts of both teachers and students are ignored. Students become objects to be acted upon rather than people with complex and differing backgrounds, abilities, motivations, and habits of mind. In extreme cases, outside sources may mandate a very structured curriculum, and teachers’ abilities to use professional judgment may be seriously curtailed (McNeil, 2000). Ironically, many of the effects of these systems of educational accountability are not what policymakers desired nor are they beneficial for students (Wise, 1979; McNeil, 2000; McDonnell, 1994).

The behaviorist view of learning as something that is doled out and mastered in discrete bits, then assessed, is a hallmark of assessment as a policy tool. “Assessment and instruction are often conceived as curiously separate in both time and purpose” (Shepard, 2000, p.4). Performing well on the tests, or, in current language, “achieving the [state standards],” is equated with learning, and the

knowledge that is deemed important is not locally negotiated (Chapter 127, 2001). Educational policy experts explain the appeal of testing as an accountability mechanism by offering the following reasons: tests are relatively inexpensive; they can be externally mandated; results of testing/assessment programs are visible; and changes in assessment can affect what happens in classrooms. (Linn, 1995; McDonnell, 1994). The implicit question is not, “What has this student learned?” but “Has this student achieved the standard?”

However, in the rush to accountability through the use of state standards, not all policymakers ignored the criticisms leveled at multiple-choice standardized testing. Many states, particularly in the early 1990s, heeded the concerns related to multiple choice standardized testing as the sole assessment for accountability purposes, and implemented a variety of performance assessments. These included portfolios, assessments of written work, constructed response items, and scientific experiments. Since one of the complaints about accountability testing was that teachers then “taught to the test,” i.e. on a low cognitive level, those who favored higher standards and assessment reform from an academic standpoint proposed a new slogan, “Build assessments towards which you want educators to teach” (Resnick & Resnick, 1992, p. 59).

Quite soon, however, testing and measurement experts were raising concerns about claims that could be made about large-scale performance testing (McDonnell, 1994). These concerns range from traditional measurement concerns about reliability and validity (Linn, 1995; Messick, 1994; Madaus & O’Dwyer, 1991) to concerns about the quality of such assessments (Linn, 2000; Hargreaves et al., 2002). Linn

(2000) cautions that “assessment systems that are useful monitors lose much of their dependability and credibility for the purpose when high stakes are attached to them” (p. 14). The cost of constructing and scoring large-scale performance assessments became an issue. In some states, such as California, the curriculum framework that provided the rationale for performance assessments was challenged and discarded, and both curriculum and assessment returned to a skills-based approach with an emphasis on multiple-choice standardized testing (Lewis, 1999). An examination of state by state information published by Consortium for Policy Research in Education indicates that states are, by and large, moving away from performance assessments on state-level exams and reinstating norm-referenced, multiple choice standardized tests. Assessment of writing competency is the one remaining performance assessment in consistent use on the state level (Goertz & Duffy, 2001).

The tangled strands of educational reform through the use of assessment and the unexamined layers of assumptions about teaching and learning that inform the various approaches to reform ensure a lack of clarity and guarantee conflicting agendas in assessment reform. In many cases, people are using the same words to talk about very different concepts. In policy talk, for example, the term “assessment” is frequently used as a euphemism for testing, especially when scores are needed to fulfill policy makers’ accountability needs. Policy makers expect the same assessment system to provide valid inferences for a variety of very different purposes, with contradictions being written into the policy itself (Coldarci, Johnson, Beaudry, Cormier, Ervin, Rosenblum & Silvernail, 2000). The allure of a tightly controlled, rational accountability system where goals, standards, and assessments are linked in a

neat package has been hard to resist for those who believe that education will improve only with external sanctions and controls.

The responsibility for the implementation of legislated reforms rests with local enactors. In spite of their centrality to reform efforts, much of the research on school improvement is curiously “teacher blind” (Smyth, 1995); that is, it fails to take account of the complexities and realities of teachers’ work. Teachers’ beliefs and concerns about themselves as agents of change and their role in reform implementation are not well documented. Teachers do not have the luxury of deciding what ought to be, in a policy sense; they are charged with taking reforms and making sense of them the best way they know how, contradictions and inconsistencies notwithstanding. In the words of Hargreaves et al. (2002), “These contradictory forces have made assessment reform a schizophrenic activity. It is hard to expect teachers to harmonize their assessment practices when policymakers and the wider public cannot” (p. 83). The careful consideration of teacher capacity within the context of reform implementation will clarify the problematic aspects of policies and reforms from the practitioner perspective.

#### *Teacher Capacity: Definitions and Understandings*

We have examined the origins of assessment reform efforts, including those emerging from within and without the education community, as well as those originating from state legislation. We have tried to uncover some of the assumptions about teaching and learning that inform various approaches to assessment. We have presented the contradictions and inconsistencies that characterize what can be termed

“assessment reform,” and now we bring our focus to the teacher. How do teachers approach assessment in the face of the above-mentioned inconsistencies and contradictions? How do they make meaning of the requirements placed on them and translate those requirements into classroom practice? In order to investigate those questions, we will turn to a discussion of the dimensions of teacher capacity.

In most of the literature about educational capacity, the term is defined in a common sense fashion. It is understood as the ability or power to accomplish some particular educational end (Floden, Goertz, & O’Day, 1995; Jamentz, 1994; Lopez, 1995). Educational capacity is generally defined in system-wide terms as the ability of the education system to help all students meet more challenging standards (Christie, 2001; Fuhrman, 2001; O’Day, Goertz & Floden, 1995). Elements of capacity are generally spoken of in tangible terms, such as personnel, money, time, and infrastructure to effect reform, but some intangible elements are mentioned as well. These include vision and leadership, collective commitment to cultural norms, and knowledge or access to knowledge (O’Day et al., 1995).

Some educational policy experts advocate that we consider interactions among all levels of the educational system when discussing capacity (O’Day et al., 1995). A complete investigation of capacity at all levels of the system is beyond the scope of this inquiry, but we will consider how the actions and policies emanating from the different system levels have an impact on teacher capacity, especially as it pertains to assessment reform.

The predominant discussion of teacher capacity relates to teachers’ ability to understand the reforms that policy makers are seeking to implement (Christie, 2001;

Snow-Renner, 1998; Spillane, 1999; Spillane & Jennings, 1997). In some cases, teacher content knowledge and pedagogical knowledge are also explicitly mentioned as a part of teacher capacity (Lopez, 1995; Bull, 2000; O'Day et al., 1995).

The link between teachers' ability to understand reform and their prior knowledge and experience is not always articulated in the discussions on the capacity to implement reform. The assumption is that the right kind of professional development will suffice as a way of improving teachers' capacity (see e.g. Jamentz, 1994; Snow-Renner, 1998).

Emerging discussions from theorists and practitioners interested in professional development provide some support for a broadened picture of teacher capacity, including teacher as learner (Smylie, 1995; Borko and Putnam, 1995), teacher as agent (Hargreaves, 1995), and teacher as reflective practitioner (Tillema & Imants, 1995). However, the research in professional development frequently extends to such areas as organizational contexts that are certainly part of capacity but beyond the scope of this inquiry.

Other more complex depictions of teacher capacity emerge in some discussions. O'Day et al. (1995) acknowledge that "teacher capacity is multidimensional and evolving" (p. 1), and they list four important dimensions of teacher capacity: knowledge, skills, dispositions, and views of self, including self-as-learner. Borko & Putnam (1995) would add pedagogical knowledge to a definition of teachers' knowledge base. Fuhrman (2001) states that the relationship between teachers' underlying knowledge, access to resources, and belief systems defines capacity and that this capacity, in turn, affects teachers' responses to policy change.

Attitudes towards students as learners and self as learner were also important in studies on reform implementation conducted by Fairman and Firestone (2001) and Spillane (2001). Some authors theorize that teachers' social contexts, such as collaboration, support groups or communities of practice affect teachers' capacity to change practice (Spillane, 1999; O'Day et al. 1995; Windschitl & Sahl, 2002).

Acknowledging the multidimensionality of capacity leads us to examine the contexts of teachers' professional lives, thus enriching the emerging picture of capacity. Teachers are not simply unquestioning implementers of reform-from-above, nor are they automatically reluctant to engage in reform efforts. Both of these assumptions can be found in discussions on teachers' implementation of reform (Tyack & Cuban, 1995; Conley & Goldman, 1998). However, teacher acceptance and implementation of reforms can be linked to several factors, such as their beliefs about students as learners (Wolf & Gearheart, 1997), their uncertainty about the efficacy and durability of the reform movement (Wilson & Floden, 2001, Conley & Goldman, 1998), and their perceptions about the value of the reform in the context of their own teaching (Datnow & Castellano, 2000). Engagement in communities of practice also affects capacity as knowledge is shared and understanding deepened across co-participants (Lieberman, 1996).

A singular definition of teacher capacity as only the ability to implement reform, the policy makers' viewpoint, undercuts teacher agency in important ways. For policy makers, state-mandated reform efforts move in one direction, from the state house to the classroom. Sometimes the final product of an educational reform contains references to public or teacher input, but participation is generally limited in

very real ways (Ogawa, Sandholtz, Martinez-Flores & Scribner, 2003). This leads to a deterministic view of teacher behavior that is characteristic of an authoritarian view of democracy (Bull, 2000). From this viewpoint, teachers are required to address standards-based curriculum and assessment in an accountability framework. In a more transformative view of teacher behavior, teachers would instead use their knowledge and abilities to design curricula that would confront the problems in their schools and communities in locally responsive ways. Unfortunately, this transformative view of teacher agency is unrecognizable in most current discussions on teacher capacity. However, Bull's analysis can provide a framework for helping teachers interrogate and critique aspects of current reforms that they may find problematic. It allows us to envision a more dynamic role for teachers in reform implementation, based on knowledge, dispositions, and their conception of their role. The definition and re-definition of these attributes across social contexts is described as a determinant of teachers' ability to change their practice (Putnam & Borko, 2000; Wells, 1999).

*Studies related to teacher capacity.*

The next section focuses on studies of teacher capacity. Some studies include assessment as a research focus, but for some aspects of capacity, little of the research pertains to assessment. In those instances, research from a wider context will be discussed.

Defining what good teachers know and do, whether by theorizing, observation, or through conversations with teachers themselves, is an elusive task.

Teachers may have difficulty recognizing what they do and articulating how they make the decisions that affect their practice.

Researchers interested in teacher education and the practice of teaching have attempted to sort out the types of knowledge that teachers employ in their craft (Shulman, 1987). Shulman elucidates seven types of knowledge that should be included in a teacher's knowledge base. Among these are content knowledge, general pedagogical knowledge, pedagogical content knowledge and knowledge of educational contexts. He deems pedagogical content knowledge to be of special interest because

it identifies the distinctive bodies of knowledge for teaching. It represents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, represented, and adapted to the diverse interests and abilities of learners, and presented for instruction (Shulman, 1987, p. 8).

Shulman goes on to delineate a model of pedagogical reasoning and action derived from both philosophical and empirical sources. A critique of this model is beyond the scope of our purpose. What is notable for this study is his inclusion of a section on evaluation that assumes knowledge about assessment is part and parcel of the overall knowledge base that teachers utilize.

The research on teachers' basic knowledge of assessment indicates that content knowledge about assessment is an area of concern. Some researchers maintain that the formal training teachers receive in assessment is weak (Stiggins & Conklin, 1992; Jett & Schafer, 1993). This is partly due to the lack of formal coursework required during their pre-service education (Schafer, 1991). Teachers

rate their education in assessment as inadequate and tend to learn assessment in an idiosyncratic fashion (Plake, et al., 1993; Wise, Leslie & Roos, 1991). Furthermore, as Airasian (1991) laments, the coursework that does exist tends to focus on traditional educational measurement instruction with little inclusion of or discussion about other ways to assess students either formally or informally.

A subset of the research on teacher knowledge and beliefs includes teachers' understanding of assessment as a focus, usually in connection with curricular reform based on disciplinary standards, or state-mandated reform, or both. Teachers' learning about assessment is idiosyncratic, and in the absence of feedback, beliefs play a large role in how teachers conceptualize their assessment practices (Wise, et al., 1991). Investigations of assessment used in classrooms uncovered teachers' uncertainties about implementation of various reforms, whether it be assessment (Delandshere & Jones, 1999; Matthews, 1995; Stiggins & Conklin, 1992) or curricular (Datnow & Castellano, 2000; Spillane & Jennings, 1997).

Related to formal knowledge within the context of teacher capacity is teacher understanding of the differing conceptions of "assessment" and the assessment process. Many groups are using the same words but in different contexts and are working from very different assumptions. Some of the issues that might be addressed in this instance would be purposes of assessment (Shepard, 2000; Assessment Training Institute, 2002; Delandshere, 2002); issues of participation (Gipps, 1999; Wolf, 1993) and optimal assessment strategies within the context of curricular change (Delandshere & Jones, 1999; Soodak & Martin-Kniep, 1994).

Assessment is not the only area that has spawned concerns about teachers' mastery of content knowledge and pedagogical content knowledge. The advent of discipline-based standards in the late 1980s and early 1990s such as those from the National Council of Teachers of Mathematics and the National Science Education Standards led to research concerning how teachers were implementing these standards and if classroom instruction was changing. Studies indicated that teachers believed they understood the new curricular approaches well, but researchers found that teachers put them into practice in ways that were inconsistent with the intended reforms (Spillane, 1999; Datnow & Castellano, 2000; Landman, 2000; Snow-Renner, 1998; Wolf & Gearheart, 1997). In some cases, incomplete understanding of the reforms kept teachers from being able to clearly specify their classroom practices in a coherent fashion (Delandshere & Jones, 1999; Wilson & Floden, 2001).

A knowledge base for teaching, comprehensive though it may be, is not sufficient to explain teachers' decisions and actions. Teacher beliefs and attitudes, though difficult to operationalize from a research stance, are acknowledged as an important factor in teachers' practice (Richardson, 1990; Nespor, 1985). Teaching is a complex profession with many demands and few right answers. Immediate feedback in certain cases is non-existent, and in the absence of hard data, beliefs and attitudes assume more importance (Hargreaves et. al, 2002; Nespor, 1985; Tyack & Cuban, 1995).

Teacher belief systems are formed through knowledge garnered through their own practice (Nespor, 1985). Since teacher knowledge cannot ever be deep enough to account for every particularity of practice, teacher beliefs have explanatory power in

seeking to determine why teachers act as they do. Belief systems overlap with other areas of teacher capacity, such as teacher knowledge of particular disciplines or knowledge of proposed reforms. Belief systems include beliefs about content, instruction, and assessment; about students as learners; about their role as professionals; about themselves as learners, and beliefs about the efficacy of external mandates and reforms. All of these inter-related aspects of teacher capacity determine practice. The studies reviewed below have some aspect of teacher beliefs as a research focus.

*Research on teacher beliefs.*

An aspect of teacher belief systems that appears in the literature pertains to teacher beliefs about students as learners. Teachers' perceived lack of high expectations for many students is frequently given as a reason for external implementation of standards (Fuhrman, 2001; Fairman & Firestone, 2001; *State of Maine Learning Results*, 1997; Whitford & Jones, 2000). Teachers who believed that students were capable of working with challenging conceptual material, even in the face of minimal background knowledge, approached classroom instruction with different expectations than teachers who felt that new approaches were too difficult for their students (Wolf & Gearheart, 1997; Wilson & Floden, 2001; Spillane, 2001; Kannapel, Aagaard, Coe & Reeves, 2001; Wolf S., Borko, Elliott & McIver, 2000). Many studies noted teachers whose low expectations, worries about "getting through" the curriculum, and views of students as compliant learners hindered their capacity for change, but there were also teachers described who worked around these

limitations to engage their students in learning experiences that more closely resembled the curricular reforms sought by the advocates of higher disciplinary standards (Wolf S., et al., 2000; McNeil, 2000; Spillane, 2001).

As interest in the situatedness of learning increased, the lack of research on teacher learning in communities of practice was noted. Some researchers posited that the focus of teacher learning needed to be broadened from a concentration on individual teachers to an examination of learning with other colleagues. Some studies report changes in teacher beliefs and practice, especially those studies that involved intensive intervention at the school level by research teams that extended over the course of a school year or more (Wolf et al., 2000; Putnam & Borko, 2000). Other researchers have more cautionary tales to tell, concluding that changing teachers' practice depends on many variables, and collaborative experiences with others may or may not lead to lasting change (Spillane, 1997; Windschitl & Sahl, 2002; Franke, Carpenter, Levi, & Fennema, 2001). For example, one study examined what the author termed "cognitive individualism" as a barrier to teachers' ability to collaborate intellectually (Torres, 1996), and Windschitl & Sahl (2002) noted the heavy workload and relative inexperience as possible factors explaining the lack of change in one of the teachers in their study. Researchers attempted to isolate particular elements of teacher understanding and belief systems that would lead to the desired enactment of reforms. Thus researchers focused on constructs such as "zones of enactment" (Spillane, 1999) or "generative change" (Franke et al., 2001) in their attempts to explain the interaction among the complex factors that are part of teachers' work. Sometimes teachers rate certain practices high on a belief scale but don't follow

through in their actual instruction (Spillane & Jennings, 1997; Wolf & Gearheart, 1997).

Finally, teachers need to believe in the efficacy of the proposed changes. Teachers may be unfairly painted as a source of the problem rather than part of the solution, but they are expected to implement many changes during their careers, and research has shown that this leads to cautious adoption and adaptation of reforms rather than wholehearted acceptance (Wilson & Floden, 2001; Conley & Goldman, 1998; Kannapel, et al. 2001; Matthews, 1995, Datnow & Castellano, 2000).

Teachers' dispositions about their role as professionals will also have an impact on assessment practices. What are the varied aspects of their role? How do they see themselves as professionals? Are they imparters of knowledge, guides to its construction, or some combination? What is their perception of the level of autonomy they have in making major decisions regarding their teaching? Are there conflicts between how they would like to envision their role in the classroom and how they think they must act due to other considerations? For example, teacher complaints about standardized testing and narrowing the curriculum are a common result of certain types of reform efforts (Whitford & Jones, 2001; McNeil, 2000; Kohn, 2000; Landman, 2000).

The distinction between teacher capacity as the skill to ensure certain student outcomes and teacher capacity as the ability to address wider educational and community issues opens up a vision of teacher agency that is ignored in most "policy talk" about capacity (Bull, 2000). This agency is tied in with the aspects of capacity

discussed here, and all of this has an impact on what teachers as learners bring to policy texts (Spillane & Jennings, 1997).

### *Research Questions*

Those who thought that assessment reform linked with challenging standards would mean that all schools would be ratcheting up the levels of instruction and learning are discovering some of the roadblocks to such reform. Nevertheless, evidence exists that reforms are taking hold in some places (Fuhrman, 2001; Wolf, S. et al., 2000). Current state initiatives are designed to bring about alignment to state standards to an unprecedented degree (McNeil, 2000; Coldarci et al., 2000). The purpose of this study is to explore how reforms related to assessment are interpreted and enacted by teachers, both individually and collectively. Questions informing the study from the perspective of individual teachers can be stated thusly: How do different aspects of teacher capacity, such as different types of knowledge and beliefs, interact as teachers interpret policy within the context of their own practice? What conceptions of teaching, learning, and assessment do teachers utilize in their work? What factors contribute to the development of teachers' understanding and use of assessment, particularly when expectations emanating from policy are unclear or contradictory? In what ways have teachers redefined, reorganized or reinvented the new regulations in assessment in order to accept and use them in their practice?

Policy decisions are interpreted and acted upon collectively as well as individually. District capacity in implementing reform is enhanced by intangibles such as knowledge and understanding that teachers and administrators bring to bear

on their collective work as well as tangibles such as money and time (O'Day, et al., 1995). How does individual capacity, including teachers' sense of their own agency, intertwine with communities of practice to support teachers' work in assessment? What elements contribute to the development of educational capacity on the local level? What factors serve to undermine capacity, both individual and collective? This study affords an opportunity to examine the interplay of individual and district-level capacity as the educators involved seek to implement assessment reforms mandated by state policy.

### Chapter Three: Situating the Study: Context and Methodology

Educational researchers and those who formulate educational policy acknowledge that changes in classroom practice are necessary for educational reform, but ascertaining the specific impacts of particular reforms on classroom practice is difficult. Success of educational reforms is usually defined in narrow terms such as improvement on test scores, increased attendance, or higher scores by teachers on self-reported instruments designed to measure reform implementation of various sorts.

Certainly these measures of improvement can contribute to a picture of reform implementation, but a reliance on quantitative measures for evaluation of reform risks obscuring complex issues and perspectives concerning teachers, classroom practice, and reform implementation. Teacher agency tends to be overlooked; the role of the teacher is assumed to be acceptance and implementation, not critique. Complex issues of teacher capacity and its relationship to capacity at systemic levels are acknowledged to a degree, but the panacea of professional development is generally proposed as a way of addressing the issue of teacher capacity (Roeber, 1999).

This study seeks to understand local level reform implementation from the perspectives of the practitioners involved through the use of case study methodology. The cases include individual practitioners as well as the district-level assessment committee. By employing research methods traditionally associated with case studies, I will explore the possible tensions between teachers' practical classroom knowledge and systemic assessment reform required by the state of Maine. There is

an increasing awareness on the part of those who generate policy that issues of teacher and school level capacity have an impact on how policy is interpreted and implemented, yet there is still a reluctance to come to grips with the many variations of implementation that may exist at the local level (Ogawa, et al., 2003; Hargreaves, et al., 2002). It is hoped that an in-depth, local-level look at the process of reform implementation will provide stakeholders with a more comprehensive picture of the challenges and complexities, from the practitioner's standpoint, involved in changing classroom practice.

Maine has embarked on an ambitious educational reform program with several components. While the Maine program has not received the attention given some of the other state initiatives, the legislation passed by the state is designed to have far-reaching effects that alter the nature of schooling for all Maine students (*Learning Results*, 1997). Two different state-level assessment programs are central to this legislation. One is a standardized test that has been given at grades 4, 8, and 11 since 1984, the Maine Educational Assessment (MEA). The other legislation is a mandate for each school district to develop a "comprehensive assessment system" (Chapter 127, 2001, p. 1). These locally developed systems were originally to be in place by fall, 2003, but this time frame has been changed due to a variety of state and national occurrences. The regulations state that all assessments should demonstrate that Maine students are making satisfactory progress towards achievement of the state standards, called the *State of Maine Learning Results*, and hereafter referred to as the *Learning Results* (Coldarci, et al., 1999; *Recommendations for Assessing Maine's Learning Results*, 1997).

The following section of the chapter gives a historical context to the legislation that is integral to this study. Next I describe the local context for the study, including the research participants. This is followed by an explanation of the methodology employed in the study, including data collection and analysis. I end with a discussion of the trustworthiness of the study.

### *Recent History of Maine Assessment*

In this section I describe the three major legislative components that form the backbone of the educational system in Maine at the present time. I examine them in chronological order. The Maine Educational Assessment (MEA) was introduced in 1984, the *State of Maine Learning Results* document was released in 1997, and the Chapter 127 legislation that laid the groundwork for the Local Assessment Systems (LAS) was passed in 2001.

Maine approached the accountability issue by mandating a state-wide standardized test before it developed an official state standards document. Thus in 1984, the state of Maine developed a standardized test for grades 4, 8, and 11 called the Maine Educational Assessment (MEA). Its stated purpose was to monitor “educational accountability and school improvement” (*The use and impact of the MEA, 1985-89*). In connection with the introduction of the MEA, Maine schools undertook programmatic decisions based on the analysis of test results.

The second major legislative undertaking with regard to education was the creation of a standards document. In 1990, a document entitled *Maine’s Common Core of Learning* was published that articulated a common vision for what Maine

students should know and be able to do upon completion of high school. This was the first attempt to tie the MEA to a document that was more oriented towards standards. In the general fervor to implement standards-based reform in the 1990s, it was decided that the *Common Core of Learning* was not sufficiently goal and standards oriented, so in 1993 the legislature directed a task force to “develop long-range education goals and standards for school performance and student performance to improve learning results...” (*Learning Results*, 1997, p.ii). The result of that task force is the document published in 1997 called the *State of Maine Learning Results*. It supercedes the *Common Core*. These are the standards that teachers must address in their classrooms at the present time.

The *Learning Results* document incorporates some of the language from its predecessor, the *Common Core of Learning*. It puts forth six guiding principles “which describe the characteristics of a well-educated person.” These stipulate that “each Maine student must leave school as: a clear and effective communicator; a self-directed and life-long learner; a creative and practical problem solver; a responsible and involved citizen; a collaborative and quality worker; and an integrative and informed thinker” (*Learning Results*, 1997, p.3). Each Guiding Principle also has up to four descriptors that serve to clarify the principles in more concrete terms. For example, one of the descriptors under “clear and effective communicator” states “Reads, listens to and interprets messages from multiple sources” (*Learning Results*, 1997, p. 3). For a complete listing of the Guiding Principles and their descriptors, see the *Learning Results* pp. two and three.

The Guiding Principles are an overarching framework for the standards in Maine, but they are not sufficiently content oriented to satisfy the desire for cohesion between standards and the MEA. In order to accomplish this, the curriculum has been divided into eight subject areas (Career Preparation, English Language Arts, Health and Physical Education, Mathematics, Modern and Classical Languages, Science and Technology, Social Studies, Visual and Performing Arts), and each subject area has been divided into content standards that include both knowledge and skills. Finally, each content standard has been divided into performance indicators that describe what students should know and be able to do at various grade spans. The four grade spans are: pre-K-2; 3 and 4; 5-8; and secondary. For example, the English/Language Arts subject area contains a content standard labeled “Process of Reading,” and for grades pre-K-2, one of the indicators (out of a total of seven) states, “Students will be able to make and confirm predictions about what will be found in a text” (*Learning Results*, 1997, p. 13).

After the adoption of the *Learning Results* by the state legislature, schools were expected to align their curriculum with these standards. The language of the *Learning Results* document explicitly disavows using it as a curriculum; it is intended to inform the curriculum (*Learning Results*, 1997, p.3). Accountability to these state standards was originally accomplished through testing students in grades 4, 8 and 11 with the MEA. The stated purpose of the MEA is to “measure achievement of the *Learning Results*” (*Recommendations for Assessing Maine’s Learning Results*, 1997). In addition, each of the questions on the MEA is referenced to a performance indicator in the *Learning Results*. For example, a question from the 2002 Grade 4

science and technology portion of the test reads, “Curtis pushed a seashell into some clay. Then he took the shell out. The shape in the clay looks like a: 1. rock 2. fossil 3. mineral 4. skeleton.” Underneath the answer to the question the corresponding Learning Result is listed, in this case D 2 which reads, “Continuity and Change— Students will understand the basis for all life and that all living things change over time. Students will be able to: Describe how fossils form” (*Learning Results*, 1997, p. 68).

One cannot escape the pervasive influence of the *Learning Results*, and its espoused flexibility is undermined by the way questions on the MEA are linked to performance indicators. Practitioners’ ideas about teaching a certain topic will be circumscribed by the interpretation favored by the state test. In addition, the breadth of some of the content standards and performance indicators, and the sheer number of performance indicators overall, make it all but impossible for teachers to consider topics outside of the scope of the *Learning Results*, in spite of the official rhetoric.

The third piece of educational policy legislation introduced was an attempt to circumvent criticism directed at the use of the MEA as a gate keeping standardized test. Thus a second aspect of the state assessment plan was proposed (*Recommendations for Assessing Maine’s Learning Results*, 1997). In addition to the MEA, each school district is charged with developing a district-wide assessment plan of which the MEA is only one piece (Chapter 127, 2002). These have come to be referred to as “Local Assessment Systems” (LAS), or “Comprehensive Local Assessment Systems” (CLAS) or “Comprehensive Assessment Systems” (CAS). For purposes of this research, I will refer to these local systems as “Local Assessment

Systems,” or LAS, even though I recognize that the system has both state and local components.

According to the Chapter 127 legislation, each district’s LAS was to be designed for three purposes: to inform and guide teaching and learning; to monitor and hold educational units accountable in achieving the *Learning Results*; and to certify achievement of Maine’s *Learning Results* (Chapter 127, 2001). The legislation originally mandated that these systems be in place by fall, 2003, but that proved to be impossible, especially after the passage of the No Child Left Behind (NCLB) in 2001.

It is worth noting that Maine officials originally attempted to persuade officials from the federal DOE that data from the Local Assessment Systems would be sufficient for the accountability purposes of NCLB. Maine officials did not want to incorporate another standardized test into the curriculum. However, LAS data were not deemed to be sufficiently comparable across the entire state to win approval for this plan (personal communication, Patrick Phillips, Maine Deputy Commissioner of Education, 3-28-04). Therefore the Maine Commissioner of Education decided to acquire a test from an outside source that would fulfill the requirements of NCLB.

The NCLB legislation was not the only roadblock to the design and implementation of the Local Assessment Systems. Information presented to leaders of assessment teams, principals, and others involved in the work has been slow in coming, confusing, and sometimes contradictory. One example of this is the interpretation of the purpose of Chapter 127, the legislation that mandated the establishment of the LAS. The law states that “the purpose of assessment is to: (1) Produce high quality information about student performance that will inform teaching

and enhance learning, (2) Monitor and hold school administrative units accountable for students achieving the content of the standards of the system of Learning Results; and (3) Certify student achievement of the content standards of the system of Learning Results (Chapter 127 section 4.01 (A)). The merging of classroom assessment and assessment for accountability was written into the statute.

Experts in educational measurement caution that attempts to design assessment systems to serve multiple purposes will serve none of them well (Popham, 1997; Delandshere, 2002; Shepard, 2001). Indeed, the Maine Department of Education (MDoE) realized that accountability would be the overriding purpose of the LAS when the following appeared in *LAS Guide* in June, 2003:

The purpose of this document... is to provide procedural guidance on how to build a fully developed framework for local assessment systems that will serve, in particular, the third purpose of assessment, namely the certification of student achievement (p. 3).

What had become increasingly obvious to local practitioners involved in LAS design and implementation was now codified.

Confusion about the purpose of the LAS also carried over into the practical work demanded of schools with regard to the design and implementation of a local system. The initial timeline given out by the Maine DoE was optimistic in the extreme, and information that was specified as forthcoming by May of 2002 was not available until June of 2003 (Policy Advisory Committee Q and A, 2001; Maine's Balance of Representation: Information and Results, June, 2003). An integral piece of the system, the data management system that would allow for both the utilization of assessment data on the local level and reporting to the state for accountability purposes, has yet to arrive as of this writing. Early estimates of the number of

assessments that would comprise a “system” have been revised upward (Local Assessment System Model Narrative, 2-06-02; Interview, # 3, 2-24-04). Originally, the state provided a “working model” for school systems to use, but as the system has developed, the model has changed significantly, and the state has become more prescriptive about the details of constructing a local system. Types of assessments, number of assessments, content standards that are to be assessed, and many other details of the Local Assessment Systems have now been established by the Maine DoE. Thus school districts who tried to meet the original time line that mandated a Local Assessment System in place by the fall of 2003 have had to redo much of their work because of new regulations and requirements (ibid).

#### *The Local Context of the Study*

Maine school districts are classified according to number of students. Most Maine districts, larger cities excepted, encompass more than one town. Many districts have an elementary school in each small town, a consolidated middle school or junior high school, and a consolidated high school. Still others have small schools that serve grades K-8 with a larger consolidated high school. Some districts still have grades K-12 in a single building.

The district where data collection took place is a small district with four elementary schools, a consolidated middle school that serves grades 5-8, and a consolidated high school that serves grades 9-12. It is a class B district, meaning school enrollment is between 500-1000 students. Socio-economic status varies across the district. One of the elementary schools has very few Title I students and as such

was in danger of losing Title I funding while the other elementary schools in the district have a larger percentage of Title I students. Two of the elementary schools are located in small towns, one is in the country, and the other is in a larger town that is almost a “suburb” of a small city. This latter town was at one time a farming community, but as housing developments took over, the percentage of farms has dropped while the percentage of white collar and professional people living in the community has grown. The other towns in the district retain more of a rural character.

*The site.*

The district chosen for this study, given the pseudonym of Beaver Pond, was recommended by the director of an organization that provides professional development to teachers. She is knowledgeable about local school districts, administrators, and, in some cases, individual teachers. Beaver Pond has a reputation for innovative curriculum, administrative support for professional development, and excellent teaching. Soon after the *Learning Results* were published, Beaver Pond established district-wide committees to design curriculum based on nationally recognized content standards and to align that curriculum with the *Learning Results*. This gave the teachers of Beaver Pond some advantages when the Local Assessment System requirement was passed into law. First of all, they were already very familiar with the *Learning Results* and how to use them in appropriate ways within the context of their own classrooms. Second, they had started a collaborative process that

reached across schools and grade levels with considerable sharing of ideas and collaborative support.

Beaver Pond is a small district, but data collection on every teacher was still beyond the scope of this study. Research participants were composed of two overlapping groups: the Beaver Pond district-wide LAS committee, and individual teachers who agreed to participate in the study. Both of these groups contributed to the articulation of state assessment policy on the local level. An examination of the struggles of the Beaver Pond assessment committee as they sought to design a system that included all the components asked for by the state but yet was not overwhelming for teachers will illuminate issues of local capacity. Uncovering the attitudes, understandings, and other elements of teacher capacity will be helpful in understanding the decisions that go into their classroom practice.

Approximately sixteen members comprised the Beaver Pond assessment committee—two from each elementary school, two each from the high school and middle school, several elementary school principals, and a curriculum coordinator from the middle school. One of the elementary principals served as the chair of the committee, and the curriculum coordinator was her assistant. In response to the Chapter 127 regulations for the establishment of a local assessment system, assessment was made a district-wide priority for the 2002-2003 school year. The committee was given release time one afternoon a month in order to meet, all in-service days were spent on assessment issues, and other curriculum committees, grade level committees, etc. throughout the district were to focus on assessment for the year. Even though the district assessment committee had secured release time for

its members, teachers' schedules are such that there was never 100% attendance at any one meeting that I attended. There was, however, a core group that attended most meetings, and this included at least one representative from each school. In addition, the chair and co-chair were excellent about keeping members informed of committee deliberations and decisions.

The teachers at Beaver Pond were eager to be of assistance in my research. I had to limit the number of teachers who participated to the fullest extent because of my own time constraints, not because they were unwilling to participate. Three elementary teachers and one middle school teacher agreed to participate in the research through observation and interviews. In addition, two other elementary teachers were interviewed and/or observed, but less extensively. Two administrators were also participants.

Sampling criteria for the participants focused on different dimensions of teacher capacity. In particular, I inquired initially about teachers with knowledge, experience, and dispositions toward reflective practice and change. These very characteristics were instrumental in the participants' desire to get involved in the research: In some cases, they used the study as an opportunity to investigate their own practice. In a certain sense, we sought each other out.

*Individual teacher and administrator participants.*

Three of the teacher participants are located in one of the elementary schools in the Beaver Pond district, and the other is in the middle school. All of the names used are pseudonyms.

John Wilkins has taught first grade for twenty years. Due to his years of experience, he is a veteran of a number of school reorganization efforts and curricular reforms. Now he is dealing with the *Learning Results* and the LAS. He has incorporated some aspects of certain curricular reforms, such as whole language, into his own practice. He is particularly impressed by the latest standards from the National Council of Teachers of Mathematics, and he and his grade one colleague have adopted a math program that is based on these standards. I observed John regularly during the school year 2001-02, interviewed him, and had numerous informal conversations with him during the course of my research.

Another participant, Elizabeth Traynor, had sixteen third graders in her class during the period of this research. Two of the students warranted extra adult help in order to deal with the varying challenges they faced. This extra adult support allowed Ms. Traynor a great deal of latitude to work with individual students.

Ms. Traynor exemplified the concept of life-long learning through her approach to professional development. For example, she spent her summers traveling around the northeast giving professional development workshops in math. In addition, she was in the process of applying for national board certification. She also believed that one could not complain about curriculum if one did not get involved, so she was willing to participate in committee work in order to have a voice.

The two remaining research participants, one upper elementary teacher and one middle school teacher, were members of the Beaver Pond assessment committee. In committee meetings both of them were advocates for shaping the LAS in a manner that was consistent with local teachers' understanding about what constitutes useful

assessment, and how that might be incorporated into a system that complied with the Chapter 127 regulations.

Sophia Laken taught fifth grade. In addition to her primary responsibility as a classroom teacher, she served on several committees, including the assessment committee. In addition, she was taking graduate courses in order to obtain her principal's certification. In her words, she had "been in education long enough to have been through a couple of these cycles" (school reform), and her coursework was giving her a new perspective on the history of education and educational reform efforts. "You try not to get cynical," she stated. (interview, 3-12-03, p. 13). Sophia's primary goal as a teacher was to "meet her students' needs," (interviews, 3-06-02, p 4; 3-12-03, p. 2).

Polly Benton was a middle school teacher who had just moved up to that level the previous year. The middle school teams were set up so that she looped with the same 48 students through 6<sup>th</sup>, 7<sup>th</sup> and 8<sup>th</sup> grades. In addition to her teaching responsibilities, Polly participated on the Beaver Pond district assessment committee, and she also represented her team on the middle school's assessment committee. As a member of various assessment committees at the state level, Polly had helped score the Maine Educational Assessments and she attended a summer assessment institute held in June, 2002. At the institute, she worked on devising assessments for use in local assessment systems as well as various issues that needed clarification such as balance of representation.

These teachers were veterans. As such, they had a great store of what Connelly and Clandinin (1994) refer to as "personal practical knowledge." This

showed in their ability to combine pedagogical knowledge with disciplinary knowledge, such as John's work on curriculum and Polly's work on assessment. All of these teachers were concerned with individual students as learners; this was evidenced in many of their comments about state mandated reforms.

All of the teacher participants are highly rated by their respective principals and others, such as university personnel, who have had contact with them. They have all served as consultants, workshop facilitators, trainers, etc. for a variety of educational programs and curricular initiatives, both in Maine and elsewhere. During the course of the study, one of the participants successfully undertook the task of obtaining national certification. Another is in the process of obtaining an administrator's license. These teachers would seem to embody the concept of the knowledgeable, reflective practitioner.

Two of the research participants, Marla Smith and Jessica Bean, were administrators in the Beaver Pond district. Marla was the principal of one of the elementary schools in the district, while Jessica held a variety of positions, all with the title of "coordinator," that added up to a full time position. In reality, each of Jessica's part-time positions should have been full-time, but such are the finances of small school districts. Jessica and Marla developed a strong working relationship as the chair and co-chair of the assessment committee. Jessica also met with other committees throughout the district from time to time as they worked on assessment. Marla and Jessica's vision for the LAS and their interpretation of the regulations coming from the MDoE were important to the assessment committee as they struggled with the task of LAS design.

All of the participants have been involved in attending and facilitating workshops on a variety of topics, but one of the teachers and an administrator have been involved in a number of conferences and workshops on assessment, including work at the state level on standards setting and discussions of performance levels. The increased familiarity with the state-level work was helpful as they worked on a local assessment system.

### *Methodology*

#### *Data collection*

The complexities of teacher decision making are not easily uncovered by data gathering techniques that focus on a narrow range of teacher behaviors and dispositions. Gathering data on assessment practices as a part of teacher decision making presents a particular challenge because of the contradictions swirling about assessment reform efforts that confront teachers.

Data were collected from a variety of sources: semi-structured interviews with participants, interviews with state Department of Education employees, notes taken during classroom observations and follow-up interviews, notes from meetings, instructional artifacts, and documents. Appendix A lists interviews, observations and meetings. These are discussed in more detail below.

#### *Interviews.*

Through the use of semi-structured, audio taped interviews I explored the participants' current understanding and practice of assessment and its relationship to

both their instructional practice and issues of assessment reform, particularly regarding the development of the local assessment system. I also addressed issues of what might be called their assessment history as I probed how current practice differs from past practice, and their understanding of why and how these changes occurred. Assessment was discussed in the larger context of state standards and other reforms, both within the district and state-wide, that have been instituted in the past ten years. Most participants were interviewed numerous times. See Appendix B for a list of questions used in initial interviews.

During interviews I explored issues of teacher agency. Many aspects of teachers' professional lives are outside their control, and I was interested in examining the extent to which the focus on assessment was perceived as contradictory to their vision of good teaching. This was especially pertinent given the definition of assessment in the official documents from which teachers would be working.

In addition to interviewing teachers, I interviewed administrators in the districts who were responsible for the design and dissemination of the Local Assessment System. I also interviewed people from the State Department of Education who enlightened me about the latest developments in state policy regarding assessment in Maine and the anticipated outcomes of those policies.

Participants were given the opportunity to read and comment on transcripts of interviews and the subsequent analysis, but the final interpretation of the data for purposes of this study rests with the researcher.

### *Observations.*

In addition to interviews with teachers and others, the researcher observed extensively in two classrooms and occasionally in three other classrooms. In the case of John Wilkins, I observed weekly for one semester and an additional five times the following semester. In Elizabeth Traynor's classroom, I observed nine times over the course of the school year. Polly Benton and Sophia Laken were each observed twice in their classrooms, but they were also observed during LAS committee meetings. Most of the observations were followed by a debriefing session during which I asked questions related to the observation and follow-up questions from previous observations and interviews. In this way I could re-visit important topics and continue to probe for participants' understandings about assessment and its use in their classrooms. My availability in the classroom encouraged wide-ranging conversations about the on-going reforms and a more reflective stance from participants.

### *Meetings and document collection.*

I attended several types of meetings in the course of my data collection. The frequency of the meetings and my attendance at them varied. The types of meetings were:

- 1) Meetings of the Assessment Committee of the Beaver Pond school system. These were scheduled monthly during the school year. The committee had members from every school in the district, although not everyone attended every meeting. I attended

three meetings during the 2001-02 school year, and every meeting except one during the 2002-03 year.

2) Grade level meetings: These were monthly meetings where all the teachers from each grade level got together for an afternoon to discuss grade level concerns. These took on increasing importance as work on the Local Assessment System progressed because of the requirement that certain assessments be “common assessments.” This meant that all teachers in the district had to give the same assessment for certain tasks. I attended four of these meetings, for various grade levels, during the course of the research.

3) District-wide workshops: The district had made a commitment that released time, at least for the 2002-03 school year, would be devoted to assessment. I attended the final work day for teachers in June, 2003. They received updates about the design of the Local Assessment System and worked on common assessments in grade level groups.

When attendance at meetings was not possible, I contacted participants who gave me information about what had transpired. I double-checked the accuracy of this information by asking more than one participant to update me on the meetings that I missed.

Documents collected include instructional and assessment artifacts, assessments published by the Maine DoE for use in Maine classrooms, publications from the Maine DoE about assessment, as well as memos, working papers, and other documents from the Beaver Pond assessment committee.

*Data analysis.*

Data analysis through the constant comparative method has been on-going during the research (Glaser & Strauss). All interviews were taped and transcribed. Interviews were analyzed and categorized initially for themes related to research questions as well as themes raised empirically as a result of meetings, interviews, and classroom observations. A rudimentary coding scheme was developed, and data analysis cycled between thinking about existing data and strategies for augmenting data through further data collection (Miles & Huberman, 1988).

As themes began to emerge from my investigation of the data, they were reanalyzed and re-categorized as a result of further data collection. Emergent themes were cross-checked with research participants as I sought discrepant evidence. Of particular importance at this stage was the decision to consider the LAS committee as another case in the study and hence broaden the definition of capacity to include collective capacity as well as individual capacity. Also at this stage, analysis shifted from a focus on individual stories to a cross-case analysis of emergent themes (Miles & Huberman, 1994).

Quotes have been checked with participants, but final interpretations rest with the researcher. I tried to include quotes that were representative of sentiments expressed across participants. Due to the researcher's proximity to the research site, follow-up conversations and interviews have taken place throughout the data analysis process.

*Trustworthiness of the study.*

Qualitative research studies use a number of verification procedures in order to augment trustworthiness (Glesne, 1999). In this study, the following procedures were used:

- 1) Prolonged engagement—the interviews and observations took place over the course of a year and a half. This allowed me to re-visit questions and issues as I looked at the data as part of the analysis procedure.
- 2) Triangulation—the study used multiple sources of data, including observations, interviews, documents, and attendance at meetings. The study also included both administrators and teachers as participants.
- 3) Attention to researcher bias—I tried to remain aware of the possibility of bias due to my “intuition” as someone with an emic perspective of the research setting.
- 4) Discrepant data—what data did not fit with other data as themes emerged through analysis?
- 5) Member checking—data and analysis were shared with participants.
- 6) Rich descriptions of the research setting, activities and participants.

These procedures enabled me to provide a contextualized description of the research that credibly supports my analysis.

## Chapter Four: Findings

The analysis of the data presented in this chapter uses the previously introduced concept of educational capacity, understood as the ability or power to accomplish some particular educational end (Floden, et al., 1995; Jamentz, 1994). Elements of educational capacity can be both tangible and intangible and are affected by interactions throughout a particular educational system. This analysis is divided into three major sections that coincide with different levels of the K-12 educational system. The first section considers individual teacher capacity; section two branches out to examine other contributing elements to the development of educational capacity on the local level; and the third section addresses the impact of state level policy decisions on educational capacity at the classroom, school and district level. A summary is included after section three.

From a policy standpoint, teacher capacity has been rather narrowly defined as the ability of teachers to understand the reforms that policy makers are seeking to implement and then to change their classroom practice to mesh with these desired reforms (Christie, 2001; Snow-Renner, 1998; Spillane, 1999; Spillane & Jennings, 1997). The analysis presented here, however, utilizes a multidimensional depiction of teacher capacity that includes knowledge, skills, dispositions and views of self to provide a more complete picture of teachers' interpretations of and dispositions toward policy (O'Day et al., 1995). Using the explanatory concept of capacity, the first section of this chapter examines the conceptions of teaching, learning and assessment that individual teachers use in their work. Specific questions include:

What conceptions of teaching, learning and assessment shape teachers' practice?  
How do different elements of capacity interact as teachers interpret policy within the context of their own professional work? What is the impact of individual teachers' knowledge and attitudes on their understanding of and willingness to implement policy changes, specifically, assessment reform?

Individual capacity is a vital component of reform implementation, but use of capacity as an explanatory tool will be extended to demonstrate that collaborative relationships among teachers and administrators are an important element of school-level capacity. The visions that local administrators bring to reform efforts as they attempt to interpret state policy within the context of local level "communities of practice" (Lieberman, 1996) are particularly important. How does individual capacity intertwine with communities of practice to support teachers' collective work in assessment? This research question is addressed in section two.

Section three, which extends the analysis to a multi-pronged examination of capacity within the context of state policy decisions, focuses on three categories of participants: individual teachers, individual administrators, and the Beaver Pond assessment committee. The analysis first details individual and collective attributes that would seem to enhance capacity. It then takes advantage of Arthur Wise's (1979) concept of hyperrationalization to demonstrate how Local Assessment System (LAS) policy undermined individual and local district capacity as practitioners sought to design and implement an assessment system with a focus on local contexts. Legislators and those who translate legislation into policy do not begin their work with the goal of undermining educational capacity. As Wise notes, however, the end

result may be the antithesis of what reformers had envisioned (p. 15). In the present instance, the policy guidelines served to derail local reform efforts.

### *Teacher Capacity and Classroom Practitioners*

Individual teacher capacity includes domains of knowledge, skills and beliefs garnered through formal and informal learning and experience. Ability in assessment is one of the core competencies teachers must acquire in order to be effective in the classroom. As detailed in chapter two, research on assessment has expanded significantly in the past fifteen years. Public attention has grown as well. Maine's Chapter 127 legislation, enacted in 2001, ensured that assessment would become the foremost item on every school's agenda.

This section focuses on the individual teachers who participated in this study and uses evidence gathered from observations and interviews to examine the following dimensions of teacher capacity that emerged from the research data: beliefs about teaching and learning and about students as learners; teacher agency, including adopting or adapting reforms such as the *Learning Results*; and beliefs, attitudes about, and knowledge of assessment. It addresses the question raised in chapter two regarding the relationship between aspects of individual teacher capacity and teachers' understanding of assessment policy. I begin by discussing participants' conceptions of learning and of students as learners. Next I discuss participants' perceptions of the state standards document in the context of capacity for reform. Then I address participants' understanding and use of assessment.

Descriptions of the research participants were provided in chapter three. This section analyzes data primarily from three of the participants, (John Wilkins, Elizabeth Traynor, and Sophia Laken), all of whom are located in one of the elementary schools in the Beaver Pond district, and secondarily from a fourth (Polly Benton), a middle school language arts teacher. Pseudonyms are used to ensure confidentiality.

*Conceptions of learning and students as learners.*

As noted in chapter two, the question, “What does it mean to know?” is seldom addressed in the literature on teaching and learning. In my discussions with these participants, it became clear that their conceptions of learning were not fixed. They sometimes spoke of learning as sequential, hierarchical, and centered around a body of material that needed to be mastered. At the same time, they were aware of developmental differences among students and knew that not all students mastered material in the same way or at the same pace, if at all.

The teacher participants had conceptions about what material would be appropriate for “mastery” at the level they taught. They seemed to have little problem in the abstract with the requirement that students be able to give an on-demand performance of discrete skills. At the elementary level at Beaver Pond as elsewhere, there is considerable agreement as to what basic knowledge and skills are desirable at different grade levels. Curriculum coordinator Jessica Bean indicated: “By the end of eighth grade, all these kids should have this core knowledge. They can have lots of other stuff too...this core body of knowledge will be covered and met, and we will

have evidence of [their] learning that” (interview, 5-07-03, p. 1 #9). Research participants, particularly those working at the primary level, took for granted that certain kinds of basic literacy and numeracy skills had to be mastered before students could move ahead.

However, individual student development was also an acknowledged aspect of the learning process, and the elementary teachers recognized that development was not lock step and was affected by many factors. As John Wilkins stated, “I want to take a child wherever they are and help them make a year’s worth of growth” (interview, 11-23-02, p. 10). John and his colleague, Elizabeth Traynor, structured their classrooms and the school day so that every day they could work with each child individually on the acquisition of important basic skills such as reading, writing and math. It should be noted that this highly individualized curricular approach was possible because of small classes and the presence of adults other than the teacher.

John’s conceptions of students as learners permeated other aspects of his classroom. During group instruction, John was careful to target his questions to students based on his conception of their present level of reading or math mastery. He worked on what he considered essential skills such as letter recognition, sight words, and graphing every day, believing that a little time spent daily on these concepts was more valuable than spending a great deal of time all at once and expecting children to retain the information. Everything from the books he chose for students to read at home to the questions he asked in class was targeted to the level of the particular individual as he understood it from his work with the child. He also

focused on developing critical thinking and other analytical skills, but he did not keep track of these to the extent that he did the basic skills of reading, writing, and math.

Like John, Elizabeth Traynor devoted considerable class time to individual meetings with students. One of Elizabeth's goals for her students was to get them hooked on reading. By meeting with students individually, she could recommend books at their level and keep everyone reading. She was of the opinion that this made for a more vibrant classroom atmosphere as individuals were more engaged in literature and writing (observation, 2-13-02, p.2).

Elizabeth believed that with the consistent modeling of desirable academic and social behaviors, student interaction would eventually improve. For example, she modeled for her students appropriate ways to respond to in-class presentations. She encouraged and practiced respectful behavior in her classroom, and as a result of this, students in her classroom frequently exhibited a concern for each other that is not always a component of third grade classrooms (observations 5-22-02; 10-07-02). Elizabeth believed that personal responsibility was important to academic success, and she emphasized this in her classroom (interview, 1-30-02, p.8).

For Sophia Laken, a fifth grade teacher, "meeting students' needs" was a top priority, but her emphasis shifted from individualized instruction to a focus on more student collaboration in the learning process. This included extending curricular units of study so that students could explore a topic in more depth as well as involving students in self-evaluation, goal setting, and designing assessment instruments. "Over the years, I have learned to respect what students think and what they tell me," remarked Sophia. Sophia encouraged students to set goals as a way of involving

them in the educational process by helping them to focus on both academic content and habits of mind that would be useful both in school and out, such as meeting deadlines.

All of the teacher participants were aware of the futility of demanding that students in the same class meet the same expectations. This theme echoed through individual conversations as well as the deliberations of the Beaver Pond assessment committee. Teachers' professional judgment stemming from both theory and practice informs conceptions of students as capable learners even as teachers realize that not all students are capable of mastering the same material at the same time. Teachers realize that to ask the question, "What does this student know?" is to invite different responses depending on the student. A combination of knowledge and experience influences a teacher's recommendation for a struggling student. Therefore policies that ignore the needs of individual learners are regarded as problematic (meetings, 2-5-03; 2-27-03; 4-2-03). Teachers may be supportive of new policies and programs in the abstract, but when these policies and programs conflict with teachers' professional expertise regarding meeting individual students' needs, teachers will look for ways to circumvent them in some fashion (Datnow & Castellano, 2000).

The participants in this research viewed students as active participants in the learning process, they recognized that individuals brought different abilities and challenges to school, and they knew that their students would learn at different rates. The research participants did not engage solely, or even primarily, in behaviorist teaching practices as characterized by Shepard (1989, 2000). The practices of these teachers were located within a perspective of psychological constructivism (Slife &

Williams, 1995) which as we have seen does not challenge empiricist conceptions of learning. In addition, the framework imposed on their practice by the state standards was within their vision of school as a place where one developed certain desired skills and competencies in order to function in society, i.e. Kliebard's (1995) social efficiency curriculum.

Unease with the idea that the proper environment will lead to the requisite learning (Slife & Williams, 1995; Delandshere 2002) surfaced primarily when participants were confronted with accountability issues. Individual participants, all of whom were familiar with the deleterious effects of high stakes accountability measures in other states, mentioned potential problems in holding all students to the same outcomes. While teachers may have wanted to proclaim that all their students had mastered the required curriculum, their experiences were at odds with the essentialist conception that knowledge, once learned, was "mastered" for all time. As Elizabeth noted, at this age, mastery levels of certain skills that had been stressed as part of the curriculum for weeks dropped significantly after a vacation (10-23-02, p. 9). These elementary teachers viewed student learning as a combination of enhancement of basic skills, gradual acquisition of dispositions thought to be useful in academic settings, and development of more robust conceptual understandings about a variety of topics that may not be easily evaluated quantitatively.

Teachers were reluctant to be held accountable for student mastery of certain material, even if required according to the *Learning Results* or *No Child Left Behind* at their particular grade level. Policymakers were telling them that they were responsible for "certification of achievement of the *Learning Results* (Chapter 127,

2001), their own conceptions of learning were within the empiricist tradition, yet their experiences with students made them wary of the implications of a system that defined learning as providing evidence of mastery of material with little or no concern for individual ability. Their worst fears were that they would be forced to focus all their instructional efforts towards securing certain scores on assessments they deemed minimally useful.

*The capacity for reform: The Learning Results.*

In order to understand teachers' capacity to implement the proposed assessment policy known as Chapter 127, it is instructive to examine teacher capacity with regard to the *Learning Results*, which was the major educational reform that preceded it and is linked to it. Since 1997, all school districts in Maine have undertaken a massive effort of aligning their curriculum to these standards. A thorough investigation of these efforts at curriculum alignment with state standards is beyond the scope of this inquiry. However, the *Learning Results* document is, in many ways, defining the curriculum and instruction in Maine schools, and a discussion of teacher capacity within this context would be incomplete without an examination of teachers' conceptions, beliefs and practices as juxtaposed with mandated state policy.

All three research participants from the elementary school in initial conversations talked of the *Learning Results* in mainly positive terms. They agreed that the *Learning Results* brought needed focus to the curriculum. John thought that certain goals of the national math and science standards were well articulated in the

*Learning Results*. Elizabeth also referred to the *Learning Results* as being focused, specific about what to teach and realistic (interviews, 10-03-01; 2-13-02; 10-23-02). “The *Learning Results* have had a big impact on my teaching. They have allowed us to focus our instruction. With science we were able to look specifically at the objectives from the *Learning Results* and it gave us a narrower teaching field that is more realistic—perhaps four concepts we are trying to teach instead of thirty-five. It makes sense to me to have all those big topics,” (interview, 10-03-01, p. 7). She also thought that the *Learning Results* forced teachers to think purposefully about the activities they provided, rather than just go with a “cute idea.” Sophia liked to integrate language arts and social studies, and she applauded the *Learning Results* for supporting her efforts (interview, 12-03-02).

The middle school teachers I interviewed had to approach the *Learning Results* in a different fashion due to the theme-based curricular approach at the middle school. Curriculum binders had been prepared with ideas for incorporating the *Learning Results* into various themes. Similar binders with assessments were being prepared. Polly Benton, a middle school teacher, explained how they were helpful, “The last theme the students chose was ‘chocolate.’ So I went to the binders, and I found information that fit with the science on classification of plants and the classification system that scientists use. We also did Mayan cultures under Ancient Cultures, and we did some health and nutritional things. We have the same kids for three years, so we developed these tracking sheets so we can list standards that they have covered for each theme. We write what the standard was and the assessment. At this point we have not talked about possible scores; we’re just starting to use these.

And some of these assessments have to be common assessments in order to comply with the regulations,” (interview, 3-10-03, p. 3). These sheets were an attempt to provide the accountability information that the state was requiring; there was no talk of using this information to improve instruction or increase learning. It was merely keeping track of what parts of the *Learning Results* had been part of the curriculum when certain themes had been studied. The middle school teachers had worried that the standards would force them to teach in ways inconsistent with the school’s philosophy, but by paying close attention to the *Learning Results* as part of on-going curriculum design, the teachers at the middle school were able to maintain their approach to curriculum and provide evidence that the students were receiving standards oriented instruction.

Not all comments about the *Learning Results* were positive. Criticisms tended to fall into three categories: sheer amount of material to be covered, developmental inappropriateness of the standards, and the difficulty of expecting students to master the same material in lock step fashion.

The amount of material listed in the *Learning Results* that is supposed to be addressed in the curriculum is daunting. The introduction to the *Learning Results* states that the standards were “not written to prescribe a minimum or passing standard” (p.v). However, in my discussions with research participants, I found they could not envision including any more material than what was already specified. Elizabeth remarked, “I’ve found it to be more than enough at third grade, especially in science and social studies. And if you consider health and career preparation, it’s just too much” (interview, 1-30-02, p. 3). In my conversations with individual

participants, no one thought of the *Learning Results* as leaving room for additional goals except high school teachers who could point to specific curricular areas that were not included, such as calculus (interview, # 1, 7-30-01, p. 3; interview, # 6, 5-10-02, p. 4; meeting 9-20-02 p. 2).

Participants were also vocal about the difficulty of some of the material to be covered in the elementary grades, especially in science. Elizabeth gave me a specific example from the Science portion of the standards. “In science we are supposed to do a little introduction to cells. It’s so abstract for kids. So we do a tiny little introduction to it. They can’t even fathom learning about cells” (1-30-02, p. 3). Sophia was also frustrated with the difficulty of presenting certain concepts from the Science standards so that fifth graders would retain them. “Any of the concepts on matter are way over their heads at this age,” she bemoaned during a discussion of the appropriateness of the standards (5-10-02, p.6).

In addition, the concerns about standardizing curriculum and learning goals that surfaced when teachers talked about assessment were also seen in discussions about the *Learning Results*. As one participant remarked, “We’re not training elephants here; we’re working with little kids.”

There were also practical complaints about the amount of teacher time consumed and all the paperwork required. Third and fourth grade teachers were especially worried about the standards, as they were responsible for an entire portion of the *Learning Results* in just two grade levels (instead of four). Fourth grade teachers in particular faced a lot of pressure since the state standardized test, the

MEA, was given to students for the first time at this grade level. Many thought that test preparation and administration took valuable time away from their teaching.

One participant had a more pedagogically and cognitively oriented criticism of the standards. John agreed with portions of the *Learning Results*, but he thought that in general they were “dumbing everything down. They were meant as a baseline, not a maximum, yet teachers take it as a maximum. Especially in areas such as science and social studies, there is an enormous crunch on time, and it’s too easy just to ‘discuss’ stuff rather than do activities,” (interview, 7-30-01, p.3). During another conversation, he again mentioned the ease with which the lofty ideals of the *Learning Results* could be subverted, “It’s too easy to look at the *Learning Results* as single pieces of information you learn—nuggets you come away with. You want kids to learn these nuggets after a whole array of experiences. Instead what happens is that teachers just focus on the nugget. It’s too easy just to think, ‘I can tell the kids this’ and not give them the experiences so that they can come to the understanding themselves,” (interview, 11-23-02, p. 16).

John’s argument about the possible deleterious effects of the *Learning Results* is cogent and powerful. Why was his the only voice to lodge a conceptual complaint against the standards? The generally positive responses to the *Learning Results* might be interpreted in several ways. People in complex occupations such as teaching always struggle with the need to validate their work. Schools and teachers, in particular, have taken much criticism since the publication of *A Nation at Risk*, and the passage of the *Learning Results* gave a focus and legitimacy to curriculum across the state that many teachers welcomed. As Elizabeth said, “Now we know what to

teach!” Also, the initial context of the *Learning Results* was one of minimal accountability; schools were to be aligning curriculum with the *Learning Results*, but there was not a great deal of definition about what that meant. In addition, many of the individual performance indicators were very broad and open to interpretation. Absent any other mechanism to enforce accountability, teachers had a lot of lee-way in the interpretation of the document.

In addition to the perceived focus and aura of legitimacy that some teachers welcomed, other teachers thought that the *Learning Results* incorporated some of the best aspects of the national standards documents, particularly in science and math. Problem solving and communication, for example, are not always emphasized to a desirable extent, particularly at the elementary level where curriculum tends to be oriented toward the acquisition of basic skills. Thus, advocates of this approach to learning looked favorably on the *Learning Results*.

It must also be noted that teachers may have been wary of responding negatively because, in spite of researcher assurances of confidentiality, they did not want their views to become known to others and perhaps cause them grief. Some may have felt that legitimate criticisms of the standards and the processes by which they were developed may have been used against teachers to portray them as opposed to accountability or unwilling to hold students to high expectations.

The introduction of the *Learning Results* both enhanced and inhibited teacher capacity. Teachers had achieved a degree of comfort with the *Learning Results* since their 1997 publication, and no longer viewed them as an extreme threat to their professional judgment. Indeed, the standards had encouraged reflective practice.

Some teachers welcomed the focus and legitimacy the standards delivered. Collaboration across the Beaver Pond district improved, and conversations about teaching and learning began. These positive developments could have been accomplished without the introduction of a standards document, but nevertheless, the *Learning Results* were the catalyst for these changes. Research participants thought that the *Learning Results* at least had the potential to be an asset in enhancing the performance of Maine students.

The *Learning Results* also inhibited capacity in important ways, chiefly through a narrowed vision of what was desirable curriculum and how learning would be defined. Teachers' professionalism regarding decisions about individual students was threatened by the assumption that all students had to meet the standards on a particular timetable. From a practical standpoint, teachers were overwhelmed by the task of relating their entire curricula to the *Learning Results*. As Elizabeth remarked at one point, "So much paperwork! Isn't it obvious?" (meeting, 4-14-03). Teachers had put a good deal of effort into aligning curriculum to the standards, and burnout was a concern.

For the most part, research participants tried to be positive about the *Learning Results*. The standards did not greatly challenge teachers' conceptions of learning as based in an empiricist epistemological framework. However, the conflation of "high standards for all," rooted in the empiricist tradition, with the constructivist position that we are all capable of "using our minds well" (Wolf, et al., 1991) has put teachers in an untenable position that is creating ambivalence as they acknowledge that all students will not learn the same material at the same rate (if ever). As standards and

assessment become increasingly linked, cognitive dissonance for teachers will also increase (Hargreaves, et al., 2002; Windschitl, 2002).

*Teacher capacity and assessment.*

We now turn specifically to teacher capacity with regard to assessment. I will show how participants' engagement in assessment activities served important formative ends. Nonetheless, participants' own conceptions of assessment purpose remained strongly linked to the mastery model of learning, viz., knowledge is external, able to be divided into discrete hierarchical segments that can be mastered and assessed. Participants' conceptions of learning as mastery of material inhibited their ability to critique the system of state-mandated controls that had been developing since the introduction of the *Learning Results*. With the development of the LAS, this system would further constrict teachers' ability to exercise professional judgment.

Clearly, the research and discussions about assessment that have taken place from the 1990s to the present are affecting teachers' practices. Assessment is frequently more curriculum-embedded and also more participatory. Particularly in grades K-8, there is less emphasis on tests and more emphasis on projects, portfolios and performances as a way of gathering information about what students know. As one participant remarked, "I'm always assessing kids....there are a lot of pieces for assessing progress as the work goes on, and that is a big change" (SH, 5-10-02, p. 1). Four practices that demonstrated the use of assessment in the service of learning were discussed by research participants and observed by the researcher. These were:

rubrics and lists of criteria; individualized assessment; portfolios; and conversations, both among colleagues and with students.

Collaboration with students was a common theme in the development of criteria and rubrics. Sophia used both teacher and student-developed criteria as assessment tools. “To me, not having criteria for students to work from is like not having a job description,” (interview, 3-12-03, p. 20). Students brainstormed criteria for some assignments, while she provided criteria for other assignments. Sometimes they collaboratively developed rubrics. In connection with this, she encouraged her students to assess themselves, and she found them to be honest in their evaluations of their work, partly because they shared their work with one another frequently. They used each others’ work as models, and they practiced formative assessment with each other based on the criteria for the assignment.

Elizabeth, at third grade, used lists of criteria, some of them generated with student input, rather than rubrics. This was part of her strategy to encourage students to take responsibility for their own learning. Students could follow lists of criteria when writing journal entries or solving math “story problems,” but at this grade level many of them could not use a rubric for self or peer evaluation, thus negating its usefulness.

Two of the research participants used a great deal of individualized assessment for the purpose of informing instructional decisions. John worked with every student individually every day on reading skills and math, and a classroom volunteer listened to children read and provided some feedback to John. The information was kept in a binder that provided evidence of student progress through

the course of the year; this method of assessment also gave John a great deal of information about each student's strengths and weaknesses at any given time. He had turned to this system of record keeping when he adopted a whole language approach to reading. He decided to individualize most of his reading instruction and eliminate reading groups for two reasons. First, reading groups labeled children, and he wanted to avoid that stigma as much as possible. Second, his current practice gives him more information about how each child is performing and how best to provide instruction for that child.

Elizabeth also spent a great deal of time conferencing individually with students, especially concerning work related to literacy. In addition to weekly conferences on their journal entries, the students selected books based on interest and reading level, and Elizabeth discussed their selections with them, asking them questions about the plot and its development, character development, etc. They were also required to answer some questions in writing, and occasionally they were responsible for a formal book report.

Through notes on these individualized reading and writing conferences, Elizabeth kept tabs on children's progress and gave them suggestions about their work. Interestingly, she did not consider this assessment. "I wouldn't say assessment is my strength," Elizabeth told me. "I'm not a person who likes to give a lot of tests. I don't see assessment as my strength although I have a good sense of what the children know and don't know," (interview, 10-3-01, p. 4). She frequently worked from an intuitive stance commonly acknowledged to be part of teachers' practices, although teacher intuition as a basis for student evaluation is generally regarded as

problematic and lacking in validity in the assessment literature (Gipps, Brown, McCallum, McAlister, 1995; Beirne, 1997).

After this conversation when she told me that assessment wasn't her strength, I observed the individual reading and writing conferences mentioned above. When I noted that she spent a lot of time in individual assessment, she said, "I hadn't thought about what I was doing as assessment. It never occurred to me to call it that," (interview, 11-14-01 p. 1). To her, this type of formative work with students was not assessment because grading was not involved.

The teachers in the middle school were beginning to experiment with portfolios as a way of documenting progress. These had not yet evolved into portfolios based on the *Learning Results*. They were the result of individual goal setting by students with the assistance of teachers. Polly had encouraged her students to set goals and provide evidence of their attainment through portfolios. She talked about the difficulty of establishing appropriate goals with middle school students. Specificity was a problem, so she and her students devised some goal setting sheets that listed specific discipline-related skills that students might want to improve. Then they could collect evidence to show they were meeting their goals. She kept the goals in mind and would sometimes remind students that a certain piece of work should go in their portfolio as evidence of meeting a certain goal.

Beaver Pond teachers were also becoming aware of the power of conversations about assessment as a way of clarifying expectations and setting performance standards. For example, Polly maintained that the work on assessment that was occurring at the middle school because of the Local Assessment System

(LAS) was beneficial both for its effects on curriculum and for the avenues of communication it had opened among faculty. Because of the nature of the curriculum, discussions about providing evidence for meeting the standards were very important. What would constitute an exemplary piece of writing, or provide evidence that a student knew certain mathematical concepts and how to apply them?

Elementary grade level meetings were another forum for teachers to discuss assessment. The impetus for these discussions was the LAS, but nevertheless, the teachers recognized the value of sharing their ideas (meeting, 4-14-03).

The idea of making requirements transparent so that students would know what was expected of them made sense to these teachers. They did not see the value of making success in school into a guessing game where the teacher knew all the answers and the students were left to figure out how to do well. These teachers were also in favor of working jointly with students to determine criteria.

The research participants also reported other examples of teacher-student conversations about assessment. After giving her students a very difficult math task that was to be commonly graded by teachers at an assessment institute, Sophia encouraged them to discuss the task and share the problems they encountered as well as their solutions. She also worked together with her students to create rubrics, and she encouraged goal setting through the use of an Individual Learning Agreement. This was an agreement between teacher, child and parent where the child set goals, reviewed them periodically with teacher and parents and decided on the next steps (interview, 3-06-02).

These data show that research participants had taken ideas and practices advocated by proponents of formative assessment to heart, but conversations revealed that they considered accountability in the form of mastery to be the overriding purpose of assessment. Elizabeth stated this quite succinctly when she said, “To me, assessment means they’ve got it,” (interview, 10-23-02, p. 4). Both John and Elizabeth were interested in assessment for mastery only for elements of the curriculum that they thought were precursors for other skills. Thus John assessed for mastery in reading and math, but he was opposed to assessing for mastery in such areas as science and social studies because students’ conceptual understanding of the subject matter is tenuous at this age. As an example, he told a story about one first grader who, after three weeks of studying Native Americans, was asked why the Plains Indians lived in tipis and did not live in long houses as the Northeast Indians did. She replied, “Because they would not fit on the planes (Plains)!” John responded, “Okay, so she hadn’t learned a thing from the last three weeks of school. I explained it to her again, but I’m not going back and start all over again in first grade. If that were a reading skill, if I thought she didn’t know her short vowels, and an assessment showed she didn’t, I’d start all over again and re-teach it,” (interview, 11-23-02, p. 9). John thought the most important part of what he should be doing for students was giving them a firm foundation in reading and basic math. Critical thinking, knowledge in other subjects, and creativity were important, but he did not think he should be held accountable for student performance in these areas at this grade level.

Elizabeth's perspective of real assessment was also tied to the mastery conception of knowledge. This was evident in her positive comments about the *Learning Results*. She believed that with enough tenacity, a competent teacher could help a student learn the appropriate grade level material. Thus she was not comfortable assessing students until she felt they "got it," and even then, she recognized that they might not remember it the next year (interview, 10-23-02). She appreciated the limited scope of accountability in the form of report card grades that she was responsible for at her level. She told me, "Luckily I don't have to give a lot of real, real grades on things" (interview, 10-23-02, p.5).

Elizabeth was not comfortable assessing students on material that they were not expected to master. Again, the *Learning Results* were helpful in this regard, as some of the performance indicators clearly indicate that students are merely introduced to certain concepts in third grade. She comfortably used projects and performance assessments in such cases, and she initiated conversations with students about criteria for these assignments. She made a distinction between assessments for information she thought the students should be accountable for, and assessments that were less evaluative for subject areas where she did not consider accountability to be an issue.

Teachers' understanding of learning and student development was one influence on their conceptions of assessment, but the rhetoric of the latest educational reform efforts also had an effect. Noteworthy across participants was the conception of assessment as serving the purpose of accountability, even among teachers like John and Elizabeth who were astute users of individual assessment in an instructional

context. This is an indication of the tenacity of this conception of the purpose of assessment, even though much has been written about other, more formative, uses of assessment during the past fifteen years. All of the participants used assessment for purposes other than accountability, but, as noted above, they frequently did not think of it as real assessment. None of the research participants directly challenged the notion that the primary purpose of assessment was accountability, although there was underlying uneasiness with the potential of this approach to affect practice in undesirable ways.

Elizabeth had information about what she considered ill-conceived assessment practices from her work with teachers in math workshops during the summers. She was fearful of accountability measures that would base teacher salaries on student test scores, for example. Another concern was a curriculum driven by a state test, where “teachers feel like they have no flexibility in instruction, that they pretty much have to teach to the test....it’s out of control, and I think that’s a big mistake,” (interview, 1-30-02, p. 7). She contrasted this with the Maine Educational Assessment. Beaver Pond administrators saw the MEA as one small part of a student’s education, and they were low key about the results, perhaps because Beaver Pond students were at or above the state averages on the MEA (#7, 10-14-01, p. 3).

All participants also worried about what would happen to individual students who would not be able to produce an adequate score on required assessments. Elizabeth’s concern was frequently echoed by all the other participants as she fretted about students who would not be able to perform up to the standard, “The idea that we have to have all these children on the same level makes me crazy. I have students

in my class having trouble with math who could never do [what the tests require]. What's going to happen to these kids?" (interview, 3-5-03, p. 2). In the age of accountability, teachers fretted that their concerns about "the whole child" would be forced to take a back seat.

The emphasis over the last fifteen years on rethinking assessment affected participants' capacity in the use of assessment, but not their ability to critique assessment purpose. These teachers were effectively using assessment techniques that involved students as learners, that went beyond multiple choice tests and that informed instruction (Shepard, 2000; Gipps, 1999). In addition, they were engaged in significant conversations with colleagues and, in some cases, with their students about learning and assessment. They were well aware of the individual differences among their students and effectively responded to these differences in planning instruction.

However, the interaction of two phenomena affected their capacity to critique both local school culture and public policy with regard to assessment. One phenomenon was the growing emphasis on teacher accountability through the rhetoric of No Child Left Behind with its adherence to standards and testing as a measure of learning. This phenomenon grew out of policy makers' desires to provide evidence of increased student achievement soon after reforms were instituted (Mc Donnell, 1994). Teachers were put on the defensive since the rhetoric insinuated that they had been content to ignore students who were struggling academically. The other phenomenon relates to the common conception of the purpose of schooling as transmission of the status quo. Even though teachers were uneasy with the

behaviorist conception of learning as the demonstration of mastery, their location within the psychological constructivist perspective was not sufficiently different for teachers to be able to articulate their unease with this conception of learning. Participants were thus unable to critique ways in which assessment within this context affects classroom practice (Hargreaves, 2002; Gipps, 1999). These two elements reinforced each other, undermined teachers' confidence in their own assessment abilities, and inhibited teachers' capacity to critique the detrimental effects of the accountability model.

*District Capacity: The Beaver Pond Assessment Committee*

Considering the role of individual teacher capacity in reform implementation is necessary but not sufficient to explain local outcomes of policy decisions. Policy decisions are acted upon collectively as well as individually. Therefore, we must consider the role of the assessment committee and its administrative leaders as we look at policy interpretation and implementation at the school district level. This section will show how the individual capacity of the administrative leaders who were involved contributed to the ability of the LAS committee to articulate a vision for an assessment system.

The Beaver Pond assessment committee was an outgrowth of district-wide curriculum committees that were established in the early 1990s. After the advent of the *Learning Results*, the curriculum committees focused on standards-based curriculum. When the establishment of a Local Assessment System became law, the already existing committees were urged to focus on assessment as well as curriculum.

In addition, a district-wide assessment committee was organized, one that would articulate the big picture for the entire district with one administrator in charge, assisted by the person who was both coordinator of the *Learning Results* and the chair of the district-wide curriculum committee (Atkins interview, 2-14-01).

I began following the work of the Beaver Pond assessment committee in February, 2002. I attended meetings monthly until June of 2003. During that time the committee worked to forge an approach to a Local Assessment System that would be consistent with their ideas of what constitutes good teaching and learning while taking into account an ever-increasing number of set guidelines regarding LAS design from the state Department of Education. In this section I will analyze the committee's attempts to come to grips with what state-level policy makers consider to be a bold new approach to accountability on the state level (*Recommendations for Assessing Maine's Learning Results, 1997*). The analysis will show how the administrators were instrumental in interpreting policy, and how their leadership enabled the committee to move ahead with recommendations for LAS design and implementation, however incremental the progress.

An acknowledged aspect of building capacity in order to implement reform involves leadership at the local level (Wolf et al., 2000; O'Day et al., 1995). The Beaver Pond district was fortunate to be able to put administrators in charge of the assessment committee who had a vision of effective assessment practice that was informed by their own experiences (interview, # 9, 5-07-03; interviews, #3, 11-05-02, 7-09-03). This previous work also ensured that these administrators carefully

considered information coming from the DoE in deciding how to craft the local system rather than adopting it *carte blanche*.

The two administrators, Marla Smith and Jessica Bean, were introduced in Chapter Three. Ms. Smith was the principal of one of the elementary schools in the district, while Ms. Bean held two positions, both with the title of “coordinator.”

Marla Smith had been principal in this district since 1995. She started teaching in 1973 and has held a variety of classroom and administrative positions, all at the elementary level. All of the administrators in the district were in charge of a particular committee related to some aspect of work in schools, and Marla was put in charge of the assessment committee.

Jessica Bean served as the Beaver Pond district’s *Learning Results* coordinator, the Gifted and Talented coordinator, and assistant chair of the assessment committee. She had worked in the district for 13 years, and had spent the last three in the coordinator positions. Previously, she had taught second grade.

Marla was determined to build on the already-established initiatives from committees that had aligned curriculum with the *Learning Results*. She also thought that certain aspects of Beaver Pond work on curriculum in response to initiatives from the state DoE had served to encourage teacher discussions and collaboration on topics having to do with assessment. In her eyes, this was very valuable. She thought these conversations were crucial to developing teachers’ understanding of assessment in the service of teaching and learning. These discussions involved topics such as performance standards, obtaining evidence of validity and reliability through looking at student work, and the adoption of common assessments. “We have a lot of

collaboration because of the way we have structured the work. Our work days have led to a sharing of ideas—they get time to talk about what they are teaching and how they are teaching it. Some of these changes in structure are to better meet students’ needs. A lot of sharing is taking place now, such as sharing student work, really discussing curriculum and instruction and assessment. It’s well beyond just sharing ideas for projects,” (interview, 4-24-04, p.2).

Other committee members and research participants concurred. The collaborative discussions of math assessments that took place during the January, 2003, Beaver Pond assessment institute were seen as valuable for their emphasis on looking at and evaluating student work as well as critiquing tasks that appeared on the DoE web site. Such work was seen as beneficial by teachers in helping them to clarify their own expectations, understand student performance and help students understand what is expected of them.

From 1985 until 2004 it was common practice in Maine to have teachers collaboratively scoring essays and constructed response answers on the Maine Educational Assessment. Not all teachers participated in this work, but those that did now understand how collaborative examination of student work leads to a clearer articulation of what “meeting the standard” might mean. These discussions helped put teachers on the same page, so to speak, and thus improved the validity and reliability of scoring such assessments. District teachers who had participated in scoring open response items on the Maine Educational Assessment were familiar with rubrics and other aspects of collaborative scoring. For Marla, this was an issue of

increased capacity that would benefit the district as the assessment committee tackled its work (meeting, 9-20-02).

Grade-level committees were also an important element of district capacity. The elementary teachers had advocated to be able to set aside time once a month to meet with colleagues from other schools in the district who were teaching the same grade. The working relationships thus established helped the elementary teachers when the assessment committee requested information for the local assessment system. Teachers were already familiar with colleagues' curriculum and approaches to instruction, and they had shared ideas and materials. They had some sense of projects, units and assessments held in common and so could comply with the initial requests of the assessment committee for common assessments. The middle school teachers were also accustomed to working in teams within grade levels. Teachers who had participated in assessment-oriented workshops and conferences were able to share their knowledge with colleagues and move the discussions about assessment in a direction that was helpful to the LAS committee.

Marla's awareness of these capacity building elements partially framed her vision for the LAS. Her ideas about the design of the LAS also stemmed from her conceptions of useful assessment. As the outline for the LAS unfolded during the 2002-03 school year, Marla took the initiative in drafting documents that were to be part of the LAS, but the committee as a group re-wrote the proposed drafts. Her efforts at consensus building were rewarded as the committee assumed ownership of the developing system after using her documents as a basis on which to build.

One of the most important decisions that the committee made was to list assessment for the purpose of furthering teaching and learning in their belief statement as the most important reason for assessment. Several committee members were adamant about this point (meeting, 2-27-03, #5, #9, p.3). Even though assessment for learning was mentioned in Chapter 127 as one of the purposes of a local system, many of the teachers were concerned that the state was more interested in accountability (#2, 1-30-02, p.7). The document generated by the Beaver Pond assessment committee detailing their beliefs about assessment was an attempt to retain control over the primary purpose of assessment as defined for the LAS. (See Appendix C).

Also of importance to several committee members was the utility of the system, for them and for their students. It should not be a collection of assessments with no purpose other than to satisfy the requirements of Chapter 127; it should indeed have the potential to inform teaching and learning. Teachers of students in elementary school were particularly concerned that assessments be opportunities for learning and not be unduly frustrating or threatening. The middle school teachers on the committee pushed for assessments that could be embedded in the themes that formed their curriculum. All teachers wanted to avoid labeling students who did not meet the standards as failures.

Marla was adept at suggesting steps in the process of implementing the LAS where a consensus could be reached to try something new. Thus for an all-day teacher workshop on assessment, it was agreed that everyone in the district would give a math assessment. The same assessment would be used for all classes at a

particular grade level such as grade five. The teachers would then discuss and score it together during the workshop. The selected assessments, however, targeted the grade span 5-8 so students in the fifth grade could not complete the task. This happened at other grades also, leading to student frustration in some cases. However, all the teachers agreed that the conversations that centered around scoring and meeting the standard that came from this endeavor were valuable (#4, 3-12-03 p.9; #1, 4-7-04 p.2).

Marla also encouraged the Beaver Pond literacy teachers to construct rubrics for different grade levels that would be used for writing assignments. Again, this was an idea that was widely agreed on by the teachers, partly because of dissatisfaction with the rubrics provided on the assessment web sites. By focusing efforts on assessments that were acknowledged to be necessary and useful, Marla kept the process moving ahead and avoided confrontational issues. The committee and the teachers in the district could feel that they were making progress, even though everyone knew that to get the entire system in place would be a task of some magnitude (meeting 4-02-03 p. 5; interview #3, 2-24-04, p. 3).

As the chair of the assessment committee, Marla had a particular opportunity to influence the emerging design of the Beaver Pond LAS. Her vision was the force behind the committee's work, and her efforts in several directions were intended to enhance district level capacity. In other hands, the LAS might have had a similar design, but teacher investment in its implementation might have been quite different.

Marla rightly envisioned herself as a mediator between several parties. She received information about state-level LAS policy, and it was up to her to share and

explain this information with the teachers and the other administrators in the district. She decided to selectively share this information for two reasons. First, presenting all the information she was receiving would have overwhelmed committee members and teachers, and she was determined not to have that happen. As it was, this tendency to despair surfaced in spite of her best efforts to keep it at bay. Second, information from the state about LAS design and implementation was subject to change without notice. Laying it all out at once would have created extra work that would need to be re-done once the final guidelines were in place.

Marla also tried to incorporate her vision of good assessment into the policy requirements. She pushed for practices with which she agreed that were also compatible with Chapter 127 regulations. This was apparent in such things as her draft of a local philosophy and an additional document on the purposes of assessment, her insistence on teacher collaboration and conversation as a venue for improving curriculum and assessment, her recognition of common assessments as a way to initiate conversations about student learning and performance, and her support of assessment committee members in their insistence that the local assessment system be useful to their purposes and not comprised of “make work” assessments.

When the demands of the state policy threatened to derail what she thought best, she re-directed the committee’s efforts in a direction that was still useful, if less ambitious than what they might have done. Encouraging the literacy teachers to develop writing rubrics while abandoning the inclusion of a portfolio in the LAS were examples of Marla’s pragmatic approach. According to Fullan (2001), the ability of

leaders to act flexibly in the face of state mandates is a sign of enhanced systemic capacity.

Marla was supportive of the state's policy to a point. She expressed support for the *Learning Results* and agreed that assessment was an important part of the learning process. She did not disagree with using the *Learning Results* to inform curriculum development and planning, but as the practical difficulties of implementing the Chapter 127 regulations became more apparent, her support was undermined. She urged the committee to continue the work they all agreed was beneficial to their teaching, but at the same time, she was worried about what ultimately would be required of teachers if and when the LAS became fully operational (interview, 4-24-04, p. 3). As an administrator, she undertook what was asked of her and attempted to articulate a vision for the assessment committee in a supportive and practical way that would allow the Beaver Pond teachers to make incremental progress toward developing an LAS without destroying morale. Teachers were grateful for her efforts to focus on implementation of assessments and procedures from the LAS that were universally acknowledged to be useful while de-emphasizing more contentious aspects of the system that could have derailed the process entirely. She balanced the need to construct the LAS with her knowledge of the limits of teachers' capacities. Perhaps someone else would have approached LAS design in a bolder fashion, but Marla thought that such a top down approach would undermine the collegial climate she had nurtured that was important for this type of work (meeting, 6-04-03).

Marla had incorporated an understanding of assessment that was articulated in previous documents that had been disseminated under the auspices of the MDoE such as *Grand Ideas and Practical Work* (Spruce, 1998) and “What is a Local Assessment System and Who Needs It Anyway?” (Goldman, 1997). These documents stressed assessment practices such as curriculum embedded assessment, looking at student work in order to set standards, and discussions among colleagues about goal setting, instructional approaches and aligning curriculum and assessment. They did not stress technical requirements such as validity, reliability, and sufficiency. These first documents were more concerned with building teacher capacity in linking appropriate kinds of assessment to instruction, and in making good evaluations of student work.

An interview with Emily Baxter, an employee of the DoE who has been involved with assessment, clarified that the original intent was to design a system that would be a combination of state and local measures, with multiple opportunities for students to demonstrate proficiency. Maine eschewed a high stakes test from the outset. The emphasis at the local level was to be on curriculum embedded performance assessment (interview, 2-10-03).

The DoE wanted to provide actual examples of what they considered to be high quality assessments, and so they enlisted teacher volunteers from across the state to design performance-based assessments for the Maine Assessment Portfolio web site. The Local Assessment Development website was added later. It was envisioned that these assessments would assist local districts as they designed their LAS. Only later was it emphasized that the assessments on these sites met “technical standards of

validity and reliability,” a determination that encouraged their use in the LAS over the use of assessments developed locally (meeting, 2-13-02).

Initially, this on-going state-level work encouraged teachers to view curriculum, instruction and assessment from new perspectives without a lot of pressure to put everything into practice at once. Teachers who participated received professional development credits and could engage in substantive discussions with colleagues about assessments, curriculum, and standard setting. As Ms. Baxter noted, it was building both individual and district capacity as teachers shared knowledge and insights gained with their colleagues. As local assessment took on a life of its own, however, many of the capacity enhancing conversations about teaching and learning that pleased Ms. Baxter were replaced by conversations about numbers of assessments and concerns about accountability. The extent to which policy decisions undermined district and individual capacity will be the focus of the next section.

#### *Chapter 127, Policy, and Effects on Capacity*

This section will show how overly prescriptive policy implementation inhibits the accomplishment of reform goals. First, I will examine the effects of MDoE policy on the Beaver Pond LAS committee’s deliberations. Next the concerns of individual research participants will be addressed. Finally, Wise’s concept of hyperrationalization will be used to clarify how state policy guidelines can inhibit rather than enhance capacity.

*The assessment committee and its work.*

Beaver Pond assessment committee participants strived to remain on task during meetings. The attitude toward the work was not one of unbridled enthusiasm, but the tone of committee meetings was not excessively negative. Concerns about the LAS surfaced mostly during individual conversations with research participants and not at committee meetings. Nevertheless, some concerns surfaced at committee meetings as well.

Originally, the state promised local districts a great deal of latitude in the design of the local assessment systems. The rationale was that it would be a large undertaking, but each district could adhere to its own philosophy. The state also realized that local districts would need some support as they proceeded with this work. It is important to keep in mind that the assessment committee began its work with the promise of further information and refinement of the basic regulations from the DoE. In particular, the DoE assured the school districts that information concerning validity and reliability, ranking of performance indicators in order of importance, and number and types of necessary assessments would be forthcoming (Local Assessment System Model Narrative, 2-06-02).

The lack of timely information from the Maine DoE about crucial aspects of LAS design became a major obstacle to the committee's work. How many assessments would be needed? Was the LAS to assess broad content standards or the more specific performance indicators? How many had to be assessed? What could be used besides tests? How would teachers be expected to keep track of all the data? An ironic comment made at one meeting summed up the frustration with this lack of

pertinent information, “First we’ll do it; then they’ll tell us how they want it done,” (meeting, 2-5-03, #4, p.2). One of the administrators in charge of the committee referred to the ongoing work as “a house of cards” because all the work could be rendered useless by one update from the MDoE changing a crucial part of the system (interview, #9, p. 6).

Time was also a frequently discussed concern in committee meetings. There were two perspectives. First, there was the issue of getting the system in place and establishing all the technical requirements. Some committee members were not willing to accept Maine Assessment Portfolio and Local Assessment Development tasks as “valid” without using them with students to determine what flaws might be present, “These things are very raw. You have to give them to kids,” (meeting, 4-02-03, #4, p.5). The time needed to identify assessments that would meet the Chapter 127 regulations seemed overwhelming. “It’ll take fifteen years!” was the comment of one committee member (ibid).

Second, committee members could see that the assessments from the LAS could easily consume as much classroom time as teachers would give them. Marla cautioned the committee members at one point, “This cannot be all you are doing,” (meeting, 9-20-02, p.3). Concerns emerged that “great stuff” that was happening in the district would be shunted aside due to the need for assessment. This concern abated somewhat when the assessment committee drafted its philosophy statement that put assessment for instructional purposes as the major focus of the LAS, but it became more prominent when the official guidelines finally appeared.

When the promised *LAS Guide* arrived in June, 2003, it dealt a serious blow to the committee's work. Gone was any major emphasis on assessment to inform learning; instead, the *Guide* finally acknowledged that the overriding purpose of the LAS was accountability (*LAS Guide, 2003, p.3*). A continuing emphasis on technical requirements coupled with the accountability focus alarmed people. The state's actions were intended to build capacity, but they had the opposite effect. Instead of enabling the committee to pursue a grand vision of a local assessment system, the *Guide* inhibited creative thinking and caused committee members to pull back. "Let's stick with what we have to do," became the *modus operandi* of the committee and characterized the advice that committee members were giving their colleagues in grade-level meetings (meeting, 6-16-03, p. 4).

The emphasis on accountability also hampered the committee in another important way. As local assessment system work became a process of counting assessments, standardizing, searching for assessments that would be considered valid, and generally seeking to fulfill the requirements as laid out in the *Guide*, important conversations that had been started by the committee languished. These included such issues as: "How do we organize assessment to improve our teaching?" "How do we know when a student is 'at grade level'?" "How good is good enough?" "What is good evidence of student learning?" "What supports will we provide for the students who cannot pass these assessments at an acceptable level?" In meetings these questions would be raised, but in-depth discussions did not occur because practical considerations about quantity, quality, and standardization of needed assessments got in the way. Discussion of these questions might have led to increased capacity with

regard to assessment in the service of instruction, but there was no incentive to take the time to consider them.

*Individual participants and the LAS: Issues and concerns.*

This section will address concerns of individual teacher participants about the requirements of the new system and how these might affect their ability as teachers to make decisions about classroom practices. These concerns about agency include professional aspects of capacity such as decision-making about students as well as practical concerns such as the time involved in LAS design.

Research participants were wary of the potential of the LAS to redefine teaching and learning in undesirable ways. As mentioned previously, Marla and Jessica, the leaders of the assessment committee, carefully considered the tasks they asked teachers to undertake. They provided as much moral support as possible as well as tangible support in the form of workshops and opportunities for collaboration. The proposed LAS, however, was a gigantic undertaking that could not be based on existing models from other states, and the uncertainty of the outcome in addition to the amount of work required to design and implement the system weighed on the minds of all participants (*LAS Guide, 2003*).

For those involved in K-12 schooling in Maine at this time, coming to an understanding of what the Chapter 127 regulations required was an arduous task for several reasons. First, specific information was not forthcoming in a timely fashion. Individual districts were told to work on certain aspects of their LAS, then they were told that the state would provide some of the information they needed, particularly

with regard to technical requirements (Local Assessment System Model Narrative, 2-06-02). How or whether to proceed with certain work in light of the state's increasing involvement in setting guidelines was an issue that vexed participants. From the outset as the administrator in charge, Marla was concerned that the state would eventually produce guidelines that would render obsolete the work done by the Beaver Pond committee (interview, 10-26-01, p. 3). In addition, certain information about meeting Chapter 127 technical requirements was confusing and had the unintended consequence of alarming people instead of assisting them. This was particularly true of the information concerning validity and reliability (Designing a Local Assessment System, Part B, 2001). Teacher-designed assessments were valid only if they were aligned with the *Learning Results* by means of a cumbersome fourteen-step assessment protocol (ibid). Teachers were also told by MDoE personnel that all work scored by using rubrics would have to be evaluated by three different people in order to call the scoring reliable (interviews, #9, 5-07-03, p. 5; #4, 2-18-03, p. 4).

As discussed previously, participants tended to equate assessment with mastery of material, even though they also used assessment in other ways. Since participants were aware of the tenuous nature of children's grasp of new material, they were reluctant to certify mastery and unwilling to assess for knowledge and skills that they would not re-teach in subjects such as science and social studies (interviews, # 1, 10-03-01; 6-16-03).

The accountability issue also collided with teachers' ideas about how to meet the needs of individual students. Teachers acknowledged the benefits of standards,

but they also knew that all students could not be held accountable to the same high standard. Succinctly stated, the question became, “What’s going to happen to these kids?” (#5, 3-10-02, p. 18). This concern was frequently voiced by individual participants and at assessment committee meetings. The mandates of Chapter 127 and NCLB that established arbitrary levels of achievement at certain grade levels ran counter to the professional instincts of research participants who preferred to look at the whole child, recognized that children grew and developed at different rates, and realized that social and emotional growth had an effect on academic progress. Teachers were worried that their professional decision-making power regarding individual students was in jeopardy, and they were concerned about the implications of this loss of agency.

Confusion also existed among local practitioners about the purposes of assessment. The original language of Chapter 127 listed assessment to inform instruction as one of the purposes of the LAS, and the Beaver Pond teachers wanted to design a system that took this seriously. As the system became more prescriptive, concerns rose about the utility of the assessments for enhancing learning, the potential of the assessments to become threatening or frustrating for students, and the paucity of information from these assessments that would be useful for the purpose of improving instruction. The term “jumping through hoops” was frequently used in connection with the amount and type of assessments required under MDoE guidelines (interview, 3-10-03, p.18 #5). Marla’s frustration is typical of comments heard from others, “Assessing should be about moving students ahead. It doesn’t always have to be a high-falutin’ test,” (interview, 2-24-04, p. 2). Certain assessment approaches,

such as observation, could not be incorporated into the LAS once the emphasis became focused on accountability. Several participants thought that much of their vision of the LAS was undermined by state guidelines, “We worked so hard to keep assessment for learning uppermost!”(interview, 6-5-03, #3).

Elizabeth Traynor, who had talked with a number of teachers from other states through her summer workshops, voiced a different accountability concern. She knew that in other states teachers were evaluated on the basis of student test scores. She thought this was a very dangerous trend, and as the LAS shifted its focus to accountability, she became wary of this potential undesirable development. It led to teaching to the test, in her opinion, and was not good for teachers or students.

Teacher input about appropriate amounts of standardization of assessments was also a concern. As mentioned earlier, many participants applauded the coherence that they thought the *Learning Results* brought to curriculum, and they did not necessarily view having the same curriculum across the district as worrisome. However, the level of standardization required by the LAS was overriding teachers’ opportunities to make decisions about individuals in their classrooms. Should common assessments have “standard” directions about how to give them? What about students who would not be able to complete the assessment satisfactorily? How could teachers legitimately take account of individual student differences, especially among younger children? There were fears that Chapter 127 requirements had “thrown a wedge” in the curricular work they were doing (meeting, 4-9-03, #2, p. 4). Concerns were emerging that the state would go too far in standardizing what teachers had to do, and that they would end up teaching to the test as had happened in

other states. As one participant stated, “Schools have to be a creative place for kids. The other part of that is, we don’t want to lose the art of teaching and teacher passion for their subject,” (interview, 5-07-03, p.5 #9). Beaver Pond teachers concurred that “when we have to do everything all the same, it will be a sad day” (6-16-03, p.1).

Some participants observed the increasing amount of state control over the LAS and were worried about what that might bring. The veteran teachers in this study had seen many types of reform come and go. As John remarked, “This is SO controlled by the state; it’s hard to just let it slide. If they want that degree of control, why not just have a state curriculum, instead of forcing all these things on us in the name of local control! I feel like telling them, just let me do this,” (interview, 6-16-03, p.2). This was the sentiment of some of the members of the Beaver Pond assessment committee who were frustrated at state requirements that were handed down after they had spent months working on the type of system that they considered the most beneficial.

Teachers were also concerned about being pressured by deadlines into adopting an assessment system that would not fit their needs. Research has shown that teachers reflect on information of all kinds that is given to them, from professional development workshops to state standards, and then they adjust their practice according to personal beliefs about teaching (Wilson & Floden, 2001). Both Elizabeth and John discussed this process as they talked about changes in their practice and how those changes had come about. Elizabeth voiced concerns about the ability to approach the LAS in this reflective manner. “My worry now with local

assessments is that we won't have time to process the information. I'm worried we'll jump on a bandwagon without thinking it through," (interview, 2-13-02, p. 2).

Because of her role as the local administrator in charge of the LAS, Marla was very much aware of issues of capacity at different levels of the system. She realized from the outset that, at the local level, amassing evidence to support the validity of every assessment in the LAS would take years (interview, 10-26-01, p.2). As the guidelines from the MDoE grew ever more prescriptive concerning assessments that were appropriate for the LAS, Marla became more concerned about the validity of the entire system, "How can you say someone met the standard, even with eight assessments? How can you tell if someone really knows life sciences? I don't know. As accountability gets bigger, I see more problems" (interview, 2-24-04, p.3).

The attempt to utilize one assessment system for different assessment purposes without acknowledging or clarifying the tensions that would occur was a further difficulty. For Marla, the combination of the tensions that arose and the lack of guidance from the MDoE about resolving them caused her to rethink her plan for system design and implementation. She stated, "Instead of trying to cover everything, we're meeting the letter of the law" (interview, 4-24-04, p.1).

The concerns of the Beaver Pond teachers involved both pragmatism and agency. If they had to develop a local system, they wanted to be able to design it in the way that was most useful to them. Including all the accountability measures that were eventually handed down from the MDoE was not part of their original plan. The utility of these accountability measures was very much an open question for the research participants. The other concern involved control of their lives as teachers.

Decisions about students, curriculum, pedagogy and a host of other professional issues were usurped by the state under the guise of the LAS. At the time of this research, teachers were only beginning to get a sense of this, but the concerns are clearly evident in remarks made by participants.

A more nuanced and complex version of teacher capacity emerges from this analysis of the Beaver Pond teachers and their concerns. The participants' capacity was not limited to knowledge, disposition, ability (O'Day et al., 1995; Floden et al., 1995) but included concerns about their decision-making powers as professionals (Bull, 2000). The contradictory approaches to assessment espoused by policymakers (McDonnell, 1994; *LAS Guide*, 2003) and researchers (Gipps, 1994; Shepard, 2000; Resnick & Resnick, 1992) led to tension in the professional lives of these teachers regarding assessment purpose. That is, is assessment used to demonstrate mastery of material, or is it used to gather information about student performance and as a guide to instruction? The Beaver Pond teachers were left to wrestle with the appropriate way to accommodate policy demands within their conceptions of best practice (McNeil, 2000; Spillane, 2001; Datnow & Castellano, 2000).

The professional decision-making aspects of capacity discussed above are not the only elements of capacity that affect a school system's ability to execute mandated reforms. Of particular concern in the present instance was the more practical issue of time. Participant concerns about time were expressed with regard to both LAS design and eventual implementation.

The Beaver Pond assessment committee, the administrators, and the teachers were all aware of the magnitude of the task before them. One of Marla's major

accomplishments as the LAS project got under way was to keep the committee and the rest of the teachers moving forward, however incrementally. However, participants were still prone to despair, as the following comment from Elizabeth demonstrates, “Everyone is freaked out! It’s huge!” (interview, 5-22-02). Other teachers concurred, “The sharing and chance to plan curriculum have been valuable. But after all these meetings, we’re maybe one-tenth of the way through English/ Language Arts. How much time will it take to do this for all the subjects?” (meeting, 4-09-03, #2, p. 1).

Other teachers worried that the emphasis on assessment would usurp instructional time in undesirable ways. For these teachers, the handwriting was on the wall. As the LAS was taking shape, they foresaw that they would spend more actual time assessing than teaching. One participant who also served on the Beaver Pond LAS committee gave voice to the frustrations of many of her colleagues with the following observation: “We see on the horizon more time being spent on assessments, eroding the little time that we have to teach. Each year we get less and less done. Due to lack of time, especially with science and social studies, we’ll create an assessment and teach to it,” (interview, # 4, 6-03-04).

*Legislative intent and hyperrationalization.*

The Chapter 127 legislation empowered local school districts by giving them a great deal of control over the design and implementation of local assessment systems. True, the amount of work involved was potentially enormous, but the promise of local control was part of the legislative intent. Beaver Pond teachers and

administrators individually and collectively brought many strengths to the table as they began work on the LAS. These included an experienced and reflective staff, a history of collaboration throughout the district, staff members who were familiar with types of assessments other than traditional tests, and a cadre of teachers and administrators who could articulate what they did and did not want from an assessment system. These capacity enhancing attributes assisted the Beaver Pond district as teachers and administrators began LAS design.

Wise's concept of hyperrationalization helps explain why the promise of Chapter 127 was not realized. The original intent of the legislation was to empower individual school districts to develop contextually sensitive assessment systems that de-emphasized standardized tests and encouraged curriculum-embedded assessment in the service of learning (Baxter, 2-10-03). Policy makers did not initially acknowledge the tension between assessment to inform instruction and assessment for accountability. However, as accountability overcame all other purposes of the LAS, local districts were required to alter the LAS in ways inconsistent with their vision of assessment. In addition to changing the locus of LAS decision making from local districts to the state DoE, other issues hindered the LAS design process. These included a lack of timely information from the DoE, overly prescriptive guidelines, erroneous information about important concepts, and a general lack of awareness about the magnitude of the task and the lack of time and money available to individual school districts to complete it within the allotted time frame, or perhaps at all. However helpful it was meant to be, LAS policy developed in such a manner that negative consequences were unavoidable for Maine schools.

A policy is operating hyperrationally when its outcome is the opposite of what is intended (Wise, 1979). In the present case, state policy had the unintended effect of disempowering the Beaver Pond district. Instead of encouraging assessment in the service of learning, instruction was co-opted to serve assessment. State policy turned the LAS into a search for assessments that could be checked off to meet the required number. Many positive elements of capacity that could serve Beaver Pond teachers well, such as the grade-level discussion groups, were ignored or underutilized in the design of the LAS. This will be further explored in the concluding chapter.

### *Summary*

In this chapter I have analyzed data for instances of capacity with regard to assessment practices and policy. My interviews of teachers and administrators in the Beaver Pond district, my observations of teachers in their classrooms, and my attendance at meetings and workshops involving discussions about assessment have informed my conclusions about individual teacher capacity, district capacity, and the impact of state level policy decisions on educational capacity at all levels.

The teachers whose classrooms I observed worked from a complex conception of student learning that acknowledged individual differences but still conceived of knowledge as a body of material to be mastered. Teachers were aware of the futility of demanding that all students demonstrate their mastery of material at the same rate, but they did not critique this conception of learning, which was reinforced by state policies such as the *Learning Results* and the Local Assessment System.

Since 1997, teachers in Maine have used the state standards, known as the *Learning Results*, to inform curriculum and instruction. Welcomed by some as bringing legitimacy and coherence to the curriculum, the *Learning Results* also promised greater accountability since schools have to certify that all students have met the standards. The introduction of the *Learning Results* affected both individual and district capacity. On an individual level, the necessity of aligning instruction with the standards increased participants' reflective practice. On the district level, teachers collaborated on curriculum alignment with the standards and initiated conversations about curriculum and instruction that enhanced capacity.

Participants' assessment practices showed the influence of research and writing on alternative approaches to assessment that was prevalent during the 1990s. However, the growing emphasis on accountability reinforced conceptions of teaching, learning and assessment that focus attention on student mastery of a body of knowledge. Participants experienced tension between this approach and their professional knowledge about students as learners, but they were unable to effectively challenge the dominant paradigm.

The members of the Beaver Pond assessment committee began their work of designing a Local Assessment System from an advantageous position. The administrative members of the committee as well as many of the teachers who were involved had a vision of assessment that they wanted to implement in their design of a local system. Collaborative relationships within grade levels and across the district had already been established through work on curriculum alignment. There was a core of teachers and administrators who, through work on state-level committees, was

familiar with the reforms envisioned by state-level personnel. All of these attributes as well as the knowledge and experience that come with a veteran teaching corps placed the Beaver Pond district at an advantage as they began their work on the LAS.

However, state policy worked to inhibit the capacity of the Beaver Pond district. Lack of information, unclear directives, overly prescriptive guidelines, and confusion about the overarching purpose of the local assessment system hampered the Beaver Pond committee in its efforts and caused the committee as a whole and individuals in the Beaver Pond district to re-evaluate the design and implementation of the LAS. This is an example of hyperrationalization (Wise, 1979) at work.

## Chapter Five: Discussion

It is no exaggeration to say that assessment is consuming Maine educators. Currently throughout the United States, educators, students and parents are trying to come to grips with the repercussions of the federal legislation known as No Child Left Behind. In addition, Maine educators must contend with the Chapter 127 legislation that requires them to develop a local assessment system tied to Maine's state standards. The findings of this research have demonstrated how individual and collective attributes that would seem to enhance capacity for enacting reform instead are derailed in the effort to bring about desired reform. Local practitioners are not acting against reform, *per se*, but are responding as professionals to aspects of policy decisions that inhibit their professional capacity.

This research provides evidence that as teachers are attempting to deal with this system, they are reacting in ways that are somewhat detrimental to their own and students' engagement in the learning process, yet by behaving in this manner they are preserving their agency as professionals. In this chapter I first discuss how all the elements of Maine's hyperrationalized educational reform combine to inhibit rather than support teacher capacity. I then discuss how Beaver Pond teachers respond to the assault on teacher agency through two mechanisms: defensive simplification and cautious implementation. Both approaches are notable for what practitioners did and did not do in their attempts to maintain the integrity of their practice. A concluding section ends this part of the chapter.

The chapter continues with the following topics pertinent to the research: usefulness of this study in other contexts, limitations of the study and ideas for further research.

### *Hyperrationalization Revisited*

Returning to Arthur Wise and his ideas from *Legislated Learning* provides a context for understanding the behavior of Maine teachers with regard to educational policy in Maine. In *Legislated Learning* (1979), Wise defined hyperrationalization as “a tight mechanism of goals and controls in a domain where such a tight mechanism can undercut the fundamental purpose of the enterprise” (p. 66). Educational policy in Maine, with its prescriptive mandates encompassing state standards, a standardized state test (however low the stakes may be for individual students), and a mandated assessment system, is a textbook example of Wise’s concept of a tight mechanism of controls.

The institutionalization of this mechanism of goals and controls has been a gradual process. The state standardized test was first administered in 1987, but the state standards were not published, let alone implemented, until 1997, and the requirement to design and implement a Local Assessment System (LAS) was not put into effect until 2001. With each step, the reification of the *Learning Results* has become more pronounced, forcing teachers to accommodate to it and stymieing attempts to examine it critically on its own merits and as a basis for the LAS. As the accountability stakes have risen, teachers have become less comfortable with the avenues of interpretation left open to them concerning the *Learning Results*.

*Hyperrationalization as capacity inhibiting: Hanging on while letting go.*

Chapter two alluded to Smyth's (1995) argument that educational policy is "teacher blind," that is, it disregards teachers and the situations in which they work. Smyth argues that certain policies encourage or require teachers to develop collaborative structures, but these structures then are used for essentially managerial ends (p.80). Smyth's analysis fits the Maine context very well. Maine's development of both the state standards and the local assessment system is an example of what Smyth described as "hanging on while letting go" (p. 79). This trend leads to an apparent relaxation of state control and seemingly allows teachers the flexibility to make decisions collaboratively about their work. In reality, the policy decisions that affect the collegial work are made outside the schools with little regard for teachers' capacity to reflect on their work.

The development of both the *Learning Results* and the LAS are examples of this. An essay on the *Learning Results* design process mentions two thousand educators who responded with their concerns, and it states that the final document reflects these concerns (*Learning Results*, 1997). The process is portrayed as open and leading to a state-wide consensus. It does not explain, however, to what extent educators' concerns were allowed to alter the final document, what parts of the document were open to critique, or whose opinions eventually held sway. The rhetoric of consensus masks a process whereby the design of important elements of the system was farmed out to other organizations, thus hiding the decision-making process (Moss & Schutz, 2001). The interpretive freedom of teachers with regard to

the *Learning Results* has eroded as additional state policies have defined and determined the meaning of this standards document, usurping teachers' professional judgment in this area.

Smyth's argument applies as well to the situation in Maine with regard to the Local Assessment Systems. School systems were told they had to develop local assessment systems, but the state did not want to appear dictatorial. Thus the legislation requiring the design and implementation of local assessment systems, as well as the initial communications from the MDoE, promised a maximum amount of latitude and local control in fulfilling the mandate. However, the *LAS Guide*, prepared in large part by outside consultants, mandated a framework for the LAS that emphasized accountability as its dominant purpose. The prescriptive requirements of *The LAS Guide* ensured that accountability concerns would play a larger role in the development of local assessment systems as a result of state assessment policy. As this research has documented, the state has mandated numbers of assessments, types of assessments, and has forced schools that want to use their own locally developed assessments to provide evidence as to their validity and reliability. During the course of this research, participants showed increasing apprehension over what they considered to be the potential dangers inherent in the number of state directives concerning the LAS framework. Some participants expressed frustration with what they considered to be a disingenuous approach by the state to the issue of local autonomy as well as a lack of consideration for teachers' professional judgments. "Letting go" i.e. allowing teachers to design the LAS, became overwhelmed by

“hanging on” i.e. ensuring that accountability mechanisms were adhered to state-wide.

*Teacher response: Defensive simplification.*

The tightening of the links between all aspects of the standards, curriculum and assessment affected teacher responses to reform efforts. When standardization and accountability become the controlling educational rationale, teachers may respond to the bureaucratization of schooling in several ways. These include omitting parts of the curriculum that they deem inessential, fragmenting the curriculum into discrete segments that are perceived to be easily taught and measured for testing purposes, and simplifying the curriculum so that teachers and students are held accountable for both less material and less challenging material (McNeil, 2000). McNeil refers to this last response as “defensive simplification” (p. 11).

Defensive simplification is a useful concept for analyzing the reactions of the participants in the research described here. Defensive simplification was noticeable in the reactions of individuals and the variously constituted committees. It figured in the deliberations regarding curriculum and assessment as the Beaver Pond committees worked on their local assessment system.

The introduction of the *Learning Results* had already resulted in a large amount of curricular revision with some simplification. Anything that could not be directly tied to a performance indicator in the *Learning Results* for a particular grade level was dropped. Teachers eliminated topics and units that they previously taught. For example, certain literary genres and also concepts in science were written out of

the curriculum as unnecessary and not worth the time since they weren't listed in the standards.

Another understandable, though disturbing, trend toward simplification was the use of the exact language of the *Learning Results* in writing curriculum. Using the standards not just as a guide but as the only framework for curriculum design imbues them with a validity and legitimacy concerning knowledge claims that is far beyond what these standards or the process that spawned them can support. Some teachers have stopped thinking beyond the covers of the *Learning Results*.

This was seen by some teachers as a sensible move, particularly when the performance indicators for certain subjects were deemed too difficult for the age levels involved and hence not developmentally appropriate. Research participants also acknowledged framing their instruction and practice assessments in the exact language of the *Learning Results* as a way of boosting student performance on the Maine Educational Assessment.

The trend toward defensive simplification that started with the overwhelming task of re-aligning curriculum with the *Learning Results* has been exacerbated by the legislation mandating the implementation of Local Assessment Systems. As the state mandated more and more elements of the LAS, the Beaver Pond committee decided, in the words of one research participant, to "stick with what we have to do. Whatever we put down, we will be accountable for." Thus portfolios for language arts were abandoned as part of the local assessment system because teachers were unsure how they and their students would be held accountable if they decided to use them. The grade level committees also adhered strictly to the language of the *Learning Results*

and did not include assessment of any performance indicators where students were merely introduced to a topic at a particular grade level. For example, the first grade committee was able to agree most readily on assessing discrete skills that they considered age appropriate, such as knowledge of certain math facts, spelling, writing conventions and decoding. More contentious issues, such as assessments for science and social studies, were left for later.

Both local district capacity and individual teacher capacity were affected by the state's handling of Chapter 127. The initial emphasis on maximum freedom for local design was gradually replaced by state control of all the important aspects. This process created much confusion and uncertainty about what was appropriate in terms of local system design and also what would be required. Defensive simplification was a natural reaction to the level of uncertainty and contradiction that emanated from the MDoE about the guidelines for Local Assessment Systems. The Beaver Pond assessment committee had some understanding of the use of assessment for other than accountability purposes, but as state policy became more prescriptive in the guidelines for the LAS, the committee grew more cautious in what they were willing to propose for adoption. When important information about the task at hand changes from month to month due to actions at higher levels, local enactors become reluctant to take bold steps. Indeed, one participant foresaw the day when teachers would select an assessment from one of the MDoE websites, plan a unit of instruction around the assessment, and proceed from there. That this is merely another form of "teaching to the test" did not occur to her.

Defensive simplification was also a pedagogical issue. The math and science portions of the standards had been expressly modeled after national standards that encourage constructivist approaches for young children, including exploration and conversation leading to a more robust construction of concepts. This type of instruction takes time. There were indications that teachers did not always take the time to present instruction in an exploratory fashion and instead reduced time spent on exploration in favor of more structured lessons and didactic explanations of the information. In the words of one participant, “The *Learning Results* are dumbing everything down.”

With the passage of the No Child Left Behind Act, the state of Maine engaged in some strategies of defensive simplification on the state level also. The MDoE announced that, starting in 2003, the Maine Educational Assessment, which had previously targeted all the content areas of the *Learning Results*, would test only language arts and math, as required by NCLB. Certification of achievement of all the *Learning Results* as determined by performance on local assessments would still be necessary in order for students to be able to graduate with an academic diploma, but the target date for implementing this was pushed back.

*Cautious implementation.*

Defensive simplification is one strategy at individual and systemic levels to the constricting requirements of standards linked to assessment and accountability that have come to dominate the educational landscape in Maine and elsewhere. However, in addition to defensive simplification, the confusing and contradictory

requirements of Chapter 127 regarding the establishment of local assessment systems, as well as the frequent changes in the most basic information about LAS design, led to a pattern of implementation on the local level that satisfied the needs of the local practitioners but fell short of a radical revision of their assessment practices. I refer to this approach as cautious implementation. It is a practitioner approach similar to what Wilson and Floden (2001) refer to in their article on standards-based reform efforts as “hedging bets.”

Cautious implementation of the LAS was a legacy of the drawn out process involved with the state’s articulation of LAS guidelines. This has been documented elsewhere in this study. As a result of the stream of constantly changing directives concerning LAS design, the Beaver Pond committee began implementing aspects of the local system that were viewed by Beaver Pond teachers and administrators as having the most local ownership, were the least controversial, and that served as a mechanism for teachers to compare student performance across schools. Thus district-wide writing prompts, scored with a district-wide rubric in collaboration with other teachers from the same grade level, were one of the first pieces of the LAS to be implemented in the Beaver Pond district. The use of writing prompts was not a new practice in this district, but the decisions to use a common prompt across grade levels and engage in collaborative scoring were related to the LAS. The time spent in scoring and discussing these, though considerable, was looked at by all participants as time well spent. However, the amount of time available for collaborative scoring was cut back, making it virtually impossible to double score all the essays in subsequent years without spending large amounts of time outside the classroom.

Elementary level teachers were able to reach a certain degree of consensus about language arts and math curricula. The need for basic math and language arts skills is not disputed. Beyond these basic areas of agreement loomed many potential problems, including reasons for assessment, the number of assessments that will eventually be required as part of the LAS, and a creeping level of standardization that most teachers wanted to avoid. “It will be a sad day when we have to do everything exactly the same,” commented one participant. Cautious implementation of a few agreed-upon components of the developing LAS and the use of data from these assessments demonstrated to the school administration that the Beaver Pond assessment committee was proceeding with their work at all deliberate speed while teachers were not pushed outside their comfort zones in dealing with aspects of the LAS that might prove more controversial than what had already been developed.

The continuing issues of time and money also affected capacity. The administrator in charge of the committee was also the principal of one of the elementary schools, and she found herself spending less and less time on issues related to running the school and more time on district assessment work. The people on the committee were exhausted from the numerous meetings, frustrations with the MDoE and their changing rules and regulations, and the need to explain the importance of the committee’s work to ambivalent and/or hostile colleagues. The committee had to plan and negotiate with the superintendent in order to gain approval for a minimum of in-service days to meet with teachers and to organize collaborative scoring. Even this minimal amount of released time was not secure, as evidenced above. Parent complaints resulted in the elimination of half of the allotted time for

collaborative scoring of district-wide writing assessments. Crucial supports for the type of work that was envisioned in the Chapter 127 legislation were not forthcoming. It is worth noting that a recent report on LAS implementation in eighteen school districts that was commissioned by the Maine DoE noted that districts “have attempted to fit the work of LAS into the school calendar by using the traditional mechanisms of early release days, in-service days, and time set aside for committee work. In almost all cases, those interviewed report that this has been insufficient” (LASIS Final Report, 2004).

*Relationship to capacity.*

The two issues described here, defensive simplification and cautious implementation, are sides of the same coin. They are understandable mechanisms employed by teachers and others when faced with professional demands that are unrealistic, confusing, contradictory, and that undermine their professional knowledge as teachers. Rather than demonstrating a lack of teacher capacity, defensive simplification and cautious implementation reflect a capacity on the part of practitioners to read the political landscape well enough that they could come up with whatever documents were necessary to give the appearance of compliance with regulations. In the case of the *Learning Results*, this meant that school districts aligned their curricula with the standards, at least on paper. What this meant in practice was only loosely defined. In the case of LAS development, we have seen Beaver Pond’s attempts to follow the letter of the law without committing to specifics except in cases where there was general consensus within the group of teachers

affected by the assessment, such as the common writing prompt. Thus teachers in the Beaver Pond district were agreeable to moving ahead with aspects of LAS implementation that they considered beneficial, regardless of the final LAS design.

The assessment system initially envisioned in policy documents such as *Grand Ideas and Practical Work* (Spruce, 1998) that gave rise to Chapter 127 may have been undermined by the *LAS Guide*, but aspects of this vision continued to be a positive influence on the Beaver Pond teachers. In particular, Beaver Pond teachers and administrators expressed gratitude for increased collaborative opportunities with colleagues, and for the opportunities afforded them to reach an understanding of excellence in student work through teacher conversations. The chair of the Beaver Pond assessment committee focused on building teacher capacity in this area as a way of improving teaching, learning, and assessment. Other research participants supported her efforts. Because of their success in this area, many of the teachers on the assessment committee focused on building a local system that included practices that were important to them. This led to an increased confidence in their ability to push local policy decisions in desired directions. The Beaver Pond committee realized they could be proactive about some aspects of LAS design rather than waiting for the state to interpret every last detail. As Fullan (2001) notes, this is an important aspect of local capacity.

The positive conversations about assessment that were established because of the committee's work were not enough to compensate for all the concerns that loomed just below the surface during the period of this research. The combination of LAS and NCLB left teachers with increasing concerns about the use of testing and the

increasing standardization of assessments and assessment procedures, even outside of standardized testing. The amount of control emanating from the state and federal levels frustrated practitioners at the local level who regarded these intrusions as dismissive of their abilities as professionals. Local practitioners also were very concerned about individual students' abilities to function in the new system of accountability on demand. Teachers worried about both the effects of testing on students who would not be able to meet the standards and the futures of those students, particularly those for whom high school graduation might be in question.

In the words of one of the Beaver Pond assessment committee members about the LAS, "Why do teachers keep doing this to themselves?" Smyth's argument reworks this statement as he points out that teachers are not really doing it to themselves. As this discussion has shown, Maine's development of both the *Learning Results* and the Chapter 127 legislation have increased outside control of teachers' work and decision making under the guise of local control. Nevertheless, teachers seem to be reluctant to address these important issues from the vantage point of their professional judgment and experience. This is somewhat explained by the politics and rhetoric of reform; it is difficult to mount an argument against something called No Child Left Behind, especially when that piece of legislation passed with overwhelming bipartisan support. Teachers who critique the reforms appear to be afraid of being held accountable, and they appear to stand in the way of reform.

Accusations that teachers do not want to be held accountable have stuck, and teachers have had difficulty redefining accountability in ways that are meaningful for teaching and learning. Contrary to rhetoric that paints teachers as unintelligent and

incapable laggards who want to avoid any examination of their performance, the teachers who participated in this study were not opposed to accountability *per se* (Selingo, 2004). They were open to accountability in the form of standards as a way of confirming that they were presenting curriculum that was deemed important by the constituencies that schools serve. One of the initial surprises of this research was that all teachers interviewed were mostly positive about the *Learning Results*.

Accountability can serve as a mechanism that validates teachers' work in an ill-structured domain, and as such can assure teachers that their work is valued and appreciated.

This validation is not without its price. Under such a system, the definition of good teaching is beyond teachers' control. Teachers' personal professional knowledge of students and their needs is marginalized if it does not fit within the accountability paradigm. Thus teachers who were competent users of classroom assessment for purposes of instruction doubted their abilities concerning assessment because their definition of learning was more complex than the measurement of discrete skills. The accountability system did not value their abilities as assessors, but the problem was within the system and its definition of learning and assessment, not with the teachers.

Teachers view children and their abilities in holistic fashion and not as the sum of their negatives. They also understand the importance of factors other than academic, such as emotional well-being. Current practices force teachers, against their better judgment, to reduce their perceptions and knowledge of students' backgrounds, prior knowledge, interests and dispositions to an evaluation of learning

that gives us only part of students' abilities, academic or otherwise. Teachers resent the disregard for their professional judgment that is part and parcel of a system with an increasing emphasis on standardization. Recall one participant's succinct remark, "We're not training elephants here."

Maine teachers are especially wary of accusations of negativity. Having been through the institution of the MEAs and the adoption of the *Learning Results*, many teachers have grown weary of having their voices silenced through accusations that they are excessively negative. Some of the tendencies toward defensive simplification and cautious implementation stem from this. As one teacher described her approach to the LAS, "I just give them a little bit. If they look at it, and they decide they need more, let them come back and ask me" (Aileen Koch, personal communication, 4-09-03).

The issue of why teachers have not mounted a full-scale critique of the Chapter 127 legislation and what it is doing to the curriculum, students, and themselves as stressed-out individuals is complex and raises a different sort of capacity issue that is at the heart of our educational system. Discussed here are four areas of concern that help explain teachers' reluctance to challenge policy and assume the mantle of change agents.

First, there is little public discourse about the assumptions underlying the knowledge enterprise that is the American educational system. To what extent does a consensus exist about the most basic assumptions concerning schooling? In this research, e.g., basic questions about the purpose and scope of the *Learning Results* were never definitively answered. As Kliebard's insightful analysis of the history of

American curriculum demonstrates, contradictory visions of the purpose of schooling have co-existed at least since the growth of public schools (1995). Teachers should not be expected to make sense of an issue that the wider public is not willing to tackle.

Second, research has shown that school culture is impervious to change (Tyack & Cuban, 1995). One aspect of this culture is teacher as transmitter of knowledge. In spite of the best attempts of teacher education programs to educate teachers to be change agents, the real enculturation of teachers occurs in classrooms and schools. The day to day ability to participate in school culture is acquired on the job and tends to override new teachers' awareness of other possible practices (Featherstone, Gregorich, Niesz & Young, 1995).

Third, to what extent are teachers, with their professional judgments based on practice and observation, marginalized in academic discourse as well as from a policy standpoint? Action research, for example, may be a great tool for individual teacher reflection and action, but it does not lead to an examination of the assumptions behind policy decisions. Most professional development continues to target the practicalities of educational reform, not the decisions that underlie the policies. Teachers who expand their ability to think critically about educational matters by obtaining advanced degrees frequently leave the classroom and thus do not have the maximum impact on their communities of practice.

Finally, teachers must be practical. A refusal to implement a required practice can result in an official sanction or worse. Historically, teachers were expected to be defenders of the status quo and transmitters of the common culture, not critics of the

system (Tyack & Cuban, 1995). This perception continues to be part of the culture of schooling.

Teachers have the capacity, both personal and professional, to adapt to many types of individuals and many situations. They are accustomed to operating in tangled and uncertain domains. Forced adherence to a standardized system that obscures individual student traits and abilities and that teachers consider counterproductive might not produce outright teacher rebellion, but it will produce an overly cautious and ambivalent approach to curriculum, instruction and assessment that will serve none of us well. The proof of successful educational reform is not in the language of the reform itself. It is in the enactment in the classroom, and, one hopes, in the quality of the students that emerge.

#### *Usefulness of the Research in Other Contexts*

This examination of LAS design reflects the individual and collaborative attempt to make sense of a complex policy mandate. It is one story. What lessons drawn from this context might be applied to other reform efforts, particularly regarding the explanatory framework of capacity?

It must be noted that the combination of state standards, a standardized test, and the LAS framework may be specific to Maine. Certainly Maine policy makers have maintained that their attempt to establish a system of local assessments is without precedent (*LAS Guidelines, 2003, p.1*). In these days of standards and standardization, however, many states are experimenting with similar policies. In addition, issues of capacity affect all reform efforts.

With regard to capacity, two conclusions from this study are noteworthy in a broader context. First, to a large extent, capacity is still viewed in terms of teachers' ability to act on reform. Teacher agency is frequently not acknowledged as an element of capacity. Certain actions taken by teachers, such as the ones discussed in this study, may seem to indicate a lack of capacity when instead they protect teacher agency. Second, more attention needs to be paid to the initial capacity of individuals and districts when implementing reform. Beaver Pond teachers were able to articulate a vision of the LAS that served their purposes. Perhaps other districts were not so fortunate. A more complex picture of capacity has implications for implementing reform. We can not assume that reforms will be understood and acted upon in the same way by all.

A final conclusion from this study that has wider implications involves the detrimental effects of standardization. Recall that Maine policy makers wanted to avoid making the state standardized test "high stakes." The standardization requirements in the LAS, however, have been an enormous burden on teachers and have sapped creativity and innovation in teaching. In addition, standardization ignores the range of interests and abilities that students as learners bring to the classroom.

#### *Limitations of the Study*

A study, even a modest one such as this, needs to give one a sense of depth concerning the context. The inclusion of more districts in the study would have

increased the understanding of the relationship between LAS design and local capacity.

A larger frustration, however, was the necessity to create a rather arbitrary time frame for the research. Maine school districts are still in the process of designing and implementing local assessment systems. The school district that served as the case study for this research was fortunate to have leaders who were knowledgeable about assessment and who managed to design a local system that contained assessments that were viewed as useful by most of the teachers. Other systems with less knowledgeable personnel may have been more influenced by state guidelines and may have crafted local assessment systems that are far more cumbersome. A local assessment implementation study recently completed by the MDoE indicated that design and implementation of local systems varies widely from district to district. (LASIS Final Report, 2004). The design of this research does not permit comparisons across districts.

In addition, anecdotal evidence continues to accumulate that indicates that the frustrations of teachers and administrators that were emerging in this research have grown. The additional testing required by No Child Left Behind, the continuing efforts of the Maine DoE to bring standardization to the local assessment systems, and the efforts to tie high school graduation to achievement of the *Learning Results* as demonstrated by adequate performance on the assessments that comprise the local system are beginning to take their toll on Maine teachers (Walker, 2004). However, providing support for this contention is outside the current scope of this research.

The troublesome educational outcomes of standardized testing are well known and documented (McNeil, 2000; Madaeus & O'Dwyer, 1991; Kohn, 2000). The present study shows the beginning of the pernicious effects of standardizing of curriculum and assessment, even with minimal high stakes standardized testing. As such it joins other cautionary tales of standardization outside the framework of standardized testing, such as descriptions of Kentucky's writing assessment (Whitford and Jones, 2000). Certainly one should not assume that lack of high stakes standardized testing means that the ill effects of standardization are not present.

The Maine DoE is proud of their pioneering status in developing an assessment system that relies minimally on standardized testing as the means to identify student achievement (*LAS Guide*, 2003). However, the interrelatedness of the *Learning Results*, LAS and MEA epitomize the hyperrationalization that Wise (1979) cautioned against. The cautionary note from the present study is that unanticipated negative effects such as narrowing the curriculum, decreased instructional and curricular creativity, and challenges to the professionalism of teachers may be present in any attempt to tightly link all areas of the school experience in the name of accountability. The use of the rational model in an educational enterprise ultimately constricts and/or eliminates the very attempts at creativity and innovation that are among the strongest aspects of schooling in the United States.

In one of the ironies of the current accountability trend, people in the world of business who were among the most concerned about the supposedly poor performance of U.S. students may ultimately be even less satisfied with the

performance of future high school graduates. As we have noted in the discussion about defensive simplification, as schools struggle with accountability guidelines that pertain to all students, the tendency is to sharply focus one's efforts in order to be able to have the maximum number of students meet the accountability standards. Creativity and lofty goals do not thrive in such an atmosphere. Instead, we should foster creativity and encourage innovation, reward teachers for their professionalism, and look for ways other than standardized systems to determine accountability. Innovation and creativity in education will not thrive if teachers are pressured to have all students reach some artificially determined standard and also threatened with sanctions if all students do not reach this target.

#### *Further Research*

This study was undertaken at a time when Maine schools were just beginning to grapple with the magnitude of designing an assessment system that would fulfill the requirements of the Chapter 127 legislation. With regard to the progress of LAS design by school districts in Maine, two findings emerge. First, the Maine DoE, which initially thought to straddle both formative and summative assessment in one system, was forced to recognize that a single assessment system could not be designed that could focus equally on both assessment for learning and accountability (*LAS Guide*, 2003). Second, the expertise of teachers and local administrators brought to local system design was superseded by onerous demands that these local assessment systems meet technical standards regarding validity and reliability that were determined by the MDoE. Maine teachers are continuing to spend many hours

on the design of local assessment systems, and the state continues to publish more restrictive guidelines about what local systems can and cannot contain and how to attain validity and reliability in ways that are time consuming for teachers. Questions to inform further research might be: What are the continuing effects of this particular focus on assessment in Maine classrooms? How are teachers juggling the requirements of the LAS with the testing requirements of No Child Left Behind? With all this emphasis on the basics and standardization, what is happening to innovation, creativity, and individualization in the classroom? How are different student populations faring, such as at-risk or gifted? How are teachers responding to the plethora of demands on their time? Are local assessment systems being designed in ways that give teachers and parents important information in a timely fashion, or are the assessments being given in “check off” fashion after cursory attempts at instruction? What factors contribute to the similarities and differences of LAS design across districts?

Given the rhetoric of No Child Left Behind, it is truly ironic that some Maine school districts are contemplating instituting two levels of diplomas, one for students who meet the requirements of No Child Left Behind and can certify that they have achieved the *Learning Results*, and another diploma or certificate for those who attended classes but are unable to pass certain requirements (Marla Smith, personal communication, 12-14-04). Chapter 127 and No Child Left Behind are similar in their attempts to hold all students to high standards, but they differ greatly in how they aspire to reach that goal. How has this dissonance affected local assessment system design and implementation?

The larger issue behind all of this is the oft-repeated perception that U.S. schools are doing a poor job, and so drastic reforms are necessary. In the case of Chapter 127 legislation, the ultimate question about validity will be if the constituencies who were critical of the K-12 schooling of Maine students indeed find students who graduate under the new system to be superior. Or will the tendencies toward defensive simplification and cautious implementation noted in this study hold sway, creating a dumbing down effect and ultimately a backlash against standardized testing and the tight circle of control that now exists with standards, curriculum and assessment? Will we see an exodus to private schools as parents become disgusted with a lack of challenging curriculum as teachers feel forced to aim for the least common denominator? Will populations in special education classes continue to increase rapidly as parents and teachers search for ways to justify why students cannot meet the standards?

Capacity as an explanatory mechanism would be useful in examining responses to the LAS across districts. The Beaver Pond committee, for all the frustrations associated with the work on assessment, was able to articulate to some extent what they did and did not want and thus shape the LAS. Other districts that lacked this capacity may have assessment systems that are more burdensome and contribute less information that teachers would use in informing instruction.

Maine officials have maintained that the development of local assessment systems under Chapter 127 guidelines is immune from the criticisms leveled at systems that use high stakes standardized tests. Indeed, the suggested role of the MEA in the local assessment system is minor compared to the high stakes decisions

that accompany such tests in other states. The problem is not with standardized tests *per se*, but with standardization and its ill effects on schooling in general.

Standardization marginalizes local practitioners' knowledge about children, child development, curriculum, assessment and the local context of education in their communities. It narrows our vision of educating the whole child into a one-dimensional, lock-step ability to meet certain standards and pass certain tests or assessments at prescribed times. It threatens the integrity of teachers' professional knowledge. The concerns expressed by participants in this study will only increase as the standardization of curriculum and assessment in Maine becomes more prevalent.

## References

- Airasian, P. (1991). Perspectives on measurement instruction. *Educational Measurement: Issues and Practice*, 10(1) pp. 13-16.
- Archbald, D. & Newmann, F. (1988). *Beyond standardized testing: Academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.
- Assessment Training Institute (Producer). (2002). *Assessment for learning: A hopeful vision of our assessment future* [Motion picture]. (Available from the Assessment Training Institute, 317 SW Alder Street, Suite 1200, Portland, OR 97204)
- Beirne, M. (1997). Seeing the invisible: Classroom observation. *Momentum*, 28 (2) pp. 34-38.
- Berlak, Harold. (1992). The need for a new science of assessment. In Allen R. Tom (Ed.), *Toward a new science of educational testing and assessment* (pp. 1-23). Albany, NY: SUNY Press.
- Borko, H. & Putnam, R (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. Guskey & M. Huberman, (Eds.), *Professional Development in Education: New Paradigms & Practices*. (pp. 35-66). New York: Teachers College Press.
- Brooks, J. & Brooks, M. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Bull, Barry. (2000). National standards in local context: A philosophical and policy analysis. In B. Jones, (Ed.), *Educational Leadership: Policy dimensions in the 21<sup>st</sup> century* (pp. 107-121). Stamford, CT: Ablex Publishing Corporation.
- Cherryholmes, C. (1988). *Power and criticism: Poststructural investigations in education*. New York: Teachers College Press.
- Christie, K. (2001). Learning from the experience of others. *Phi Delta Kappan*, 83, 105-106.
- Cobb, P. (1996). Where is the mind? A coordination of sociocultural and cognitive constructivist perspectives. In Catherine Fosnot (Ed.), *Constructivism: Theory, Perspectives & Practice*. (pp. 20-49). New York: Teachers College Press.
- Coldarci, Theodore, Johnson, J., Beaudry, J., Cormier, M., Ervin, R., Rosenblum, J., & Silvernail, D. (2000). *Measured measures: Technical considerations for developing a local assessment system*. Augusta, ME: Maine Department of Education.
- Conley, D. & Goldman, P. (1998). How educators process and respond to state level educational reform policies : The case of Oregon. San Diego, CA: AERA conference, April 13-17, 1998.
- Connelly, F. & Clandinin, D. (1988). *Teachers as curriculum planners: Narratives of experience*. New York: Teachers College Press.
- Darling-Hammond, Linda. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64, 5-30.

- Datnow, Amanda & Castellano, Marisa. (2000). Teachers' responses to Success for All: How beliefs, experiences, and adaptations shape implementation. *American Educational Research Journal*, 37, 775-800.
- Davies, A., Cameron, C., Politano, C., Gregory, K. (1992). *Together is better: Collaborative assessment, evaluation and reporting*. Winnipeg, Canada: Peguis Publishers.
- Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy, Practice*, 8(2), 113-133.
- Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, 104, pp. 1461-1484.
- Delandshere, G. & Jones, J. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum & Supervision*, 14, 216-240.
- Driscoll, M. (1994). *Psychology of learning for instruction*. Needham Heights, MA: Allyn & Bacon.
- Elementary & Secondary Education Act, U.S.C. #PL 107-110, 2001.
- Fairman, J.C., & Firestone, W.A. (2001). The district role in state assessment policy: An exploratory study. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states*. (pp. 124-147). Chicago: University of Chicago Press.
- Featherstone, H., Gregorich, P., Niesz, T. & Young, L. (1995). The scary part is it happens without us knowing. East Lansing, MI: National Center for

- Research on Teacher Learning, Michigan State University.
- Floden, R., Goertz, M., & O'Day, J. (1995). Capacity building in systemic reform. *Phi Delta Kappan*, 77, 19-21.
- Franke, M., Carpenter, T., Levi, L. & Fennema, E. (2001). Capturing teachers' generative change: A follow-up study of professional development in mathematics. *American Educational Research Journal*, 38, 653-690.
- Fuhrman, S., Ed. (2001). Conclusion. In *From the capitol to the classroom: Standards-based reform in the states*. (pp. 263-278). Chicago: University of Chicago Press.
- Fullan, M. (2001). *The new meaning of educational change* (3<sup>rd</sup> ed.). New York: Teachers College Press.
- Ginsburg, H. & Opper, S. (1988). *Piaget's theory of intellectual development* (3<sup>rd</sup> ed.). Englewood Cliffs, N.J.: Prentice Hall.
- Gipps, C.V. (1994). *Beyond testing*. London: Falmer Press.
- Gipps, C.V. (1999). Socio-cultural aspects of assessment. In A. Iran-Nejad & D. Pearson (Eds.), *Review of Research in Education* (pp. 355-392). Washington, DC: American Educational Research Association.
- Gipps, C.V., Brown, M., McCallum, B., McAlister, S. (1995). *Intuition or evidence? Teachers and national assessment of seven year olds*. London: University Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.
- Glesne, C. (1999). *Becoming qualitative researchers: An introduction*. New York:

Addison Wesley Longman.

Goals 2000: Educate America Act. P.L. 103-227, 1994.

Goertz, M. (2001). Standards-based accountability: Horse trade or horse whip? In

S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states*. (pp. 39-59). Chicago: University of Chicago Press.

Goertz, M. & Duffy, M. (2001). Assessment and accountability systems in the 50 states: 1999-2000. University of PA Graduate School of Education:

Consortium for Policy Research in Education.

<http://www.cpre.org/Publications/rr46.pdf>. (Retrieved 3-12-04).

Goldman, C. (1997). What is a local assessment system and who needs it anyway?

(Available from the commissioner of education, Maine Department of Education, Augusta, ME 04333).

Hargreaves, A. (1995). Development and desire: A postmodern perspective. In T.

Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms & practices*. (pp. 9-34). New York: Teachers College Press.

Hargreaves, A.(Ed). (1997). *Rethinking educational change with heart and mind*.

Alexandria, VA: ASCD Yearbook.

Hargreaves, A., Earl, L. & Schmidt, M. (2002). Perspectives on alternative

assessment reform. *American Educational Research Journal*, 39 (1) pp. 69-100.

Jamentz, K. (1994). Making sure that assessment improves performance. *Phi Delta*

*Kappan*, 76, 55-57.

Jett, D. & Schafer, W. (1993). Ready or not, teachers K-12 move to center stage in

- the assessment arena: Implications for state educational policymakers.  
*Measurement & Evaluation in Counseling & Development*, 26 (1) 69-80.
- Kannapel, P., Aagaard, L., Coe, P. & Reeves, C. (2001). The impact of standards and accountability on teaching and learning in Kentucky. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states*. (pp. 242-262). Chicago: University of Chicago Press.
- Kliebard, H. (1995). *The struggle for the American curriculum: 1893-1958*. 2<sup>nd</sup> Ed. New York: Routledge.
- Kohn, A. (2000). *The case against standardized testing*. Portsmouth, NH: Heinemann.
- Koretz, D., McCaffrey, D., Klein, S., Bell, R. & Stecher, B. (1996). The Reliability of scores from the 1992 Vermont portfolio assessment program. Rand Institute on Education and Training. Cresst. Retrieved from <http://www.cresst96.cse.ucla.edu/cresst/reports/TECH355.pdf>. 4-20-04
- Landman, J. (2000). A state mandated curriculum, a high stakes test: One MA high School history department's response to a very new policy context. Unpublished qualifying thesis, Harvard University.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Lewis, A.C. (1999). Educational policy analysis: The treads behind, the trends ahead. In G. Cizek, (Ed.), *Handbook of Educational Policy*. (pp. 521-531). New York: Academic Press.
- Lieberman, A. (1996). Practices that support teacher development: Transforming

- conceptions of professional learning. In M. McLaughlin & I. Oberman, Eds., *Teacher learning: New policies, new practices*. pp. 185-223. New York: Teachers College Press.
- Linn, Robert. (1995). *Assessment-based reform: Challenges to educational measurement*. Princeton, NJ: Educational Testing Service.
- \_\_\_\_\_ (2000). Assessments and accountability. *Educational Researcher*, 29,(2), 4-16.
- Local Assessment Development. <http://www.maineled.org>.
- Lopez, O.S. (1995). The effect of the relationship between classroom student diversity and teacher capacity on student performance. Executive summary prepared for The Strategic Classroom Learning Enterprise. Austin, TX.
- Madaus, George & O'Dwyer, Laura. (1991). A short history of performance assessment. *Phi Delta Kappan*, 80, 688-695.
- Maine Assessment Portfolio. <http://www.maptasks.org>
- Matthews, B. (1995). The implementation of performance assessment in Kentucky classrooms. Louisville, KY: University of Louisville School of Education.
- McDermott, R. (1996). The acquisition of a child by a learning disability. In S. Chaikin & J. Lave (Eds.), *Understanding practice: Perspectives on activity & context* (pp. 269-305). Cambridge, UK: Cambridge University Press.
- McDonnell, L. (1994). Assessment policy as persuasion and regulation. *American Journal of Education*, 102, 395-420.
- McLaughlin, M. (1993). Rand change agent study revisited. *Educational Researcher*, 19(9), 11-16.

- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- \_\_\_\_\_ (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher* 23(2) 13-23.
- Miles, M. & Huberman, A.M. (1988). Drawing valid meaning from qualitative data: Toward a shared craft. In D. Fetterman (Ed.), *Qualitative Approaches to Evaluation in Education*. (pp.222-243). New York: Praeges Press.
- Miles, M. & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2<sup>nd</sup> Ed.). Thousand Oaks, CA: Sage.
- Miller, L. & Smith, D. (2002). Southern Maine Partnership's Comprehensive Local Assessment System Primer. Retrieved 2-04-03 from <http://www.elm.maine.edu/assessments/class/primer>.
- Millman, J. (Ed). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 243-247). Thousand Oaks, CA: Corwin Press.
- Moss, P. & Schutz, A. (2001). Educational standards, assessment, and the search for consensus. *American Educational Research Journal*, 38, 37-72.
- National Committee on Excellence in Education. (1983). *A nation at risk*. Washington, DC: Author.
- National Council of Teachers of Mathematics. (1991). Professional standards for teaching mathematics. NCTM Commission on Teaching Standards: Author.
- National Science Education Standards: Observe, interact, change, learn. (1995).

National Academies Press: Author.

Nespor, J. (1985). The role of beliefs in the practice of teaching: Final report of the teacher beliefs study. Austin, TX: Research and Development Center for Teacher Education.

Noddings, N. (1992). *The challenge to care in schools: An alternative approach to education*. New York: Teachers College Press.

O'Day, J., Goertz, M., & Floden, R. (1995). Building capacity for education reform. Consortium for Policy Research in Education. Retrieved 7-9-02, <http://www.ed.gov/pubs/CPRE/rb18>.

Field Code Changed

Ogawa, R., Sandhotz, J., Martindz-Flores, M. & Scribner, S. (2003). The substantive and symbolic consequences of a district's standards-based curriculum. *American Educational Research Journal*, 40, 147-176.

Perkins, D. (1999). The many faces of constructivism. *Educational Leadership*, 57 (3) 6-12.

Plake, B., Impara, J. & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 6,(4) 10-12.

Popham, J. (1997). The moth and the flame: Student learning as a criterion of instructional competence. In Jason Millman, (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 264-278). Thousand Oaks, CA: Corwin Press.

Putnam, P. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29 (1), 4-16.

- Ramirez, A. (1999). Assessment-driven reform: The emperor still has no clothes. *Phi Delta Kappan*, 81, 204-208.
- Ravitch, D. (1995). *National standards in American education: A citizen's guide*. Washington, D.C.: The Brookings Institution.
- Resnick, L. & Resnick, D. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford & M. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction*. pp. 37-76. Boston: Kluwer Academic Press.
- Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher*, 19 (7) 10-18.
- Richardson, V. & Placier, P. (2001). Teacher change. In V. Richardson, (Ed.), *Handbook of Research on Teaching* (4<sup>th</sup> ed.) pp. 905-950. Washington, DC: American Educational Research Association.
- Roeber, E. (1999). Standards initiatives and American educational reform. In G. Cizek (Ed.), *Handbook of Educational Policy*. pp. 152-183. San Diego, CA: Academic Press.
- Rowling, J.K. (1998). *Harry Potter and the Sorcerer's Stone*. New York: Arthur A. Levine Books.
- Schafer, William. (1991). Essential assessment skills in the professional education of teachers. *Educational Measurement: Issues and Practice*, 10, 1-12.
- Selingo, J. (2004). Secretary Paige calls National Education Association a "terrorist organization," *The Chronicle of Higher Education* 24 February.

- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7), pp. 5-20.
- \_\_\_\_\_. (2001). The role of classroom assessment in teaching and learning. In V. Richardson, (Ed.), *Handbook of Research on Teaching* (4<sup>th</sup> ed.) pp. 1066-1101. Washington, DC: American Educational Research Association.
- \_\_\_\_\_ (1989). Why we need better assessments. *Educational Leadership*, 46, (7), 4-9.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57, 1-22.
- Slife, B. & Williams, R. (1995). *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, CA: SAGE Publications.
- Smith, M.L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8-11.
- Smylie, M. (1995). Teacher learning in the workplace: Implications for school reform. In T. Guskey & M. Huberman (Eds.), *Professional Development in Education: New Paradigms & Practices*. (pp. 92-113). New York: Teachers College Press.
- Smyth, J. (1995). Teachers' work and the labor process of teaching: Central problematics in professional development. In T. Guskey & M. Huberman (Eds.), *Professional Development in Education: New Paradigms & Practices*. (pp. 69-91). New York: Teachers College Press.

- Snow-Renner, R. (1998). Mathematics assessment practices in CO classrooms: Implications about variations in capacity and students' opportunity to learn. San Diego, CA: AERA conference, April 13-17.
- Soodak, L. & Martin-Kniep, G. (1994). Authentic assessment and curriculum integration: Natural partners in need of thoughtful policy. *Educational Policy*, 8, 183-201.
- Spillane, J. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, 31, 143-175.
- \_\_\_\_\_ (2001). Challenging instruction for "all students": Policy, practitioners, and practice. In S. Fuhrman, (Ed.), *From the capitol to the classroom: Standards-based reform in the states*. (pp. 217-241). Chicago: University of Chicago Press.
- Spillane, J., & Jennings, N. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record*, 98, 449-481.
- Spruce, M., Maine Department of Education & the Northeast and Islands Regional Educational Laboratory. (2000). *Grand ideas and practical work: The Maine local assessment system resource guide*. Retrieved 7-26-04 from: <http://janus.state.me.us/education/g2000/pilot%20assess1.htm>.
- State of Maine. Department of Education. (2003). Balance of representation: Establishing the relative scope and importance of standards to inform curriculum, instruction, and assessment. Augusta, ME: Author.

- State of Maine. Department of Education. Chapter 127: Instructional program, assessment and diploma requirements. 05-071. (2001).
- State of Maine. Department of Education. (2001). Designing a Local Assessment System. Part A: Taking Stock: Examining system status against standards. Augusta, ME: Author.
- State of Maine. Department of Education. (2001). Designing a Local Assessment System. Part B: Using alignment criteria to build validity and reliability into assessment development. Augusta, ME: Author.
- State of Maine. Department of Education. (2003). *LAS guide: Principles and criteria for the adoption of Local Assessment Systems*. Augusta, ME: Author.
- State of Maine. Department of Education. (2002). Local Assessment System model narrative. Augusta, ME: Author.
- State of Maine. Department of Education. (1990). *Maine's common core of learning: An investment in Maine's future*. Augusta, ME: author.
- State of Maine. Department of Education. (2001). Policy Advisory Committee questions and answers. Monograph. Augusta, ME: Author.
- State of Maine. Department of Educational and Cultural Services. Division of Educational Assessment. (1989). *The use and impact of the Maine Educational Assessment, 1985-89*. Augusta, ME: Author.
- State of Maine. (2004). LASIS final report. Retrieved 9-13-04 from [http://www.maine.gov/education/las/LAS/LAS\\_study.htm](http://www.maine.gov/education/las/LAS/LAS_study.htm).
- State of Maine. (1997). *Recommendations for assessing Maine's Learning Results*. Augusta, ME: Author.

- State of Maine *Learning Results*. (1997). Augusta, ME: Maine Department of Education.
- Stiggins, R. & Bridgeford, N. (1985). The ecology of classroom assessment. *Journal of Educational Measurement*, 22, 271-286.
- Stiggins, R. & Conklin, N. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: SUNY Press.
- Terwilliger, J. (1997). Semantics, psychometrics and assessment reform: A close look at "authentic" assessments. *Educational Researcher*, 26(8), 24-27.
- Tillema, H. & Imants, J. (1995). Training for the professional development of teachers. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms & practices*. (pp. 135-150). New York: Teachers College Press.
- Torres, M. (1996). Cognitive individualism: An impediment to teachers' collaborative intellectual work. New York, NY: AERA conference, April 8-12, 1996.
- Tyack, D. & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Walker, R. (April, 2004). Outraged? Organize! Maine Education Association Newsletter, 64 (8) p. 2.
- Wells, G. (1999). *Dialogic inquiry: Toward a sociocultural practice and theory of education*. Cambridge, England: Cambridge University Press.

- Whitford, B., & Jones, K. (2000). How high stakes accountability undermines a performance-based curriculum vision. In B. Whitford & K. Jones, (Eds.), *Accountability, assessment, and teacher commitment: Lessons from Kentucky's reform efforts*. pp. 9-24. Albany, NY: SUNY Press.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46 (7), 41-47.
- Wiggins, G. (1998). *Educative assessment*. San Francisco, CA: Jossey-Bass.
- Wilson, S. & Floden, R. (2001). Hedging bets: Standards-based reform in the classroom. In S. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states*. (pp. 193-216). Chicago: University of Chicago Press.
- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72, 131-175.
- Windschitl, M. & Sahl, K. (2002). Tracing teachers' use of technology in a laptop computer school: The interplay of teacher beliefs, social dynamics, and institutional culture. *American Educational Research Journal*, 39, 165-206.
- Wise, A. (1979). *Legislated learning: The bureaucratization of the American classroom*. Berkeley, CA: University of California Press.
- Wise, S.L., Leslie, E. & Roos, L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, 42, (1), 37-42.
- Wolf, D. (1993). Assessment as an episode of learning. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in*

*constructed response, performance testing, and portfolio assessment* (pp. 213-240). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Wolf, D., Bixby, J., Glenn, J. III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), *Review of Research in Education* (pp. 31-74). Washington, DC: American Educational Research Association.

Wolf, D. & Reardon, S. (1996). Access to excellence through new forms of student assessment. In J. Baron & D. Wolf (Eds.), *Performance-based assessment: Challenges and possibilities*. (pp. 1-31). Chicago: University of Chicago Press.

Wolf, S., Borko, H., Elliott, R. & McIver, M. (2000). "That dog won't hunt!": Exemplary school change efforts within the Kentucky reform. *American Educational Research Journal*, 37, 347-396.

← - - - Formatted: Indent: First line: 36 pt

← - - - Formatted: Indent: First line: 36 pt

Wolf, S. & Gearheart, M. (1997). New writing assessments: The challenge of changing teachers' beliefs about students as writers. *Theory into Practice*, 36, 220-231.

## Appendix A: List of Interviews, Observations and Meetings

### *Interviews*

John Wilkins: 7-30-01; 10-03-01; 10-28-01; 1-23-02; 1-30-02; 11-23-02; 3-05-03; 3-12-03.

Jessica Bean: 5-07-03

Marla Smith: 10-26-01; 11-05-02; 7-09-03; 2-24-04; 3-03-04

Elizabeth Traynor: 9-19-01; 10-03-01; 1-30-02; 2-13-02; 5-22-02; 10-07-02; 10-23-02; 3-05-03

Emily Baxter: 2-10-03

Susan Howard: 5-01-02

Polly Benton: 3-10-03; 4-07-03

Sophia Laken: 12-03-01; 3-06-02; 2-18-03; 3-12-03

Don Atkins: 10-14-01

### *Observations*

John Wilkins: 15 observations during 2001-02 school year.

Elizabeth Traynor: 8 observations between 9-01 and 5-03

Polly Benton: 3-10-03; 4-07-03

Sophia Laken: 12-03-01; 3-06-02; 3-12-03

### *Meetings*

Beaver Pond Assessment Committee Meetings: 2-13-02; 9-20-02; 2-05-03; 2-27-03; 4-02-03; 5-07-03; 6-04-03; 6-16-03

Grade level committee meetings: 1-24-02; 4-14-03; 4-28-03; 4-30-03.

## Appendix B: Sample Interview Questions for Teachers

1. What sorts of assessments do you use on a daily basis? What sorts of records do you keep of these?
2. Do you assess for different purposes? Which assessments do you use to give feedback to students? Parents? Others?
3. What are your expectations regarding standards for your students? How do you communicate your expectations to students? Do your students participate in the assessment or reporting process at all e.g. student-led conferences? What opportunities for self-assessment do you offer them?
4. Do you assess differently now than when you started teaching? Do you report to parents differently? Can you give me some examples? What contributed to the changes?
5. What is your best source of information on assessment? Has this changed over the years?
6. What are some new curricular and/or instructional ideas you have tried recently? Have any of these had an impact on the way you assess?
7. Have you implemented any of the assessments that are part of the LAS? Please elaborate. How has this changed your teaching?
8. How have the MEA and/or the Learning Results affected your teaching?
9. Are there some things that the MEA does well? Other things it does less well? How about the Learning Results?

**Purposes of Assessment**

**Our Beliefs**

**INFORMS LEARNERS/PROVIDES FEEDBACK**

- allows for self evaluation and provides evidence of learning
- helps students set goals for themselves by identifying personal strengths/weaknesses
- allows for new connections and learning to occur through the assessment itself

**GUIDES INSTRUCTION**

- shows effectiveness of curriculum and/or instruction
- shows what students already know and where to begin instruction
- helps guide teachers in decisions re: curriculum & instruction
- monitors student understanding/development of a concept or skill

**MEASURES STUDENT PROGRESS**

- documents student growth
- determines to what extent a student has met a standard
- certifies achievement of *Learning Results*

**PROVIDES DATA WHICH DRIVES DECISION-MAKING**

- shows trends & establishes norms over time
- provides direction for school improvement
- gives feedback to students, parents, teachers

## Rebecca H. Berger

147 Vienna Road  
New Sharon, ME 04955  
207-778-4704

[Rebecca.berger@maine.edu](mailto:Rebecca.berger@maine.edu)

### Profile

Five years experience college and university teaching as adjunct professor, assistant instructor, and supervisor of student teachers; fourteen years teaching experience in early childhood and elementary settings; Indiana endorsement in the teaching of gifted and talented students; lifetime teaching licenses in Indiana and New York.

### Education

- |         |   |
|---------|---|
| 2005    | Ph.D. Indiana University; School of Education<br>Curriculum Studies with a minor in Inquiry/Assessment<br>Dissertation title: <i>Teacher Capacity and Assessment Reform: Assumptions of Policy, Realities of Practice</i> |
| 1997-98 | Indiana University Fellowship Recipient.  |
| 1991    | Certificate in Gifted and Talented Education<br>Purdue University<br>Lafayette, IN  |
| 1979    | M.S., Elementary Education<br>Bank Street College of Education<br>New York, NY<br>Thesis: <i>Bicultural Education in Highland Sardinia</i>  |
| 1973    | A.B., History and Philosophy of Religion<br>University of Chicago<br>Chicago, IL  |

### Professional Experience

- |           |   |
|-----------|---|
| 2005-     | University of Maine at Farmington, Assistant Professor of Education |
| 2001-2005 | University of Maine at Farmington, Instructor of Education          |
- Teaching Responsibilities  
*Introduction to Theory and Practice in Elementary Education*  
*Elementary Social Science Multicultural Education*

University Committee Work  
*Interdisciplinary General Education, 2004-*  
*Human Subjects Review Committee, 2005-*

State-level Service  
*Maine Council of Social Studies Executive Committee*  
*Member 2002-*  
*Vice-president 2003-*  
*Learning Results Review Advisory Committee*  
*Member 2004-*

1997-2000 Indiana University, Assistant Instructor

Teaching Responsibilities  
*Social Studies for Early Childhood*  
*Curriculum Models in Early Childhood Education (Masters*  
*Level) - offered via video-conferencing.*

Supervisory Responsibilities  
Responsible for Supervision of Student Teachers and  
co requisite seminar

1997-2000 Franklin College, Franklin, IN, Adjunct Professor of  
Education

Teaching Responsibilities  
*History and Philosophy of Early Childhood Education*  
*Health and Physical Education for Elementary Teachers*

Supervisory Responsibilities  
Supervisor for junior year multicultural field experience

1995-6 Grade Four Teacher  
Rensselaer Central School Corporation  
Rensselaer, IN

1994-5 Professional Leave of Absence (student at Lancaster  
Theological Seminary)  
Lancaster, PA

1982-94 Kindergarten Teacher  
Rensselaer Central School Corporation

1978-9                    Grade Three Teacher  
West Central School Corporation  
Francesville, IN

1974-6                    Assistant Teacher, Kindergarten and After School Program  
Lexington Children's Center  
New York, NY

### Papers and Presentations

“You Don't Always Get What You Want:  
Hyperrationalization and School Reform” October, 2005  
UMF Faculty Forum

“Problematizing a Gendered Approach to an Ethic of Care:  
Issues for Educators” October, 1999  
Journal of Curriculum Theorizing Conference, Dayton, OH

“Putting It All Together: Pre-Service Teachers Tackle  
Assessment” September, 1999  
Poster presentation at Indiana University's School of  
Education annual retreat

“The Ethic of Care: Implications for Schools” April, 1999  
Kentucky State University 10th Interdisciplinary  
Conference on Science and Culture

### Grants

*The Wabanaki: Exploring the Past, Imagining the Future.* Conference held at University of Maine at Farmington September 27-28, 2005. Funded in part by the Maine Humanities Council. I was conference organizer and director.

*We Can Work It Out: Cooperative Collaboration for Sustainable Science Enrichment.* Two year grant received from the Indiana Commission for Higher Education, Eisenhower Mathematics and Science Education Program, 1992. I assisted in this project by writing curricula, training teachers, and collecting data.

## **Professional Organizations**

American Educational Research Association  
Association for Supervision and Curriculum Development  
National Education Association  
Maine Council for Social Studies  
*Vice President, September 2003-present*